Phonology and Second Language Acquisition

EDITED BY

Jette G. Hansen Edwards

Mary L. Zampini

STUDIES IN BILINGUALISM

36

JOHN BENJAMINS
PUBLISHING COMPANY

Phonology and Second Language Acquisition

Studies in Bilingualism (SiBil)

The focus of this series is on psycholinguistic and sociolinguistic aspects of bilingualism. This entails topics such as childhood bilingualism, psychological models of bilingual language users, language contact and bilingualism, maintenance and shift of minority languages, and socio-political aspects of bilingualism.

Volume 36

Phonology and Second Language Acquisition
Edited by Jette G. Hansen Edwards and Mary L. Zampini

Phonology and Second Language Acquisition

Edited by

Jette G. Hansen Edwards
The Chinese University of Hong Kong

Mary L. Zampini
Le Moyne College

John Benjamins Publishing Company
Amsterdam / Philadelphia

∞™ The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Phonology and second language acquisition / edited by Jette G. Hansen Edwards, Mary L. Zampini.

p. cm. (Studies in Bilingualism, ISSN 0928-1533 ; v. 36)

Includes bibliographical references and index.

1. Second language acquisition. 2. Grammar, Comparative and general--Phonology. I. Edwards, Jette G. Hansen. II. Zampini, Mary L.

P118.2.P52 2008

418--dc22 2007046500

ISBN 978 90 272 4147 4 (Hb; alk. paper)

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands

John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

Introduction

Jette G. Hansen Edwards and Mary L. Zampini

Overview

The past three decades have witnessed a resurgence and growth in interest in research on the acquisition of a second language (L2) sound system, along with an ever-expanding and changing repertoire of techniques and models for studying L2 speech. Major research findings have shown that predicting areas of difficulty and explaining L2 phonological acquisition is much more complex than a straightforward contrastive analysis of the first language (L1) and the second (Lado 1957). Research has also shown that there are numerous factors that affect the relative level of ease or difficulty in L2 phonological acquisition, as well as the relative accuracy (or "nativeness") of L2 speech, that go far beyond a general consideration of the learner's age at the onset of acquisition. In addition, technological advances have changed the ways in which researchers collect their data and conduct their analyses as well as develop pedagogical applications, especially in recent years. These advances range from the emergence of faster and more powerful computers, along with inexpensive, and even free, software for detailed speech analysis, to the growth and availability of more sophisticated and safer techniques for the examination of neurophysiologic aspects of speech. Despite these advances, however, there have been very few works that have provided a broad and thorough overview of the field of L2 phonology. This book attempts to fill that gap by providing a comprehensive discussion of state-of-the-art research in L2 phonology. It contains 13 chapters written by experts in the field, each devoted to a particular issue that is crucial to our understanding of the way learners acquire, learn, and use an L2 sound system. In addition, it spans both theory and application in L2 phonology. Many of the chapters devoted to research also address the implications of the findings for applied linguistics and the teaching of pronunciation, while other chapters are more centrally devoted to issues related to training and teaching.

The study of the acquisition and teaching of L2 phonology is vast, not only in the breadth of the foci – from L1 transfer to the use of ultrasound equipment in training – but also in the depth of work in each area of focus and, therefore, it is impossible to address everything in one volume. However, this volume attempts to provide a broad, but rich, overview of key approaches and advances in this field. As such, the volume touches on, if not directly addresses, significant theories and models in L2 phonology in specific, as well as theories and models in linguistics and second language acquisition (SLA) in general, since many of the major constructs in L2 phonology reflect theories and research in linguistics and SLA at a given point in time. In order to provide a framework for the concepts discussed in this volume, a brief overview of major constructs in L2 phonology is given below.

Major constructs in L2 phonology

The construct of transfer – the effect of previously learned languages on subsequently learned languages – has held a major role in theory construction and pedagogical developments in L2 phonology since work first began being documented in this area in the 1950s (e.g., Lado 1957; Weinreich 1953). Since that time, transfer has been considered to be a dominant influence, both positively and negatively, in the acquisition of an L2, specifically in the area of phonology. While other domains of SLA research such as morphology, syntax, and pragmatics have also focused on transfer, it is within the domain of L2 phonology that transfer has been most heavily researched, due to the recognition that it is within this area of acquisition that transfer is most prevalent.

The construct of transfer has undoubtedly existed long before it began to be documented in the linguistics and SLA literature. This was not done with great consistency, however, until the work of Fries (1945), Weinreich (1953), and Lado (1957), which led to the development of the Contrastive Analysis Hypothesis (CAH). The CAH focused on error production, specifically their explanation and prediction, and grew out of structural linguistics and behaviorism. In its most basic form, the CAH predicted that those aspects (or features) of the L2 that were similar to the L1 would be easy to acquire, while those aspects that were different in the two languages would be difficult to acquire. From a pedagogical standpoint, therefore, the goal of contrastive analysis was to compare languages based on their features (and in particular, contrastive features for phonetics), and teach those L2 features which were different from the L1 in a series of drill-based activities aimed at creating 'good' habits (e.g., correct pronunciation) and getting rid of 'bad' habits (e.g., incorrect pronunciation); this was done under the methodological 'leg' of behaviorism, the Audiolingual Method (ALM). Structuralism and behaviorism, and consequently the CAH and ALM, began to decline in popularity in linguistics and SLA in the late 1950s. This was due in part to Chomsky's (1957) groundbreaking work shifting the focus from behaviorist to cognitive approaches in linguistics, as well as lack of empirical support for the tenets of the CAH. Nevertheless, contrastive analysis still holds some sway in L2 phonology theory and pedagogy. To give just one example, many L2 pronunciation texts and pedagogical guides, such as *Learner English: A Teacher's Guide to Interference and Other Problems* (2001, Swan & Smith, Cambridge University Press), feature discussions of differences in the L1 and L2 in order to aid the L2 teacher in diagnosing and correcting learners' errors.

Eckman's (1977) Markedness Differential Hypothesis (MDH), while bearing the name of 'markedness' (an equally important construct to be discussed below) is in fact a reformulation of the CAH that incorporates the notion of typological markedness. In this reformulation, predictions of difficulty are still based on a contrastive analysis of the L1 and the L2 but with the additional criterion that it is the level of markedness of different sounds that creates learning difficulty, not the differences in and of themselves. Therefore, unlike the CAH, which predicts that different L2 sounds will be difficult to learn, the MDH postulates that different sounds are only difficult to learn if they are typologically marked; if typologically unmarked, these sounds should not create learning difficulty. The MDH has been tested in numerous research studies, which are outlined in detail chapter 4 of this volume, to mostly positive results.

Transfer has also played a dominant role in other theories specific to L2 phonology. Several influential L2 speech perception theories, such as Flege's (1995) Speech Learning Model and Best's (1995) Perceptual Assimilation Model, are all based on the premise that the L1 shapes how the learner perceives the L2. In the latter model, it is postulated that L2 sounds are assimilated into L1 phonological categories based on similarities and that sounds in the L2 will be difficult based on their level of perceived similarity – or lack thereof – from the L1. Flege's model also posits that the acquisition process begins with L1 perceptual categories, but states that these categories may change as a result of more L2 experience. Additionally, L2 sounds may be perceived as being new, similar, or identical to L1 sounds, and the degree of similarity or dissimilarity determines whether new L2 categories can be established and/or whether equivalence classifications between the L1 and L2 sounds may be made. Both of these theories have had a major impact in L2 phonology; they are discussed in more detail throughout the volume and particularly in the chapter on L2 speech perception (Chapter 6) by Strange and Shafer.

The Ontogeny Model (OM) (1987a) and the Ontogeny Phylogeny Model (OPM) (Major 2001) also consider transfer as a major factor in L2 phonology acquisition. The OM posits that transfer is initially the major influence in L2 phonology, but that this effect decreases over time as developmental processes increase. In the OPM, a more recent version of the OM, transfer is still viewed as a dominant effect in the initial stages of acquisition; however, the effects of transfer are greater on unmarked L2 features than those that are marked. As in the OM, transfer effects decrease across time as markedness constraints increase and then decrease as the L2 is acquired.

Optimality Theory (OT), developed by Prince and Smolensky (1993, 1997, 2004) and a focus on research in L2 phonology by Hancin-Bhatt and colleagues (e.g., Hancin-Bhatt & Bhatt 1997), among others (see Chapter 5 by Hancin-Bhatt, this volume), also posits that transfer is a major factor in L2 phonology acquisition. In this constraint-based approach to L2 acquisition, learners begin the L2 learning process with their L1 constraint rankings and must learn/acquire the rankings of those constraints in the L2. Markedness also plays a major role in acquisition, as it is posited that in the process of reranking the constraints from the L1 to the L2, the least marked structures' rerankings will emerge before those that are more marked.

Another major factor in L2 phonology is universals, an important corollary of which is markedness, which had its beginnings in the work of the Prague School of Linguistics by linguists such as Roman Jakobson (1941) and Nikolai Trubetzkoy (1939). Markedness concerns universal preferences in language for certain forms or features – e.g., voiceless over voiced sounds. One approach to markedness has been the work of Greenberg (1966, 1976) on typological markedness, which focuses on the frequency of distribution of sounds across the world's languages, and implicational hierarchies (e.g., if a language has voiced stops in a coda position it will also have voiceless stops in a coda – that is, if a language has the more marked sound, it will also necessarily have the less marked sound by implication). This approach to markedness has influenced the work of Eckman in his Markedness Differential Hypothesis (MDH) (Eckman 1977), which as discussed above, incorporates the notion of typological markedness into a reformulation of the CAH by refining the hypothesis that sounds that are difficult to acquire in the L2 are difficult not simply due to being different from sounds in the L1, but by being different and more marked. Less marked sounds that are also not in the L1 would therefore be less difficult to acquire, since the level of difficulty of a sound hinges on its degree of markedness (see Chapter 4 of this volume for a more detailed discussion of the MDH).

Universals have also been important constructions in other L2 phonological theories, such as Eckman's (1991) Structural Conformity Hypothesis (SCH), which postulates that interlanguages (IL) are natural languages and governed by the universals that all natural languages are governed by. Therefore, error patterns in the learner's IL may be due to universal tendencies rather than L1 transfer or markedness.

Major's (1987a) Ontogeny Model and (2001) Ontogeny Phylogeny Model (OPM), as discussed above, also focus on the relationship between transfer and developmental processes, or universals (the OM) and markedness (the OPM); in the OM, transfer is posited to affect the early stages of acquisition, decreasing in prominence as developmental affects first increase in influence and then decrease. In the OPM, while both transfer and markedness are considered to affect L2 phonological acquisition, Major posits that those features that are unmarked in the L2 are more affected by transfer than those features that are marked. Similarly to the OM, transfer effects are posited to be more dominant in the beginning stages of acquisition, whereas markedness effects increase in dominance as transfer effects decrease, and then also decrease.

More recently, and as discussed above, OT (Prince & Smolensky 1993, 1997, 2004) has been employed in L2 phonology research (e.g., Broselow, Chen, & Wang 1998; Hancin-Bhatt 1997) to examine the effect of markedness constraints, as well as transfer, on learners' acquisition of constraint rankings in the L2. In this approach, acquisition of the L2 entails a process of reranking constraints from L1 rankings to those in the L2, with both markedness (output should be unmarked) and faithfulness (output should be faithful to the input) constraints also affecting the reranking process. It is posited that features of the L2 that are unmarked emerge before those that are more marked.

There have been other formulations of universals/markedness in L2 phonological theory, such as work on Minimal Sonority Distance, originally by Selkirk (1982) in L1 phonology and researched by Broselow and colleagues (e.g., Broselow & Finer 1991) in L2 phonology. In this theory of sonority-based markedness constraints, onsets and codas which are less marked on a sonority hierarchy are easier to acquire than those that are more marked (see Chapter 8 by Zampini for a more detailed discussion of this theory).

Another crucial concept in L2 phonology is age of acquisition, and specifically, a critical or sensitive period for SLA. The concept of a critical period for language learning was first developed by Lenneberg (1967) for L1 acquisition; he posited that language learning capabilities would start to decline at age of 2 and close at puberty. This concept has been influential in SLA, mostly in the domain of L2 phonology, with the recognition that while adult language learners may perfect their syntax and other domains of language, it is highly improbable (though possible in some extreme cases) for their L2 pronunciation to become indistinguishable from a native-speaker if L2 learning begins later in life. While questions of when the optimal period for L2 learning starts to decline and why such a period exists have not been answered, L2 researchers commonly believe that few adult L2 learners will attain the L2 pronunciation of a native-speaker.

While recent work in L2 phonology has been less concerned with native-like acquisition of the L2 and more with comprehensibility and intelligibility (cf. Cook 2002, as well as Munro, Chapter 7, and Derwing, Chapter 13, this volume), along with other factors, such as social identity (see Hansen Edwards, Chapter 9, this volume), age is still an important construct in L2 phonological theory. Transfer is an important element of the CPH – after a period of time, the L2 learner is unable to acquire new L2 forms and therefore, the forms he/she has in their repertoire will form the basis of the L2, both for production and perception. This phenomenon is addressed in some of the major theories in L2 phonology such as Flege's (1995) Speech Learning Model and Kuhl's Native Language Magnet Model (1992). Flege's model, discussed above, posits that while learners begin to perceive the L2 with L1 perceptual categories, these categories are capable of being changed towards the L2 with more experience. However, this ability to create new categories decreases with age, so that while children may be capable of developing new L2 categories perceptually, and as a result produce the L2 sounds, older learners may not have this ability. Kuhl's model of perception postulates that infants form L1 speech perceptual prototypes by age one and that as new sounds are encountered that are not in the L1, these sounds are, like a magnet, drawn towards the closest L1 speech prototypes and thus perceived as this prototype. The result is that non-L1 sounds are perceived in light of the L1 – e.g., in any later L2 learning, the L2 sounds are in effect transferred towards L1 perceptual categories.

Sociolinguistic and sociocultural theory has had a major impact on work in L2 phonology as well, as researchers attempt to understand how the social context and socio-biological constructs affect language learning. However, it is within the area of variation that sociolinguistics has had the greatest effect on L2 phonology. Variation is a critical issue for SLA, since any valid model of L2 acquisition must address questions of how variation in production is to be explained, and whether it is a feature only of a given task or speech style or type of interlocutor (and therefore a by-production of production), or a feature of competence (and indeed, whether there is a difference between production and competence). Several major approaches have been utilized in the study of these issues: social network theory (Milroy & Milroy 1992) (cf. Lybeck 2002); Speech Accommodation Theory (SAT), developed by Giles and colleagues (cf. Giles & Powesland 1975) and researched in L2 phonology by Beebe and colleagues (cf. Beebe 1977; Beebe & Zuengler 1985); Tarone's (1979) Capability Continuum, based on Labov's (1969, 1972) Observer's Paradox, and variable rule analysis of social and linguistic factors and their effect on the production of a given feature (cf. Preston 1996). These approaches are discussed in more detail in Chapter 9 of this volume.

In summary, the research and teaching of an L2 phonology, while shifting with and at times being at the forefront of new movements and trends in linguistics and SLA, can be connected across time and space by a number of key constructs that have shaped and continue to shape L2 phonology research and pedagogy: transfer, universals/markedness, the critical period hypothesis, and variation. These concepts are, not surprisingly, central themes in the chapters of this book. Although the chapters cover a wide array of topics, they are nevertheless unified in focusing on the most critical aspects of L2 phonology. An overview of the organization and discussion of these aspects is given below.

Organization of the volume

The book is divided into three parts, with each section unified by broad thematic content. Each chapter examines theoretical frameworks, major research findings (both classic and recent), methodological issues and choices for conducting research in a particular area of L2 phonology, and major implications of the research findings for more general models of language acquisition and/or pedagogy. Part I, "Theoretical Issues and Frameworks in L2 Phonology," lays the groundwork for examining L2 phonological acquisition. First, Ohala (Chapter 1) provides an overview of L1 phonological acquisition, a topic not often discussed in books on L2 phonology. It is only through an understanding of L1 acquisition, however, that one can begin to determine ways in which L1 and L2 acquisition may be similar and different. In addition, any truly adequate model of linguistic competence must be able to account for the ways in which we acquire, process, use, and internalize language, regardless of whether it be L1, L2 or L*x*. The next three chapters focus on crucial theoretical issues in L2 phonology that have had a major impact in the field: the role of age in L2 phonological acquisition (Ioup, Chapter 2), native language transfer (Major, Chapter 3), and typological markedness (Eckman, Chapter 4). The notion of markedness is central to several more general models of phonological theory for L1 as well. In this way, Eckman's chapter serves as a lead-in to the final chapter of this section, which explores the relationship between L2 phonology research and Optimality Theory, one of the more recent and influential models of L1 phonological competence (Hancin-Bhatt, Chapter 5). Thus, the final chapter returns to a consideration of the L1, but in this case, considers the ways in which L2 data can be used to inform more general models of phonological competence.

Part II, "Second Language Speech Perception and Production," examines these two aspects of L2 speech in more detail. The first two chapters examine perception from different perspectives. First, Strange and Shafer (Chapter 6) examine research on L2 speech perception – that is, the ways in which ***learners*** perceive the sounds of their L2. Munro (Chapter 7), on the other hand, focuses on issues of foreign accent and intelligibility – that is, the ways in which ***listeners*** perceive L2 speech, whether those listeners be L1 speakers of the language in question or other L2 learners. The next two chapters in Part II focus on L2 speech production. Zampini (Chapter 8) provides an overview of research on the production of individual speech sounds in the L2, focusing primarily on the characteristics of L2 speech and ways in which they differ from the speech of native speakers. Finally, Hansen Edwards (Chapter 9) examines social factors that contribute to the ways in which learners produce L2 speech – both consciously and subconsciously – along with variation in L2 speech production.

Finally, Part III, "Technology, Training, and Curriculum," bridges the gap between theory and practice. Bradlow (Chapter 10) begins with a consideration of general issues, problems, and findings related to training L2 learners to more accurately perceive and/or produce L2 speech sounds. Gick, Bernhardt, Bacsfalvi and Wilson (Chapter 11) then examine ultrasound imaging as a tool for examining the articulation of L2 speech sounds, as well as an aid for training L2 learners to manipulate the movement of the articulators in an attempt to produce more native-like speech. Thus, this chapter provides an example of how improved access to technology and advanced imaging techniques have expanded opportunities for studying speech in promising new ways. Chun, Hardison and Pennington (Chapter 12) shift the focus of study from individual speech sounds to L2 prosody, including discourse intonation, in research and training. Finally, Derwing (Chapter 13) uses major research findings in L2 phonology to examine issues of curriculum and materials development for the L2 pronunciation classroom.

While each chapter in this volume examines a particular aspect of L2 phonological acquisition in detail, it is impossible to treat each topic as an isolated and autonomous feature of L2 phonology. Thus, the reader will find that the same topics or studies may be mentioned and discussed in more than one chapter, and links are made between those chapters that complement, contrast, or have implications for another. This is particularly true of the chapters in Part I that address particular theoretical issues of L2 phonology. For example, while Ioup (Chapter 2) focuses on the role and effects of age in L2 acquisition, along with their theoretical and practical implications, Zampini (Chapter 8) also surveys a number of studies of L2 speech production that either indirectly or directly consider age of acquisition in the research design and the discussion of research findings. Strange and Shafer, in their chapter on L2 speech perception (Chapter 6), and Munro, in his chapter on foreign accent and intelligibility (Chapter 7), also consider age effects. Similarly, L1 transfer can have a profound effect on L2 speech, and while Major (Chapter 3) dedicates an entire chapter to this issue, several other chapters also address the role of L1 transfer as it relates to research findings in a particular domain or area of focus, including, for example, Eckman (Chapter 4), Hancin-Bhatt (Chapter 5), Strange and Shafer (Chapter 6), Zampini (Chapter 8), and Hansen Edwards (Chapter 9). The chapters in Part III also draw on research findings in related domains of L2 phonology. For example, many training approaches are based on or respond to more general findings in the literature, and Bradlow (Chapter 10) surveys some of these, especially with respect to L2 speech perception (cf. Strange & Shafer, Chapter 6). In addition, Derwing (Chapter 13) relates research findings such as those discussed by Munro (Chapter 7) and Chun, Hardison and Pennington (Chapter 12) to issues of curriculum design.

A key feature of this book is its attempt to provide a comprehensive overview of the field of L2 phonology. At the same time, however, an attempt has been made to keep the volume readable and of reasonable length. As a result, some significant topics do not receive chapter-length status, and it may appear at first glance that some critical issues have been omitted. However, in those cases where important issues have not received chapter-length treatment, they have nevertheless received attention within one or more chapters. For example, Gick et al. (Chapter 11) reports on the use of ultrasound imaging as an emerging technology for speech research, due to its relatively recent portability and reduced cost. Yet, it is not the only technology that is changing the ways in which researchers examine L2 speech, and Strange and Shafer (Chapter 6) and Zampini (Chapter 8) address other imaging techniques for studying L2 speech, such as fMRI. In a similar vein, Hancin-Bhatt (Chapter 5) discusses the relationship between L2 phonology research and Optimality Theory, which is only one of several theories/models of L1 phonological competence. However, Zampini (Chapter 8) discusses studies that have employed generative models of phonology for examining L2 speech, including metrical phonology for the acquisition of L2 stress and feature geometry and distinctive feature theory for studies of L2 sound substitutions. In addition, Major (Chapter 3) and Eckman (Chapter 4) discuss aspects of Universal Grammar as they relate to L2 phonology. Important theoretical frameworks in L2 phonology also receive treatment in a number of individual chapters, even though they may not appear as the focus of an entire chapter. For example, Flege's (1995) Speech Learning Model has been extremely influential in L2 phonology and is discussed in a number of chapters, including Ioup (Chapter 2), Strange & Shafer (Chapter 6), Munro (Chapter 7), Zampini (Chapter 8), and Bradlow (Chapter 10). Ioup, Strange & Shafer, and Bradlow also address important theories in L2 perception and the perception/production interface, including not only the Speech Learning Model, but also the Perceptual Assimilation Model (Best 1995) and the Native Language Magnet Theory (Kuhl 1992). Thus, it is hoped that this book will provide the reader with a sense of the breadth of the field of L2 phonology, as well as an in-depth discussion of the most critical issues and fruitful research findings in recent years.

Each chapter in the volume has a similar organizational format, although the overall focus toward theory or practice may vary. First, the specific domain of L2 phonology to be detailed in the chapter is introduced, and the authors provide an overview of the overriding research and/or pedagogical questions. Through a survey of the literature, the authors then present some of the most important research findings; while they generally focus on more recent findings, other classic and fundamental findings are also reviewed, as necessary. The chapters also contain an examination of the methodological choices and options for conducting research in a particular area; while there are obvious similarities across topics, important differences and trends for conducting research on a particular aspect of L2 phonology also emerge. In addition, the authors discuss outstanding problems and concerns for conducting research in their area, and they consider ways in which the research findings contribute to researchers' understanding of the acquisition of an L2 sound system, as well as related disciplines. Finally, each chapter concludes with a brief exposition of what researchers still do not know in a particular area, as well as what future directions a particular area may move in, given emerging trends in research and/or teaching and promising avenues of investigation.

References

Beebe, L. M. (1977). The influence of the listening on code-switching. *Language Learning* 27: 331–339.

Beebe, L. M., & Zuengler, J. (1985). Accommodation theory: An explanation for style shifting in second language dialects. In N. Wolfson and E. Judd (Eds.), *Sociolinguistics and language acquisition.* (pp. 195–213). Rowley, Mass.: Newbury House Publishers, Inc.

Best, C. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In C. Goodman & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA: MIT Press.

Best, C.T. 1995. A direct realist view of cross-language speech perception. In W. Strange (ed.) *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.

Broselow, E., Chen, S-I., & Wang, C. (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition* 20: 261–280.

Broselow, E., & Finer, D. 1991. Parameter setting in second language phonology and syntax. *Second Language Research* 7: 35–59.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Cook, V. (2002). *Portraits of the L2 user.* Clevedon: Multilingual Matters.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning* 27: 315–330.

Eckman, F. 1991. The Structural Conformity Hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition* 13: 23–41.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-linguistic research* (pp. 233–277). Timonium, MD: York Press.

Fries, C. C. (1945). *Teaching and learning English as a foreign language.* Ann Arbor: University of Michigan Press.

Giles, H., & Powesland, P. F. (1975). *Speech style and social evaluation.* London: Academic Press.

Greenberg, J. (1966). *Language universals.* The Hague, The Netherlands: Mouton.

Greenberg, J. (1978). Some generalizations concerning initial and final consonant clusters. In J. Greenberg, C. Ferguson, & E. Moravcsik (Eds.), *Universals of human language* (Vol. 2, pp. 243–279). Stanford, CA: Stanford University PressHawkins 1984.

Hancin-Bhatt, B. & Bhatt, R. M. (1997). Optimal L2 syllables: Interactions of transfer and developmental effects. *Studies in Second Language Acquisition* 19: 331–378.

Jakobson, R. (1941). *Kindersprache, aphasie, und allgemeine lautgestze*. Translated 1968 as *Child language, aphasia, and phonological universals* (A. R. Keiler, Trans.). The Hague: Mouton.

Kuhl, P. (1992). Psychoacoustics and speech perception: internal standards, perceptual anchors, and prototypes. In L. Werner & E. Rubel (Eds.), *Developmental psychoacoustics* (pp. 293–332). Washington, DC: American Psychological Association.

Labov, W. (1969). The study of language in its social context. *Studium Generale* 23: 30–87.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.

Lenneberg, E. (1967). *Biological foundations of language*. New York: John Wiley & Sons.

Lybeck, K. (2002). Cultural identification and second language pronunciation of Americans in Norway. *The Modern Language Journal* 86: 174–191.

Major, R. C. (1987a). A model for interlanguage phonology. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 101–125). New York: Newbury House/Harper & Row.

Major, R. C. (2001). Foreign accent: The ontogeny and phylogeny of second language phonology. Mahwah, NJ: Lawrence Erlbaum Associates.

Milroy, L., & Milroy, J. (1992). Social network and social class: Toward an integrated sociolinguistic model. *Language in Society* 21: 1–26.

Preston, D. (1996). Variationist perspectives on second language acquisition. In R. Bayley and R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 1–45). Amsterdam: John Benjamins Publishing Company.

Prince, A., & Smolensky, P. (1993). Optimality Theory: Constraint interaction in generative grammar. Unpublished manuscript, Rutgers University, New Brunswick, NJ and University of Colorado, Boulder, and Rutgers University Optimality Archive.

Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to Universal Grammar. *Science* 275: 1604–1610.

Prince, A. & Smolensky, P. (2004). Optimality Theory: Constraint Interaction in Generative Grammar. Malden, MA/Oxford, England: Blackwell.

Selkirk, E. (1982). The syllable. In van der Hulst & Smith (Eds.), *The Structure of phonological representations, Part II*. Dordrecht: Foris.

Swan, M., & Smith, B. (2001). *Learner English: A teacher's guide to interference and other problems.* Cambridge: Cambridge University Press.

Tarone, E. E. (1979). Interlanguage as chameleon. *Language Learning* 29: 181–191.

Trubetzkoy, N. 1939. *Principles of phonology.* Paris: Klincksieck.

Weinreich, U. (1953). *Languages in contact.* New York: Linguistic Circle of New York.

PART I

Theoretical issues and frameworks in L2 phonology

Preface

Part I provides a basis and context for investigating the acquisition of a second language (L2) sound system by addressing some of the key factors that affect L2 phonological acquisition, as well as major theoretical issues that any adequate model for L2 phonology must take into account. First, Chapter 1 examines the primary mechanisms in first language (L1) phonological acquisition and discusses the major research findings in this area. An understanding of L1 acquisition is essential in order to more fully understand L2 acquisition, and Chapter 1, therefore, lays the groundwork for examining ways in which L1 and L2 phonological acquisition may be similar and ways in which they may differ. The next three chapters survey in depth three important issues related to L2 phonological acquisition: age, native language transfer, and typological markedness (Chapters 2, 3, and 4, respectively). Each of these topics has received extensive treatment in the literature, and as will be seen in subsequent chapters, all have implications for L2 phonology research in a variety of domains. Regardless of the particular aspect of the L2 sound system that a researcher may examine, transfer, markedness, and/or age are often implicated in an analysis and discussion of the research findings. Finally, the last chapter of this section (Chapter 5) demonstrates how some the theoretical issues discussed in earlier chapters (in particular, transfer and universal markedness) may be examined within a current framework for the study of phonology in general – Optimality Theory. Part I thus concludes by showing how research in L2 phonology may inform more general models of phonological theory and vice versa.

In Chapter 1 ("Phonological acquisition in a first language"), Diane K. Ohala addresses how infants begin to develop perceptual abilities, first for language sounds in general, and then for the ambient language specifically. She also outlines how productive processes develop in various stages. In her discussion of each of these areas, Ohala critically outlines the most commonly employed research methodologies and discusses the major areas of research in perception and production, as well as the major findings in each of these areas. Finally, Ohala draws comparisons between the findings for L1 phonological acquisition to those for acquirers of an L2 phonology. In particular, she argues that, although some aspects of phonological acquisition are specific to infants and children acquiring their first language, there are nevertheless similarities in the L1 and L2 acquisition of speech sounds that merit further investigation.

In Chapter 2 ("Exploring the role of age in the acquisition of a second language phonology"), Georgette Ioup addresses the complicated issue of whether age is a factor in L2 acquisition. She first reviews research that indicates that many adult learners of an L2 have an identifiable accent, usually due to their L1, and that production may be affected by imperfect L2 perception. She also examines research that compares child and adult L2 learners to determine whether there are differences between early-onset L2 speech (i.e., child L2 acquisition) and late-onset L2 speech (e.g., those who acquire L2 after the age of 16). While stating that differences have been found and that these findings are fairly robust, Ioup nevertheless argues that methodological weaknesses may have influenced the outcomes of some studies. She then evaluates different models and proposals to account for age-related differences in L2 phonological acquisition, including the Critical Period Hypothesis, the Perceptual Assimilation Model, the Native Language Magnet Model, and the Speech Learning Model. While the Critical Period Hypothesis provides a biological explanation for age-related differences, the remaining models propose that interference from the L1 contributes most significantly to the difficulty that adults face in trying to master an L2 phonology. Ioup also discusses other factors (e.g., amount of L1/L2 use, attitude, length of residence, etc.) that may lead to not only age-related, but also individual, differences in L2 phonological acquisition. Finally, she concludes the chapter by putting forth a number of directions for future research.

In Chapter 3 ("Transfer in second language phonology"), Roy C. Major surveys in more detail research on the role of L1 interference, or transfer, in L2 phonological acquisition. He begins with a historical perspective on research involving transfer in learning more generally and continues with a detailed discussion of L1 transfer in L2 speech specifically. This discussion spans the work of several decades, beginning with the Contrastive Analysis Hypothesis and earlier, and culminating with current approaches to L2 phonological research, such as Optimality Theory. Within this setting, Major provides an overview of different conceptualizations on the conditions for transfer; discusses levels of analysis and the distinction between abstract and surface transfer; and addresses the issue of 'similarity' in transfer – that is, how similarity has been defined and investigated, as well as how similarity impacts transfer. He then outlines a number of theoretical approaches that focus on the interaction between transfer and other linguistic phenomena, such as markedness. Major concludes by providing directions for future research in terms of both methodological choices and areas of study.

The role of markedness in L2 phonology in considered in greater detail in Chapter 4 ("Typological markedness and second language phonology") by Fred R. Eckman. Eckman first provides an overview of different conceptualizations of markedness and discusses two major theoretical frameworks in L2 phonology that incorporate markedness as a major element: the Markedness Differential Hypothesis and the Structural Conformity Hypothesis. In discussing each of these, Eckman first outlines each approach and then presents evidence both for and against each framework. Within this context, he also addresses the question of whether or not typological universals constitute viable explanatory principles for the analysis of L2 speech data. Eckman concludes the chapter by discussing future directions in the conceptualization of and research on the role of markedness in L2 phonology. He places particular emphasis on the role that markedness principles play in Optimality Theory, a theoretical model discussed in the next chapter by Barbara Hancin-Bhatt.

In Chapter 5 ("Second language phonology in Optimality Theory"), Barbara Hancin-Bhatt explores Optimality Theory (OT) as a promising theoretical framework for the analysis and explanation of L2 speech data and phonological phenomena. Hancin-Bhatt first outlines the issues that any adequate theoretical model of L2 phonology must address and provides a general overview of OT, including its underlying assumptions, basic principles, and mechanism for evaluating the relationship between input and output forms. She then details how OT may account for L2 phonological phenomena by discussing L2 research that employs an OT framework. In her discussion, Hancin-Bhatt highlights the role of transfer and markedness, and she makes use of the data from the studies cited to illustrate how OT can account for both effects in L2 phonological acquisition. She also discusses how learning algorithms formulated within an OT framework may be used to examine and make predictions about different stages of acquisition. In addition, she argues that linguistic variation is not incompatible with OT, and that the theory may therefore prove to be a promising tool for the description and explanation of the variability found in L2 speech. Hancin-Bhatt concludes by discussing the theoretical implications of OT for L2 acquisition and by providing concrete suggestions for further research.

CHAPTER 1

Phonological acquisition in a first language[1]

Diane K. Ohala
University of Arizona

Introduction

Our first experience with language arguably comes in the womb, where from that safe haven we are exposed to the (albeit muted) sounds of the language or languages being spoken around us. From the moment of our birth, those sounds suddenly become louder and more distinct, and we are awash in a confusing cacophony of myriad voices – all saying things that we must ultimately come to understand and be able to produce ourselves. At the same time, we must also be able to differentiate speech sounds from all the other sounds and disturbances that take place in the world around us. As daunting as this may seem, the tasks involved in learning to perceive and produce the sounds of a first language are accomplished much more rapidly than the overall work involved in learning a language in its entirety (where the latter takes roughly twice as long). By the age of three most children learning one language can perceive and produce the majority of the sounds of that language, with a handful of more troublesome sounds (like English [ɹ] and [ð]) perhaps taking three or four years more.

One likely reason for such quick success is that the work involved in phonological acquisition does not wait to begin with the ability to produce speech. In fact, much is accomplished before that point, such that the task of speech production is best considered the *second* of two major tasks that a child must achieve in this domain. The first is to segment the speech stream into its component parts. Without an entrée into the complicated and overlapping sequences of sounds that comprise continuous speech, it would be impossible for a child to recognize words, let alone begin to comprehend their meaning. Fortunately, although our newborn

1. Many thanks are due to the editors of this volume. This chapter has also benefited from comments from two anonymous reviewers and the support and advice of Mike Hammond and LouAnn Gerken. Thanks also to Erin Good and Ernie Ohala-Hammond, whose empathy was much appreciated. All errors are my own.

vocal anatomy may not be capable of anything more sophisticated than gurgles and cries, our hearing is actually quite good and allows us to begin the task of unraveling the speech stream long before we can articulate the sounds we hear (Kuhl 1987).

As we shall shortly see, pre-linguistic babies are highly adept at discriminating among speech sounds and at many other perceptual tasks relevant to phonological learning. Of course, it is fair to ask how it is possible to even know what knowledge babies possess (linguistic or otherwise) before they can actually tell you themselves. In this regard, researchers have developed several rather clever methods for investigating what babies have learned before they begin producing speech. As these methods are complex, it is worth taking a moment to discuss them.

Infant speech perception

Methods in the investigation of infant speech perception

Several, now standard, techniques have been devised by researchers as a means of exploring what infants may learn before speech production begins. Two of the most widely-used methods to accomplish this are the High Amplitude Sucking (HAS) technique (Eimas, Siqueland, Jusczyk, & Vigorito 1971) and the Head-Turn Preference Procedure (HTPP; Kemler Nelson, Jusczyk, Mandel, Myers, Turk, & Gerken 1995). The main question addressed by both methods is to what extent babies are aware of changes in aural speech stimuli. Any demonstrated awareness (usually stated as a preference for one sound or set of sounds over another) is taken to reflect a baby's ability to perceive a difference among the relevant speech sound stimuli and, thus, to reflect what a baby knows about his/her language.

The HAS Technique

The HAS technique is used with very young babies – sometimes only hours or days old – and takes advantage of the fact that infants (a) like to hear sounds and (b) will readily suck on a pacifier. In this method, infants are given a pacifier that is connected to a pressure transducer, a machine that measures the sucking rate of the infant. When the infant sucks hard enough on the pacifier, a sound is played. This, of course, is a surprise to the infant, who typically sucks harder in order to hear the sound again and again. The rate at which the sound is played directly corresponds to the infant's sucking rate, which declines as he/she loses interest. When the baby ceases sucking on the pacifier altogether or the sucking rate falls to a predetermined level of disinterest (called habituation), a new sound is played. Upon hearing the new sound, an infant usually will begin to suck harder in order to hear the new sound repeated over and over (called dishabituation). The sucking rate of infants exposed to the new sound after habituation is then compared to the sucking rate of infants who heard the same sound again. If there is a statistically significant difference in sucking rates between the two groups (one group shows dishabituation and the other does not), then researchers conclude that the infants who heard the new sound must have been able to perceive a difference between it and the original sound.

Thus, the HAS technique allows us to discover, among other things, whether infants can perceive differences among the sounds of the ambient language or, indeed, among the sounds of any language. Such abilities are arguably directly related to what an infant must do to begin to understand the speech stream. However, as lucky as we are to have such an ingenious research method at our disposal, there are several cautions in using HAS. Studies have shown that the technique is not as successful with babies older than four months because they tend to fidget more and will not hold the pacifier (Kuhl 1987). Also, if an infant shows no dishabituation after a new sound is played, it is difficult to know whether the lack of increased sucking is due to an actual inability to discriminate a difference among sounds or to any of a number of possible causes of disinterest, such as infant discomfort, distress, or tiredness (Hoff 2001).

The HTPP

For babies older than four months, for whom the HAS technique is no longer viable, the HTPP is used. This technique cannot be used with newborns or younger babies because it requires them to have sufficient muscular control over their head and neck and is therefore most successful with babies older than four months. Like HAS, this method takes advantage of a baby's interest in hearing sounds, but also capitalizes on the fact that children will naturally *look* in the direction of a sound source when one is presented – and will continue looking if motivated to do so by some stimulus. For this technique, the baby sits facing forward on the parent's or caretaker's lap in a soundproof booth. The booth has a one-way viewing window to allow experimenter observation and recording from outside the booth. The parent or caregiver wears headphones through which masking music is played, so they cannot inadvertently cue their child. Directly in front of the baby is a yellow light; at the start of an experiment, this light begins flashing in order to orient the baby's gaze in that direction. Once the baby's gaze is captured at center for a sufficient amount of time, that light is extinguished and another light begins flashing either to the baby's left or right (which light begins flashing first is randomized). At that point, a sound or sequence of sounds is played and, whereas the flashing light is extinguished, sound continues to play for a predetermined amount of time. This is referred to as the familiarization or training phase, during which babies just listen to one set of speech sound stimuli and are trained in the stimulus-response behavior required for the experiment.

After training is complete, the baby's gaze is re-oriented to the center light. At this point, one of the side-lights begins to flash and either the same or a new sound is played. This is referred to as the test phase, where the new and old stimuli are presented in random order and are also randomly associated with one of the side-lights. Stimuli are played as long as the baby's gaze continues to focus in the relevant direction and ceases when the baby looks away for too long, at which point the baby is re-oriented to the center light and the process begins again. The amount of time the baby listens to each stimulus (as measured in looking-time) is recorded. If there is a significant difference in the looking times for one stimulus versus another, then researchers conclude that the baby was able to perceive a difference between the two sounds or sequences of sounds.

As with HAS, the HTPP allows us to investigate which sounds or series of sounds babies are able to discriminate. It also allows us to ask more sophisticated questions, such as whether older babies can learn patterns in a language-like sequence of sounds (for example, patterns containing phonotactic information) and apply their knowledge of those patterns to new sequences. This type of ability is arguably one which infants must acquire in order to begin understanding the phonological patterns present in the ambient language. However, despite the increase in the body of questions that this method allows us to ask, there are several cautions for its use. Older babies (16–18 months) tend to get restless and may disrupt the experiment by getting up and wandering around during testing. In addition, as with HAS, the absence of any discrimination is difficult to decipher because other, non-linguistic factors may focus or draw away a baby's attention. For this reason, for both methods, a number of babies are usually tested in order to obtain reliable results.

In sum, although there are several other methods available for testing infants and toddlers before they can talk, the HAS and HTTP techniques are most commonly used to answer questions regarding infant speech perception (Hoff 2001). An overview of results from studies using these techniques follows, focusing on those that have had a significant impact on the field of infant speech perception.

Findings from studies in infant speech perception

Discriminating among languages and speakers

Since the late 1950s, researchers have sought to determine the extent of infants' knowledge of their language prior to the production of speech. Decades of research have shown that infants demonstrate remarkable perceptual acuity. For example, DeCasper and Fifer (1980) showed that newborns (less than 24 hours old) preferred to listen to their mother's voice over another, unfamiliar female voice – suggesting that recognition of the defining characteristics of a parent's voice begins in utero. In a follow-up study, DeCasper and Spence (1986) showed that 3-day-old infants whose mothers had read passages to them prior to birth preferred to listen to those familiar passages versus new, unfamiliar passages when tested – suggesting that babies begin attending to some aspects of speech, such as prosody, while still in the womb.

Other studies have shown that this sensitivity to prosody extends to the ability to discriminate the native language from another language. Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, and Amiel-Tison (1988) played tapes of French and Russian speech to four-day-old babies whose mothers spoke French. Results showed that infants preferred listening to utterances spoken in their native language to those spoken in Russian. Crucially, babies exposed to the same stimuli whose parents spoke neither French nor Russian did not show any preference for one language over the other. These results show that infants are not only able to hone in on a mother's voice versus any other, similar voice, but that they are able to discriminate utterances in the mother's *language* from utterances in another language. This ability to recognize characteristic prosodic patterns in the native language is one of several, critical first-steps in a baby's journey to phonological competence. In fact, studies suggest that it is not until 9 months that babies begin to attend to more than just prosodic cues in the speech stream (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk 1993; Morgan & Demuth 1996).

Discriminating among native and non-native speech sounds

In addition to asking whether infants could distinguish among languages and speakers, researchers were asking whether infants could discriminate among *individual* speech sounds, both in the ambient language and in unfamiliar languages. One of the earliest of studies (Trehub 1973) showed that four-week-old infants could, in fact, distinguish the vowel [ɪ] from [a], and the vowel [u] from [ɪ]. Other studies have shown that infants are equally adept at discriminating consonants that differ on some dimension (such as place of articulation or voicing). For example, in a landmark study, Eimas, Siqueland, Jusczyk, and Vigorito (1971) tested babies' ability to discriminate among different tokens of the syllables *pa* and *ba*. These two syllables differ in only one phonetic contrast – the amount of time (called voice-onset-time or VOT) that elapses before the vocal folds begin to vibrate after the lips open for the initial sounds *p* and *b*; in all other respects, the two syllables are identical. Babies in the study were played sounds on a continuum from *pa* to *ba*, with some tokens beginning with a more /p/- like sound and others with a more /b/-like sound due to the manipulation of VOT. They found that babies not only detect a difference between *pa* and *ba*, but that the manner in which they do so is adult-like and categorical. That is, like adults, babies perceive different tokens of /p/ and /b/ as either one or the other and never as something in between (this is referred to as categorical perception). Because such fine phonetic contrasts apply almost exclusively to speech, these findings generated a great amount of debate over whether infants were born with an innate perceptual capacity that was specifically honed for language. Later studies showed that this was not the case given that stimuli on a non-speech continuum (such as a noise-buzz continuum) can also be perceived categorically (Miller, Wier, Pastore, Kelley & Dooling 1976) and that other animals (such as chinchillas) whose auditory mechanisms are much like humans also demonstrate categorical perception (Kuhl & Miller 1975). Regardless, a large body of research has made clear that infants can initially perceive sounds that contrast at almost every phonetic level (such as voicing and place and manner of articulation), whether those contrasts occur in their native language or not (for a summary see, e.g., Kuhl 1976; Jusczyk 1997; Werker & Pegg 1992).

The effect of the native language

What seems equally clear by now is that this ability is adversely affected the more native-language experience a baby obtains. For example, Trehub (1976) tested Canadian infants and adults on their ability to discriminate sounds in Czech. As with previously-mentioned studies, the test-sounds were phonetically similar and differed on only one phonetic dimension. Results showed that the infants in the study were able to perceive the distinction between the two Czech sounds, whereas the adults were not able to do so. This latter fact seems to indicate that the ability to make finer perceptual distinctions among nonnative speech sounds is weakened – at some point – by exposure to the native language.

Werker and Tees (1984) explored just when in development this native-language interference begins to appear. They tested English-speaking babies, aged 6, 8, 10, and 12 months, on their ability to discriminate nonnative sounds in Hindi and a Salish language (Nthlakapmx). Results showed that 6- and 8-month-old babies could discriminate the nonnative speech sounds but 10- and 12-month-old babies could not. On the face of it, these results – by demonstrating that children appear to lose perceptual acuity very early in life – challenged the popular notion that it is not until puberty that loss of language-learning skills is observed (see Ioup, this volume, for a discussion of the Critical Period Hypothesis as well as Penfield & Roberts 1959; Lenneberg 1967). In addition, in the same study, Werker and Tees showed that this weakening in perceptual acuity is not due to any physical deterioration of the human auditory mechanism but rather is a function of how listeners adjust their phonetic categories to "tune out" those contrasts that are not relevant to the native language. To further support this claim, they showed that when the experimental procedure was adjusted by, for example, providing longer intervals between stimulus presentation, adults could, in fact, discriminate nonnative contrasts.

Numerous studies since then have supported these findings, in particular that with enough training, adults can "re-learn" contrasts not present in the native language (e.g., Logan, Lively, & Pisoni 1991; MacKain, Best, & Strange 1981; Maye

2000; Tees & Werker 1984). However, a few studies show that some nonnative contrasts appear to be easier to discriminate than others. Best, McRoberts, and Sithole (1988) tested English-speaking adults and 6-to14-month-old English-learning babies on their ability to discriminate a nonnative contrast among Zulu clicks. They found that the adults and all the babies, no matter the age, could distinguish the sounds in question. They and others (e.g., Werker & Pegg 1992) have suggested that one possible reason for this lies in the fact that Zulu clicks are highly dissimilar from any English sounds, whereas sounds tested in other studies more closely resembled, but were not identical to, native sounds. In the latter case, the argument is that the similarity of tested nonnative contrasts to sounds present in the native language causes those sounds to assimilate to phonetic categories already available in that language. In the case of Zulu clicks, this is not possible as there is no native-language category similar enough. This lack causes the Zulu sounds to "stand out" in a way that too-similar contrasts do not – making the discrimination task an easier one.

These findings led to a much larger claim that the onset of developmental changes in infant speech perception, and the way in which the perception of nonnative contrasts are differentially affected, coincides with the beginning of phonological acquisition proper. That is, it is arguably at this stage in development that infants set up speech sound categories that might provide the foundation for a phonological system as opposed to solely a perceptual one. This claim, as well as the findings from which it arose, were met with numerous counter-proposals ranging from acoustic and articulatory to cognitive (see Werker & Pegg 1992, for a review; also Maye, Werker, & Gerken 2002 for a more recent proposal) and the issue of whether such findings are evidence for the emergence of a phonological system or a result of fine-tuning in some other domain is still highly debated.

Learning a phonological system

Tangential to this debate is still the question of exactly *how* an infant might begin to acquire a phonological system, whether or not we assume such a system to be present or even possible prior to speech production. Recent research suggests that there are at least two pathways to information helpful in phonological learning: distributional learning and rule learning. In a demonstration of distributional or statistical learning, Saffran, Aslin, and Newport (1996) played 8-month-old infants sequences of four nonsense words, each composed of three syllables, such as *bidaku*, *golabu*, *podati*, and *tupiro*. The stimuli were played in random order for two minutes, but the transitional probabilities among the syllables composing each word were always maintained, such that certain syllables followed or preceded others with greater regularity. For example, babies heard sequences like *tupirogolabubidakupadotipadotitupiro* where *pi* was always followed by *ro* but never by *go* or *ti*. In order to discover whether babies had attended to the transitional probabilities among the syllables presented at training (and thus learned the nonsense words), they then played the same babies new sequences of the same syllables that either followed the distributional properties modeled in the training sequences or followed some novel pattern. They found that babies listened longer to the sequences that followed the novel pattern, suggesting that the babies could detect a difference between the two types of sequences. The implication is that babies can learn (with a minimal amount of exposure) distributional patterns in the input language. This result was of significant importance as it addressed just how an infant might begin to learn word boundaries and other relevant information, like language-specific phonotactics.

As remarkable as such statistical skills are in infants, there is evidence that these abilities are even more sophisticated. Gomez and Gerken (1999) showed that infants are able to further apply knowledge gained from such input to the point that they are able to extract an abstract pattern present in language-like stimuli *and apply it to new sequences that follow the same abstract pattern*. Specifically, they played sequences of grammatical strings that followed from an artificial, finite-state grammar to 1-year-olds. Like the Saffran et al. (1996) study, stimuli consisted of a series of nonsense words (such as *VOT-PEL-JIC* or *VOT-PEL-PEL-JIC-RUD-TAM*), but the rules governing which syllables could follow or precede another were more complex (e.g., *PEL* could be followed by itself, *TAM,* or *JIC* but never by *VOT* or *RUD*). After babies were trained on one grammar, they then tested babies' ability to discriminate between two *new* grammars. The new grammars both had different "lexical" items, but one followed the same grammatical rules as the training grammar and the other violated those rules. They found that babies listened longer to utterances consistent with the rules of the training grammar, despite the fact that both grammars used all new vocabulary. This suggests that infants could not only hear a difference between the two sets of stimuli, but that they were able to *abstract the grammatical rules* present in the training grammar, recognizing a similarity in one of the new grammars but not in the other. This finding – that babies are able to learn abstract patterns or rules – was further supported in subsequent work by Marcus, Vijayan, Bandi Rao and Vishton (1999) and suggests another way in which babies might begin to learn complex phonological patterns in language, such as vowel harmony, allophony, or allomorphy. In fact, both distributional and rule learning are powerful skills that babies are able to bring to bear on the complex task of language segmentation.

Summary of infant speech perception

The first task a child must face in phonological acquisition, and, indeed, in language acquisition in general, is no simple matter. However, as we have seen, there is a significant body of evidence that the foundations for achieving speech segmentation are laid very early in life. Even before birth, infants begin to attend to salient prosodic patterns in the ambient language; this sensitivity continues to sharpen and expand after birth, such that the perception of almost any phonetic contrast maintained in languages is initially possible – by infants as young as four weeks old for some contrasts. At around 10–12 months continued exposure to the native language begins to have an effect. At this time, an infant's perceptual organization of speech sounds appears to focus almost exclusively on those phonetic contrasts that occur in the native language. Simultaneously, the ability to distinguish among other, irrelevant or highly similar nonnative contrasts is lost, at least temporarily. In addition, children have been shown to learn a great amount of distributional information present in the native language in very little time and appear further able to extract abstract patterns or rules from that distributional information. All of these skills provide numerous paths by which infants can begin the complex task of language acquisition in general, and of phonological acquisition in particular.

Comparisons to second language acquisition

Although many of the speech-related tasks that babies must initially face are specific to first-language (L1) learning only, others are clearly (or at least arguably) relevant to second acquisition (L2) as well. Segmenting the speech stream is certainly a task that second-language learners must also conquer, but with, perhaps, more comprehensive baggage: namely, depending on the age of acquisition of the L2, significant interference from L1 phonology. This latter circumstance is well documented (beginning with Weinreich 1953), although given that weakening effects of the native language are seen as early as 10–12 months, it is curious why children are nevertheless better at learning languages than adults (for some thoughts on this matter, see, e.g., Sebastián-Gallés & Soto-Faraco 1999; for a discussion of age and L2 acquisition, see Ioup, this volume). Finally, studies showing that native-language effects on a learner's perceptual organization of phonetic contrasts can be improved with training may be of particular relevance to L2 pedagogy. In this line, recent studies seem to indicate that the way in which distributional information regarding nonnative phonetic contrasts is presented – and with what additional cues – strongly affects successful learning of such contrasts (Hayes 2003; Maye 2000; Maye, Werker, & Gerken 2002).

Speech production in children

Let us now turn to the child's second major task in phonological acquisition: mastering the production of native-language speech sounds. Because adult-like competence in the production of speech sounds requires the maturation of the articulators in the mouth and other relevant physical and neurological systems, babies do not begin producing what can arguably be labeled true phonemes before 18 months (Schwartz 1983). However, as we shall see, prior to and after this point children advance through a number of definitive stages before adult-like productions are achieved. Although investigations of children's early speech productions can be accomplished quite straightforwardly – now that asking children to talk is an option – there are nevertheless a variety of methods that have been employed over the years to collect speech-production data in children. Before discussing the various stages in the development of speech sounds, it seems wise to review these techniques.

Methods in investigations of children's speech production

Diary studies

The most commonly used method for collecting speech production data in children is also the simplest one. In a diary study, documentation of a child's spontaneous speech productions are kept over a prolonged period of time, usually by means of audio or video recording (although in some early cases, only in longhand). The speech samples are typically collected in settings where the child feels most comfortable and/or when they are engaging in activities they enjoy, such as playing with toys. The recorded information is then transcribed manually, in the case of phonological studies, into phonemic or phonetic transcription. Of course, accuracy of the data is a primary concern and researchers employ a variety of methods to ensure the reliability of transcribed data. The frequency with which the data is collected in any time period is also of concern, especially as children's phonological acquisition happens so rapidly and can change considerably from one time to the next. Thus, speech samples may be collected on a daily, weekly or monthly basis, depending on the research question. Similarly, the length of time such data continues to be collected ranges over weeks to years (the latter, in the case of longitudinal studies).

One singular advantage to the diary study is that it is an extremely versatile method; due to its simplicity, it can be used in a child's home, school, or in a child-friendly, research laboratory. However, there are some disadvantages to this method as well. Transcribing speech samples at reliable levels of accuracy is a time-consuming process. Even 20 minutes of recording can take six times as long to transcribe. Also, reliability measures usually involve more than one transcriber and a careful consensus process, increasing the amount of time necessary to obtain dependable data. Also, despite the field's rigorous attention to transcription reliability it may simply be the case that listeners are not accurate enough, in which case acoustic measurements should be used – another path that is extremely time-consuming. Nevertheless, the diary study (or speech sample) remains the most common method used to obtain speech production data because it can provide not only a moment-to-moment record of the development of a child's phonological system, but also a record of how the child progresses from one stage to the next. Fortunately, many researchers that employ this method share their data with others. In fact, one of the most widely-used resources in language acquisition research is the Child Language Data Exchange System (MacWhinney & Snow 1985), which is a computer-based archive of transcripts of the speech of both monolingual and bilingual children that can be accessed online (http://childes.psy.cmu.edu/).

Controlled experiments

Another method that is used to investigate children's phonological acquisition is the controlled experiment. In controlled experiments, a number of children's speech productions of particular phonological stimuli are elicited through tasks they typically enjoy, such as picture-naming or game-play, either with an experimenter or via a computer. These studies usually take place in a laboratory setting but with the prevalence of laptops and portable recording devices can also be conducted in a child's home or school. This method is ideal for investigations of common patterns in children's phonological development – as a means of ascertaining developmental norms, for example. Elicited responses are usually recorded and transcribed following the same guidelines as are used in diary studies; however, in this case, the quantity of data is usually significantly less per child because only target utterances need be transcribed.

One of the primary advantages to the controlled experiment is fairly obvious: researchers can control the productions that children make by carefully selecting for elicitation only utterances that exhibit a particular phonological or phonetic property. However, although transcription time can be limited in this way, it is still a lengthy process that is directly related to the number of children involved in the study: the more children, the greater the transcription time required. Also, recruiting children at precisely the age when a particular phenomenon is occurring is a tricky business because children exhibit high degrees of variability in the onset and/or cessation of various stages in phonological development. Thus, the researcher must target an approximate age range and hope that children will display the phenomenon in question. For this reason, many children are usually tested to ensure the phenomenon is being accurately investigated.

Findings from studies in children's speech production

As everyone who has been around children for any length of time knows, children just beginning to talk do not produce speech sounds with immediate success, nor do they initially produce adult-like syllables or words. In fact, despite its relative rapidity, children's progression to adult phonological competence is complex and varies greatly from child to child. However, since Jakobson

(1941/1968), researchers have noted that children learning any language nevertheless go through similar stages in phonological development, although not at precisely the same ages.

Babbling

Babies' first attempts at producing speech sounds begin around four to six months, during a period referred to as vocal play (Stark 1986). Vocal play is so-called because babies truly seem to be experimenting – playing – with the range of speech-like noises they can produce. As if testing their control over their vocal apparatus, the range of sounds a baby produces during this period is various, including different vowel and consonant sounds that fluctuate in length and that are sometimes combined into rudimentary consonant-vowel sequences (Hoff 2001; Menn & Stoel-Gammon 2005). Prior to vocal play babies generally produce only vegetative noises, such as burping and crying (newborn to two months), and laughter and cooing (two to four months). Although not considered speech-like, cooing sounds are most akin to vowels but are usually longer in length than those of adults. Cooing sounds are additionally characterized by the inclusion of back consonants, like /k/ and /g/; hence, this period is alternatively referred to as the gooing stage (Menn & Stoel-Gammon 2005; Yavas 1998).

It is not until around six to eight months that syllables with adult-like timing emerge in infants' babble. In this landmark period, referred to as the canonical or reduplicated babbling stage (Oller 1980), babies begin producing strings of the same consonant-vowel sequence repeated over and over, such as *bababa* or *dadada*. In fact, because the syllables produced at this stage are so adult-like, many parents are convinced that their child has begun producing words when, in fact, there is evidence that babies actually produce these syllables without any semantic intent (Stark 1986). From around eight months, babies begin producing what is called variegated babbling, which is characterized by the production of *different* consonant-vowel sequences (e.g., *dabadi*) instead of the same one produced repeatedly as in canonical babbling. In this stage, the baby's repertoire of consonants and vowels is greatly expanded and utterances begin to take on the prosodic characteristics of fluent speech. In fact, babies in this stage often sound like they are having conversations, except that none of the utterances are meaningful.

The influence of the native language on babbling. For some time it was thought that babies produced all possible speech sounds during babbling, including ones not present in the native language (Jakobson 1941/1968). This claim is now believed to be erroneous; although some nonnative sounds may appear in infant babbling, it is certainly not the case that every possible speech sound is present. Also, by the age of 8 months (and perhaps earlier), influences of the native language on sounds in babble can be seen. For example, de Boysson-Bardies, Sagart, and Durand (1984) showed that naïve French listeners could correctly identify the babble of 8-month-old French babies versus the babble of Arabic or Chinese babies and that phonetically-trained listeners could make the same distinction among the babble of 6-month-old babies. This suggests that language-specific influences on babbling are already noticeable at that time. Also, de Boysson-Bardies and Vihman (1991) showed that the frequency of certain common speech sounds in the babble of babies learning French, English, Swedish, and Japanese differed relative to the frequency of those sounds in the native language. These and many other studies make clear that later stages of babies' babble begin to reflect an influence of the native language – interestingly at around the same time that infant speech perception is similarly affected. This is not to say, however, that there are no cross-linguistic similarities in babbling. For example, Locke (1983) coalesced the results of a number of studies on babbling in different languages and showed that for babies 11 to 12 months old, around 95 percent of the consonants present in cross-linguistic babble are the same: namely, /p, t, k, b, d, g, s, m, n, w, j, h/. He claimed that there is at least some biological basis to sounds in babbling, which would account for the similarity in the sounds produced.

In sum, infants begin the process of producing speech sounds by first getting to know what parts of their mouths are used for which sounds (vocal play). The initial repertoire of sounds is arguably a function of both biological predisposition and the native language. Once babies have rudimentary control over their articulators, more adult-like syllables begin to occur, albeit with little variety in the sound strings produced (reduplicated canonical babbling). Finally, more diverse consonant-vowel combinations (variegated babbling) are produced that exhibit more stability and adult-like prosody than previous utterances.

Transition to first words

For most children the transition between babbling and first words takes place around 12 months. This transition is typically relatively smooth, with many of the sounds and/or sound patterns produced in babbling appearing also in the child's first attempts at meaningful utterances (Vihman 1992, 1996). Some of these attempts at speech result, not in true words, but in what are referred to as proto-words (Bates 1976). Proto-words are utterances that are used like real words in that there is a consistent meaning associated with a particular sequence of sounds, but they do not resemble any real word of the language. Typically, the articulation of such utterances is less well controlled than for true words, resulting in a higher degree of variability in the pronunciation of proto-words (Menn & Stoel-Gammon 2005).

As with other phases in speech production, there is generally overlap in the kinds of utterances produced during this transition period. In this case, late babbling may coincide with proto- and genuine words, with the proportion of the latter steadily increasing as the child's motor control and other relevant abilities stabilize. These findings contradict an earlier claim by Jakobson (1941/1968), who argued that there are no similarities between babbling and first words, and that the transition is marked by a silent period – a claim which is now generally considered to be false, except in a very few cases.

Sounds in first words

Children's first words tend to be simple in structure, consisting of consonant-vowel (CV) syllables for the most part. In addition, the sounds in first words usually reflect only a small portion of the total consonant/vowel inventory present in the target language (Ingram 1999; Yavas 1998). The particular order in which they are acquired has been a subject of much study with considerable focus on the acquisition of consonants, as vowels are generally thought to be much more difficult to categorize (for summaries, see, e.g., Locke 1983; Hoff 2001; Templin 1957; Yavas 1998). As an example, the typical developmental trajectory for English is characterized by the production of the stops and nasals starting prior to the age of two years, with successful acquisition of those sounds and perhaps /h/ and /w/ by the age of three. Following these are additional fricatives, such as /f/ and other sonorants, such as /j/ and /l/. These latter sounds may begin to be produced correctly at around three years of age but are often not fully mastered until later, especially in the case of /l/. Other sounds, like /ɹ/, the affricates, and the remaining fricatives may not appear until as late as four years of age and may not be mastered until the ages of seven or eight. The criteria for successful acquisition of a sound varies among researchers but typically requires that the child be able to produce the sound in at least two of three possible syllabic positions (initial, medial, final). Regardless, the majority of consonants are generally acquired by the age of three or four.

The acquisition of consonants in English bears some similarity to the acquisition of consonants in other languages, in that certain sounds are simply more difficult to articulate (e.g., /ð/ and /ɹ/) and are therefore late-acquired in many languages. However, although there are also certain sounds that appear relatively early (like /p/ and /m/) in any language, the presence or absence of sounds in children's first words is subject to language-specific influence, much as we have seen in other domains. For example, Ingram (1999) compared the acquisition of /v/ in children learning English, Estonian, Bulgarian, and Swedish. The sound /v/ in English is relatively infrequent and also tends to be later-acquired. In contrast, /v/ is more frequent in the other three languages, and results showed that /v/ is, in fact, acquired earlier in those languages. Recent work by Zamuner (2003) confirms the influence of the frequency of sounds in the native language on the production of those sounds in early speech. Specifically, she compared the frequency of occurrence of final consonants in the productions of English-speaking children to their cross-linguistic versus target-language frequency. She found that the children's productions more accurately reflected the frequency of final consonants in the target language.

The representation of sounds in first words

Despite the focus on the acquisition of individual speech sounds that can be seen in the literature, the issue of whether early words are actually represented as sequences of individual segments is a subject of some debate. Ingram (1992) has long maintained that children's representations of early words is adult-like and can best be characterized in terms of the acquisition of featural contrasts (consistent in some respects with the proposal of Jakobson 1941/1968). This notion is further argued in Menn and Stoel-Gammon (2005), who maintain that important generalizations are missed when featural analysis is excluded. For example, a child who produces *pot* as [bat] may be characterized as simply having an unsuccessful articulation of the initial segment, /p/. However, an analysis in terms of features shows that the child has all features correct in this sound except for voicing (Menn & Stoel-Gammon, p. 80).

In contrast, Oller and Steffens (1994) argue that early words are represented in terms of syllables and not in terms of individual segments or features. This finding is arguably supported by infant speech perception research which shows that babies discriminate strings with differing numbers of syllables but not strings with the same number of syllables but a differing number of consonants (Bijeljac-Babic, Bertoncini, & Mehler 1993). Oller and Steffens' proposal, however, is largely based on evidence that consonants and vowels do not initially combine freely in early words – as might be expected if they were functioning as individual units – but instead tend to adhere to co-occurrence restrictions, for example, back consonants with rounded vowels only. The possibility of the syllable as the basic representational unit in early words was also explored by Vihman (1992), who examined the syllable productions of twenty-three children learning different languages. She concluded that while some children appeared to use the syllable as a basic unit of representation for early words, not all children did so. Her ultimate conclusion is that children are highly individualistic in how they construct their early lexicons and may base their representations on phonetic characteristics of the target language, articulatory gestures, and/or syllables. This conclusion cannot be taken too lightly as the degree of variability in children's phonological acquisition in general is known to be quite large.

The emergence of phonology

By the time a child has achieved a lexicon of around 50 words (at approximately 18 months of age), evidence of all levels of phonological structure can generally be found. Perhaps one of the most striking sources of evidence for a phonological system can be found in the kinds of errors children make in speech production (for summaries see, for example, Hoff 2001; Ingram 1989; Yavas 1998). Interestingly, these error patterns are consistent across children learning all languages, suggesting that there is some similarity in how young children construct their phonologies. For example, children learning languages that allow syllables to contain final consonants (VC, CVC, etc.), typically omit the final consonant from those utterances – pronouncing *dog* as [da], for instance. Similarly, children producing syllables that contain sequences of consonants (CCV, VCC, etc.), typically omit one of the consonants – pronouncing *skate* as [ket], for instance. These two processes, referred to as final consonant deletion and cluster reduction, respectively, show that children have an awareness of syllable structure that allows them to manipulate parts of the syllable independently, but not randomly (i.e. sounds in all positions are not randomly omitted, just those in particular syllabic positions). Children's tendency to produces words containing only CV syllables, the end-product of these omissions, has been taken as evidence for a default structure towards which all children gravitate in early speech production (Fikkert 1994; Ohala 1992). Regardless of whether this is a function of some innate predisposition or is a result of language-specific factors, children eventually overcome this tendency (if required) in the acquisition of the native language syllable structure (for further details see Fikkert 1994; Levelt, Schiller, & Levelt 1999).

Another process that is present in children's early speech production that is revealing of emerging phonological structure is referred to as weak-syllable omission. This occurs when a child omits an unstressed (or weak) syllable that precedes a stressed (or strong) syllable. For example, a child might pronounce the word *giraffe* as "raffe", where the unstressed first syllable, "gi", is omitted. Work by Gerken (1994) further showed that children learning English do not omit just any weak syllable, but only those that are inconsistent with the dominant stress pattern of the language (for English, most words have strong-weak stress). Specifically, she examined children's weak-syllable omissions in English words and found that children omit weak syllables less often from words like *monkey*, which has a strong-weak stress pattern, than from words like *giraffe*, which have a weak-strong stress pattern. Thus, "raffe" was a common reduction whereas "monk" was not. The fact that the child omits weak-syllables over strong ones – and only certain weak syllables at that – suggests that children are aware of stress differences among syllables and that this information must somehow be encoded in the words they produce.

Finally, although we have already seen arguments for evidence of a featural level of organization in children's early words, there are a number of other processes that support such a conclusion as well. For example, the first words of many children undergo a process referred to as fronting, where sounds produced at the back of the mouth are substituted by ones produced at the front of the mouth. For instance, a child might pronounce *key* as [ti], where dorsal /k/ has been substituted by coronal /t/. Also, many early words exhibit what is called gliding, where the liquid sounds, /ɹ/ and /l/ are substituted with the glides /w/ and /j/, respectively. Both processes indicate that children's representations must encode, for example, differences in place and manner of articulation (Yavas 1998). As mentioned earlier, all of these processes are typical for children learning any language and, taken as a whole, provide strong evidence for emerging phonology. The fact that most of these processes disappear by the age of three attests to the rapidity and sophistication with which children accomplish the complex process of phonological acquisition.

Summary of speech production in children

As in the task of speech segmentation, we have seen that the child's path towards phonological competence in the production of speech is a complicated one, but a task that is nevertheless achieved relatively quickly. By six months of age, babies begin producing sounds that have adult-like timing and prosody, a repertoire that expands and refines as the child progresses through the later stages of babbling to the production of proto-words and finally true words around 12 months. Native-language influence is seen in the nature and frequency of sounds produced in babbling and first words. And although the initial nature of phonological structure is debatable, there is a large body of evidence that suggests that by the age of two years, children manipulate the productions of words in such a way that evidence of the existence of all levels of phonological structure is apparent.

Comparisons with second language acquisition

As with L1 speech perception, some of the aspects of speech production in L1 are clearly applicable only to that domain. However, other aspects bear similarity to processes in L2 acquisition. For example, the influence of the native language on the production of speech sounds is clearly something that strongly affects the production of L2 sounds as well (see the chapter by Major, this volume, for a discussion of transfer; see also Lado 1957; Major 1987). Furthermore, the processes that children undergo in the acquisition of a L1 phonology, such as cluster reduction, can also be seen in emerging L2 phonologies (e.g., Sato 1984; Hansen 2004). These similarities suggest that the occurrence of such phenomena are not the explicit domain of either L1 or L2 phonological acquisition and that the reasons behind these similarities may lie in universal tendencies (see Eckman, this volume, on universals). Although comparisons of L1 and L2 phonologies have long been undertaken as a means of predicting which sounds in the L2 might, for example, prove particularly difficult (e.g., Stockwell & Bowen 1983), less attention has perhaps been paid to the similarities in the acquisition process itself (but see Eckman 2004).

Conclusions

As newborns, we are introduced into a disconcerting and overwhelming world of sound, only some parts of which are necessary for speech and provide us with the means to communicate with others. We are fortunate that our first forays into the maelstrom that is speech begin prior to birth and progress with astonishing rapidity after we are born. Although equipped with a perceptual acuity that allows us to initially discriminate almost all possible phonetic contrasts in languages, we quickly become attuned to the specifics of our native language to the exclusion of all else. For this task, we are able to use sophisticated learning mechanisms that take advantage of statistical information in the language we hear and that allow us to apply that information at an abstract level to new utterances. From six to eight months onwards, our advances in perceptual tasks coincide with our first attempts at speech production. Although our initial attempts sound most unlike adult productions, they nevertheless allow us to explore the limits of our articulatory apparatus. From our first success with adult-like syllables, we continually expand our repertoire, with sounds present in babbling typically appearing in our first true words as well. Which sounds or structures first appear in our early words is heavily influenced by the frequency of the sounds in the native language relative to each other, although cross-linguistic patterns in phonological acquisition are also attested. At around 18 months, when we have successfully produced at least 50 words, our phonological knowledge coalesces and the emergence of true phonological organization is apparent. From this point on, our progress towards adult-like competence in this domain continues to expand, until by the age of six we have completed this first and most fundamental aspect of language acquisition.

Suggested readings

Bernhardt, B. & Stemberger, J. P. 1998. *Handbook of phonological development: From the perspective of constraint-based non-linear phonology*. San Diego, CA: Academic Press.

de Boysson-Bardies, B. 2001. *How language comes to children: From birth to two years*. Cambridge, MA: MIT Press.

Ferguson, C. A., Menn, L. & Stoel-Gammon, C. (eds.). 1992. *Phonological development: Models, research, implications*. Timonium, MD: York Press.

Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. 2000. *The scientist in the crib: Minds, brains, and how children learn*. New York: Harper Collins.

Ingram, D. 1989. *Phonological disabilities in children*. London: Cole and Whurr.

Jusczyk, P. W. 1997. *The discovery of spoken language*. Cambridge, MA: MIT Press.

Karmiloff, K., & Karmiloff-Smith, A. 2001. *Pathways to language: From fetus to adolescent*. Cambridge, MA: Harvard University Press.

Vihman, M. M. 1996. *Phonological development: The origins of language in the child*. Oxford, UK: Blackwell.

Yavas, M. 1998. *Phonological development and disorders*. San Diego, CA: Singular Publishing Group, Inc.

References

Bates, E. 1976. *Language and context: The acquisition of pragmatics*. New York: Academic Press.

Best, C. W., McRoberts, G. W., & Sithole, N. N. 1988. The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14: 345–360.

Bijeljac-Babic, R. J., Bertoncini, J., & Mehler, J. 1993. How do 4-day-old infants categorize multi-syllabic utterances? *Developmental Psychology* 29: 711–721.

de Boysson-Bardies, B., Sagart, L., & Durand, C. 1984. Discernable differences in the babbling of infants according to the target language. *Journal of Child Language* 8: 511–524.

de Boysson-Bardies, B., & Vihman, M. M. 1991. Adaptation to language: Evidence from babbling and first words in four languages. *Language* 67: 297–319.

DeCasper, A. J., & Fifer, W. P. 1980. Of human bonding: Newborns prefer their mothers' voices. *Science* 208: 1174–1176.

DeCasper, A. J., & Spence, M. J. 1986. Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development* 9: 135–150.

Eckman, F. 2004. From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second Language Acquisition* 26: 513–549.

Eckman, F. R. 2008. Typological markedness and second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. 1971. Speech perception in infants. *Science* 171: 303–306.

Fikkert, P. 1994. *On the acquisition of prosodic structure*. Leiden: Holland Institute of Linguistics.

Gerken, L. A. 1994. A metrical template account of children's weak syllable omissions. *Journal of Child Language* 21: 565–584.

Gomez, R. L., & Gerken, L. A. 1999. Artificial language learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* 70: 109–135.

Hansen, J. G. 2004. Developmental sequences in the acquisition of English L2 codas: A preliminary study. *Studies in Second Language Acquisition* 26: 85–124.

Hayes, R. 2003. *How are second-language phoneme contrasts learned?* Unpublished doctoral dissertation, University of Arizona, Tucson.

Hoff, E. 2001. *Language development* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Ingram, D. 1999. Phonological acquisition. In M. Barrett (ed.), *The development of language*, East Sussex: Psychology Press, 73–98.

Ioup, G. 2008. Exploring the role of age in the acquisition of a second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Jakobson, R. 1968. *Child language, aphasia, and phonological universals* (R. Keiler, Trans.). The Hague: Mouton. (Original work published in German in 1941).

Jusczyk, P. W. 1997. *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. 1993. Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language* 32: 402–420.

Kemler Nelson, D., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A. E., & Gerken, L. A. 1995. The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development* 18: 111–116.

Kuhl, P. K. 1976. Speech perception in early infancy: The acquisition of speech-sound categories. In S. K. Hirsh, D. H. Eldredge, I. J. Hirsh, & S. R. Silverman (eds.), *Hearing and Davis: Essays honoring Hallowell Davis*, St. Louis, MO: Washington University Press, 265–280.

Kuhl, P. K. 1987. Perception of speech and sound in early infancy. In P. Salapatek & L. Cohen (eds.), *Handbook of Infant Perception*, New York: Academic Press, 275–382.

Kuhl, P. K., & Miller, J. D. 1975. Speech perception by the chincilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science* 190: 69–72.

Lado, R. 1957. *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.

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

Levelt, C. C., Schiller, N. O., & Levelt, W. J. M. 1999. A developmental grammar for syllable structure in the production of child language. *Brain and Language* 68: 291–299.

Locke, J. L. 1983. *Phonological acquisition and change*. New York: Academic Press.

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

MacKain, K. S., Best, C. T., & Strange, W. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics* 2: 368–390.

MacWhinney, B., & Snow, C. 1985. The child language data exchange system. *Journal of Child Language* 12: 271–296.

Major, R. 1987. A model for interlanguage phonology. In G. Ioup & S. Weinberger (eds.), *Interlanguage phonology*, New York: Newbury House/Harper & Row, 101–124.

Major, R. C. 2008. Transfer in second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. 1999. Rule learning by seven-month-old infants. *Science* 283: 77–80.

Maye, J. 2000. *Learning speech sound categories from statistical information*. Unpublished doctoral dissertation, University of Arizona, Tucson.

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

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. 1988. A precursor of language acquisition in young infants. *Cognition* 29: 143–178.

Menn, L., & Stoel-Gammon, C. 2005. Phonological development: Learning sounds and sound patterns. In J. Berko Gleason (ed.), *The development of language* (6th ed.), New York: Pearson.

Miller, J. D., Wier, C. C., Pastore, R. E., Kelley, W. J., & Dooling, R. J. 1976. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America* 60: 410–417.

Morgan, J. L. & Demuth, K. 1996. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Erlbaum.

Ohala, D. 1992. A unified theory of final consonant deletion in early child speech. In C. Gerken & P. Piñar (eds.), *Coyote papers: Working papers in linguistics from A to Z*, Tucson: University of Arizona Press.

Oller, D. K. 1980. The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (eds.), *Child phonology* (Vol. 1), New York: Academic Press, 93–112.

Oller, D. K., & Steffens, M. L. 1994. Syllables and segments in infant vocalizations and young child speech. In M. Yavas (ed.), *First and second language phonology*, San Diego: Singular Publishing, 45–61.

Penfield, W., & Roberts, L. 1959. *Speech and brain mechanisms*. Princeton, NJ: Princeton University Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. 1996. Statistical learning by 8-month-old infants. *Science* 274: 1926–1928.

Sato, C. 1984. Phonological processes in second language acquisition: Another look at interlanguage syllable structure. *Language and Learning* 34: 43–57.

Schwartz, R. 1983. Diagnosis of speech sound disorders in children. In I. J. Mertus & B. Weinberg (eds.), *Diagnosis in speech-language pathology*, Baltimore, MD: University Park Press, 113–149.

Sebastián-Gallés, N., & Soto-Faraco, S. 1999. Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition* 72: 111–123.

Stark, R. E. 1986. Prespeech segmental feature development. In P. Fletcher & M. Garman (eds.), *Language acquisition* (2nd ed.), Cambridge: Cambridge University Press, 149–173.

Stockwell, R., & Bowen, D. 1983. *The sounds of English and Spanish*. The Hague: Mouton.

Tees, R. C., & Werker, J. F. 1984. Perceptual flexibility: Maintenance of recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology* 38: 579–590.

Templin, M. C. 1957. Certain language skills in children: Their development and interrelationships. *Institute of Child Welfare Monographs*, *26*. Minneapolis: University of Minnesota Press.

Trehub, S. E. 1973. Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology* 9: 91–96.

Trehub, S. E. 1976. The discrimination of foreign speech contrasts by infants and adults. *Child Development* 47: 466–472.

Vihman, M. M. 1992. Early syllables and the construction of phonology. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (eds.), *Phonological development: Models, research, implications*, Timonium, MD: York Press.

Vihman, M. M. 1996. *Phonological development: The origins of language in the child*. Oxford, UK: Blackwell.

Weinreich, U. 1953. *Languages in contact*. The Hague: Mouton.

Werker, J. F., & Pegg, J. E. 1992. Infant speech perception and phonological acquisition. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (eds.), *Phonological development: Models, research, implications*, Timonium, MD: York Press, 285–311.

Werker, J. F., & Tees, R. C. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7: 49–63.

Yavas, M. 1998. *Phonological development and disorders*. San Diego, CA: Singular Publishing Group, Inc.

Zamuner, T. 2003. *Input-based phonological acquisition*. Outstanding Dissertations in Linguistics Series. New York: Rutledge.

CHAPTER 2

Exploring the role of age in the acquisition of a second language phonology

Georgette Ioup
University of New Orleans

Introduction

Scovel, in his comprehensive overview of the critical period in 1988, stated that phonological accents in a second language (L2), more than other linguistic skills, would most exhibit age effects because accent was the only part of language that was physical and demanded neuromuscular programming. He further stated that phonological age effects were the result of maturational changes in the brain. How accurate are Scovel's claims? Does the research show that phonology more than any other aspect of language is influenced by the age at which the learner first encounters the L2? Are early L2 learners in reality better at producing native-like accents than adult learners? Are any age effects the result of neurological changes in the brain, as Scovel has argued, or are they caused by other factors such as first language (L1) transfer, the disposition of the learner, or the environment of learning? To address these questions, this chapter is divided into five sections. In the first section, the empirical data relevant to age effects in acquiring an L2 phonology are reviewed. In specific, this section addresses the following questions: What is 'accent'? To what extent does age affect pronunciation more than other areas of language? What is the role of perception in accent production? What differences exist between child and adult L2 learners with respect to L2 accent? What is the age of onset of L2 accent? The second section of the chapter examines several models developed to explain age effects in L2 phonological acquisition. It provides an overview and critical discussion these models, and also addresses individual differences in L2 accent. The third section of the chapter discusses methodological issues in age research. In the final two sections, suggestions for future research and conclusions are given.

Research issues in age and L2 phonology

What is 'Accent'?

A question that has been addressed by the research is what the properties of accent are – in other words, what is 'accent'? One approach to investigating this question is to compare the acoustic properties in the speech of L2 learners with equivalent sounds in the speech of monolingual native speakers (see Munro, Chapter 7 of this volume, for an extensive discussion on the detection of foreign accent). Research has focused on several aspects of phonology, three of which will be discussed here as they have received the majority of attention in this area: voice onset time of consonants in English, vowel duration in English, and syllable structure production. This type of research has typically focused on adult L2 learners, with the overall finding that learners from the same L1 background have been found to make similar deviations from native norms.

Several studies (e.g., Flege 1980; Flege & Hillenbrand 1987; Nathan, Anderson, & Budsayamongkon 1987) have measured the way adult L2 learners produce the voice onset time (VOT) of stop consonants in English (see also Chapter 8 by Zampini for an extensive discussion of VOT). In comparison to most other languages, English voiceless stops in syllable initial position have a very long lag after the stop release before the onset of voicing. Research has found that learners may transfer their L1 VOT values in the production of L2 stops: in research on Saudi Arabic learners of English, Flege (1980) found the learners to be using their L1 values to produce the L2 stops. Nathan, Anderson, and Budsayamongkon (1987) obtained similar results for L1 speakers of Spanish. However, some studies established that learners do not substitute the exact L1 values for L2 sounds, but use a value somewhere between the two. For example, Flege and Hillenbrand (1987) ascertained that L1 French learners of English produced English stops with a VOT value that was neither French nor English, but was influenced by French.[1] As Zampini notes in her discussion of VOT in Chapter 8 of this volume, this may also be due to equivalence classifications (Flege 1995) the learners make between sounds in the L1 and sounds in the L2.

A second area of research within this domain is vowel duration. One study in this area is Mack (1982), who examined vowel duration in English, which varies systematically before voiced and voiceless consonants. English vowels are lengthened before a voiced consonant in comparison to those produced before a compa-

1. Also interesting was that they had altered the values of their native French stops as well, to make them more English-like. Similar influence of each language on the other was observed in Caramazza, Yeni-Komshian, Zurif, & Carbone (1973), Flege & Eefting (1987), Mack, Bott & Boronat (1995) and Moen (1995, cited in Mack et al. 1995).

rable voiceless consonant. French, like most other languages, does not exhibit this type of vowel lengthening. Using minimal pairs that differed only in the voicing of the syllable-final consonant, she determined that adult French learners of English transferred French vowel length values to their production of English vowels. Likewise, Flege, Munro, and Skelton (1992), measuring the acoustic output of adult L1 Mandarin learners of English, ascertained that even very experienced learners could not accurately produce the English vowel durations, nor, in addition, could they reliably produce an obstruent (stop, fricative, and affricate) voicing contrast in word-final position. Flege, MacKay, and Meador, (1999) determined that adult Italian learners of English produced English vowels that were more Italian-like in that they had the formant movement characteristics of equivalent Italian vowels.

Munro (1993) also examined the productions of 10 English vowels, in research on adult Arabic learners of English. Both new vowel sounds and those that are similar to L1 vowels were found to be different, even when produced by learners with several years of L2 experience. Acoustic analysis of their pronunciation revealed that their accent was the result of both temporal (formant movement) and spectral properties. Vowel length also varied significantly from native English pronunciation. In general, the vowels produced by the learners were shorter than equivalent vowels produced by native speakers and vowel length was held constant before voiced and voiceless consonants. These non-native features of their production could be attributed to the characteristics of the Arabic vowel system.

Finally, at the syllabic level, adult-onset learners have been shown to transfer the phonotactic constraints of the L1. One of the first to investigate L2 syllabic processes was Broselow (1984), who determined that English syllable structure was produced differently by Iraqi and Egyptian Arabic speakers according to the differing constraints of their native dialects. A more recent study by Altenberg (2005) found that adult Spanish-speaking learners of English who could demonstrate metalinguistic and perceptual awareness of permissible word-initial consonant clusters in English continued to transfer the syllabification constraints of Spanish in producing them (see also Chapter 8). There has also been some research on intonation and prosodic domains, as discussed in Munro, Chapter 7 of this volume.

In summary, one may conclude from these and other studies examining the acoustic output on late-onset L2 learners that their accents retain features of the native language in that acoustic values in the L2 may be based on parameters of the L1. Additionally, studies have shown that values may be somewhere in between the L1 and the L2. However, as Zampini notes in Chapter 8 of this volume, other factors, such as length of stay and experience in using the L2 as well as perceived phonetic similarity between sounds may also be a factor.

Age, pronunciation, and other aspects of language

Another question of posed by researchers is whether age influences pronunciation more than other language areas. Several studies have addressed this question. Flege, Yeni-Komshian, and Liu (1999) studied the L2 English of 240 native speakers of Korean with varying ages of arrival (AOA) in the US. Native-speaking judges were asked to rate the quality of their accents. The level of the learners' morphosyntax was also assessed, through a grammaticality judgment test. Results showed that only the scores for degree of accent were completely dependent on age of arrival. Morphosyntax scores were found to be influenced by factors in addition to age; both the amount of education the Korean subjects had received and the degree to which they used the L2 were significant variables.

It has also been shown that late-onset learners can be identified using only their pronunciation. Ioup (1984) asked linguistically trained judges whether they could separate adult learners of English from two unidentified L1 backgrounds into distinct groups based on either phonological or syntactic cues. Subjects produced a two-minute discourse and read a prepared text. The short discourse was reread by a native speaker of English, preserving only the subjects' syntax. The prepared text was used to provide information on subjects' pronunciation. Results showed that the identification task was quite simple using only phonological cues, but impossible with just syntactic information.

The role of perception in accent production

Difficulty in producing new sounds is often attributed to imperfect perceptual ability. Evidence indicates that if the phonological contrasts cannot be perceived, speakers will have difficulty producing them (Rochet 1995). Several studies have documented the inability of adults to discriminate speech contrasts that do not exist in their native language. An early study by Miyawaki, Strange, Verbrugge, Liberman, Jenkins, and Fujimura (1975) investigated whether adult speakers of Japanese could distinguish between synthetically generated pairs of English /r/ and /l/ which varied acoustically in their proximity to the English /r-l/ boundary, and determined that they could not. Native speakers of English, on the other hand, had a strictly defined boundary between the two and were able to correctly identify members of the pair no matter how close to the boundary they were produced. Interestingly, however, the same comparisons presented as music contrasts, rather than language, evoked highly accurate judgments by the Japanese subjects, indicating that these acoustic contrasts are accessible when divorced from language. Subsequent studies of Japanese adults have replicated the difficulty in differentiating the acoustic properties of the English /r/ and /l/ segments (Best & Strange 1981; Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, & Siebert

2003; Mochizuki 1981; Sheldon & Strange 1982; Strange & Dittmann 1984) (see also Chapter 6 on perception by Strange & Shafer). These results hold even among those with extensive exposure to L2 English (Flege, Takagi, & Mann 1996). When a training component is incorporated into the studies, discrimination ability continues to be limited and is of restricted duration (Bradlow, Pisoni, Yamada, & Tohkura 1997; Bradlow, Akahane-Yamada, Pisoni, & Tohkura 1999; Lively, Logan, & Pisoni 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada 1994; Logan, Lively, & Pisoni 1991; MacKain, Best, & Strange 1981; McCandliss, Fiez, Protopapas, Conway, & McClelland 2002 [for a full discussion of the effects of L2 speech training see Chapter 11 by Bradlow]).

Flege, Bohn and Jang (1997) also found that adult Spanish learners of English differed from native English speakers in their identification of members of a synthetic /i/ to /ɪ/ continuum. Native speakers based their responses on spectral characteristics of the sounds whereas native Spanish speakers tended to judge the sounds in terms of vowel duration. As predicted above, they found a relationship between accuracy in vowel perception and production by speakers of German, Mandarin, Korean, and Spanish.

Comparing child L2 learners to adults

The preceding studies have focused predominately on adult L2 learners, i.e. those who began learning the L2 in adulthood. The question remains whether child learners are any different. Is early-onset L2 speech less accented than late-onset speech? In the research paradigms, early onset is usually defined as L2 learning before the age of 8, while late onset addresses learners over age 16. There are two approaches to comparing early and late learners. The first uses large numbers of adult subjects divided into groups according to the age at which their L2 acquisition commenced. Subjects produce speech using a variety of elicitation measures, depending on test design. Their productions are then evaluated by native speakers using a linear scale with one end indicating highly accented speech and the other native-speaker pronunciation. Speech output produced by native speakers is interspersed with the nonnative speech for control.

One of the first such studies was that of Asher and Garcia (1969), which assessed the accents of 71 Cuban learners of English. Results indicated that those who had entered the US at age 6 or earlier had the highest probability of being judged native. A similar early study was conducted by Oyama (1976). She tested 60 Italian immigrants with lengths of residence between 5 and 18 years, comparing their verbal output to native-speaking controls. Only those who began L2 acquisition before age 10 were judged to perform in the range of the native speakers. Other variables measured, such as length of residence in the L2 environment and degree of motivation to learn English, did not correlate with performance.

More recent large-scale accent comparison studies found the same negative correlation with age in the accent ratings (e.g., Flege, Bohn, & Jang 1997; Flege & Fletcher 1992; Flege, Munro, & MacKay 1995; Flege, Yeni-Komshian & Liu 1999; Munro, Flege, & MacKay 1996; Patkowski 1990; Piske, MacKay, & Flege 2001; Thompson 1991). Although in some of these studies other variables such as the quality of the input or the amount of either L1 or L2 use had some effect, overall the onset age of L2 remained the most important predictor of degree of foreign accent in the assessment of child and adult onset learners.

The second approach to comparing early and late learners relies on acoustic measurements of speech output. These studies are not as numerous as the rater experiments for acoustic studies require a well-equipped phonetics lab. The main focus of these studies has been on VOT and vowels. In terms of VOT, a study by Flege (1991) compared Spanish speakers who had acquired English either in childhood or as adults in their production of VOT values of the English phoneme /t/. He found that the early learners' VOT values were equivalent to those of monolingual English speakers, while the late learners produced values that were midway between the monolingual Spanish and English values. Thornburgh and Ryalls (1998) also tested early and late Spanish/English bilinguals on English VOT performance, obtaining similar results. Again, age was the determining factor in the ability to produce native English values. As Zampini notes in her discussion of VOT in Chapter 8, early learners appear to have separate phonetic categories for L1 and L2 VOTs while late learners have compromise values due to equivalence values. Finally, as Zampini also notes, speaking rate may be a factor in VOT values.

In a series of experiments using acoustic measurements, Flege, MacKay, and Meador (1999) and Flege, Schirru, and MacKay (2003) examined Italian immigrants to Canada with varying age of arrival (AOA) on their production and perception of English vowel pairs. Accuracy was inversely correlated with AOA, with only the performance of a subset of the earliest arriving participants producing English vowels that were not influenced by their native Italian.

Kim (1995) also examined vowel production, in this case by Korean learners of English grouped into AOA before and after age 16. Using the /i-I/ distinction which does not exist in Korean, she found that the late arriving learners used only vowel length to distinguish the pair, whereas early learners, like the monolingual English controls, used both vowel length and the spectral differences of the vowel pair to categorize the two vowels.

The onset of a phonological accent

Based on the studies reviewed above, it appears that younger is better in acquiring the phonology of a L2. But at what age does an accent begin to appear? Long (1990) suggested that the cut-off age for the ability to acquire native-like pronunciation is age 6. Several experiments support this offset, at least in the area of perception. Pallier, Bosch, and Sebastián-Gallés (1997) tested two groups of Spanish/Catalan bilinguals on their ability to perceive a vowel contrast that occurs only in Catalan. One group had been exposed to both languages from birth while the other began Catalan with the start of schooling at ages 5 to 6. At the time of testing, subjects in both groups were fluent speakers of both languages. Subjects heard synthetically produced vowels and were to decide which member of a Catalan vowel pair they matched. Only subjects in the group that had been exposed to Catalan from birth were able to consistently discriminate the Catalan vowel contrast.

Mack (2003) examined Korean/English bilinguals' perception of English vowel contrasts. The subjects, college students who were fluent bilinguals, were divided into four groups of 15 based on their age of arrival in the U.S. (0–4, 5–9, 10–14, and 15+). They were compared to a group of 15 native speakers. When asked to discriminate computer synthesized continua of the vowel pair /i-I/ (a distinction which does not exist in Korean), only those who had first exposure to English before age five perceived the boundary between the two phonemes in the same range as the native speakers. Additional variables such as length of exposure or degree of Korean proficiency did not strongly correlate with task accuracy. Mack hypothesized that if an L2 learner must form an entirely new category, as the /I/ vowel is for Korean speakers, only those with extremely early exposure will perform like monolingual native speakers.

Studies focusing on production also provide support for the importance of very early onset in native-like pronunciation. A study by Mack, Bott, and Boronat (1995) revealed that even children who were bilingual from birth did not necessarily acquire native-like phones in the non-dominant language. Subjects were 7 French/English bilingual children who grew up in France with native English-speaking mothers. In all cases the dominant language was French though English continued to be spoken in the home. At the time of the study the children were in elementary school. The VOT values of stop consonants in both languages were analyzed and compared to those produced by native-speaking children. Over half the bilingual children produced VOT values in their English pronunciation that were similar to French. This research indicates that, in spite of what the research theories predict, even the youngest bilingual learners are not guaranteed native-like phonetic productions.

Two large-scale studies of Italian immigrants to Canada by Flege and his colleagues (Flege, Frieda, & Nozawa 1997; Flege, MacKay, & Piske 2002) also identified many learners who had immigrated as children and still maintained a detectable accent. They, too, concluded that acquisition of native-like pronunciation was not guaranteed with early onset. In general, subjects in their studies who made greater use of their native language were more likely to have detectable accents in the L2.

To summarize, results from the two types of age comparison studies confirm that late learners are not likely to achieve native-like pronunciation, but also indicate that native-like L2 phonology is normally found only with very early onset and the likelihood diminishes as the onset age increases.

Theoretical positions on age and L2 phonology

Explanations for age effects in phonology

The question that now concerns us is what causes the attested age differences in the ability to acquire native-like pronunciation. There are several schools of thought on this; two of the major ones will be reviewed here. One argues that as it matures, the brain undergoes biological changes that make it impossible for the learner to perceive and produce novel sounds. The other contends that the categories developed for the L1 interfere with the perception of new categories in the L2.

The brain maturation position was first developed by Lenneberg (1967) in the form of a **critical period hypothesis** (CPH) for language acquisition with an onset age of two years and a close at puberty. Lenneberg hypothesized that the decline in ability to acquire a natural language at puberty resulted from the end of neural plasticity and thus the completion of hemispheric lateralization in the human brain. Lenneberg drew major evidence for his hypothesis from data on recovery from aphasia (language loss) after major brain trauma, as well as acquisition patterns among the congenitally deaf and Down's syndrome children. He saw additional support in the observed difficulty to acquire a L2 after the onset of puberty, but clearly his claim was only for primary language acquisition.

Most researchers agree that the evidence supports a biological time frame for L1 acquisition such that after the window of opportunity closes, natural L1 acquisition cannot be achieved. However, there is insufficient information to allow us to know whether specific neurobiological changes are responsible for this end to the ability to acquire a mother tongue, and if so, what they are. Some argue that Universal Grammar (UG) as defined by Chomsky (1981) has atrophied either partially or completely, but again, no biological changes have been found which correlate to any difficulty in accessing UG.

There is much less agreement on the applicability of the CPH to L2 acquisition. Some researchers argue that once a mother tongue is acquired, the cognitive mechanisms that allow for language acquisition are still intact and that L2 acquisition is just as possible from a neurological perspective. Others argue that the neurocognitive mechanisms of language acquisition become defective by the close of the critical period so that native-like attainment in both L1 and L2 cannot occur. Additional debate focuses on whether this time frame is a *critical* or a *sensitive* period. Mack (2003) defines the distinction as follows. The *critical period* applies to the period when **complete** acquisition of some property of language is possible, while *sensitive period* refers to the time frame during which only **partial** acquisition is possible. However, this way of defining the terms is not very enlightening since with this distinction the critical period could be said to last throughout the life span as no one has shown that with sufficient effort and training a particular property cannot be acquired in a late stage of learning. I will not distinguish between the two terms in this overview.

The alternative explanations for phonological age effects rely on perceptual and/or production difficulties caused by interference from the L1 phonology. I will briefly review four models that have been proposed to explain this theoretical approach. First, Best's (1994) *Perceptual Assimilation Model* (PAM) incorporates a combination of perception and production factors. In the very early stages of language acquisition, an infant establishes categories for native language sounds by learning to articulate them (see also Chapter 1 by Ohala for a detailed discussion of child L1 acquisition). Once the categories have been established, phonemic categories that are nonnative will be assimilated to native categories on the basis of articulatory similarities. The more a nonnative sound can be assimilated to a native category, the easier it will be to perceive and then acquire. However, if the L2 contains a phonemic contrast in which both members are perceived as a single native language sound, establishing different categories for the L2 will be extremely difficult.

A second framework is Kuhl's *Native Language Magnet Model* (1992), which is a model of perceptual assimilation. It is developed around phonetic prototypes established by the infant learner. These prototypes are idealized representations of phonetic categories and act as anchors that interfere perceptually with the acquisition of nonnative higher-level phonemic categories. The establishment of these perceptual prototypes occurs at an early phonetic level prior to the categorization of speech into phonemic units. The prototypes shape the mapping between acoustics and perception and reduce perceptual sensitivity near the distributional peak of the prototype. Thus, when the L2 learner encounters a new sound that is similar to a native sound, the prototype acts as a magnet forcing the learner to perceive the new sound as the prototype.

Both of these models account for the changes that occur in phonetic perception by age one, but are unable to offer an explanation for the fact that children beyond age one still possess facility with an L2 phonology and that a decline in ability takes place gradually as the individual ages. The *Speech Learning Model* (SLM) developed by Flege (1995) addresses this problem. He theorized that the mechanisms needed to produce new sounds remain intact, but as with the other models, it is perception that changes with development. He further argues that the ability to discern new contrasts decreases with age because children do not have the native-language perceptual categories as firmly fixed in their phonological system as older learners do. As a result, the younger the learner, the greater will be the likelihood that sounds in the L2 will be perceived on their own terms, without reference to the L1.

Furthermore, Flege (1995) hypothesizes that those phones that do not contrast in the L1 will be the ones that are difficult to perceive in the L2; thus, like Best (1994), he concludes that it is the sounds that are similar but not quite the same in both languages that will be hardest to master. Conversely, the greater the dissimilarity between the L1 and L2 phones, the more likely the learner will notice the difference and thus not rely on the L1 to produce the L2 phone. But he does not tell us what type of continued linguistic or biological change is responsible for the decline in perceptual ability as the learner matures. Thus, his model does not offer insight into the mechanisms that allow child L2 learners to acquire phonology more easily than adult L2 learners (see Chapter 6 by Strange and Shafer for a more extensive discussion of these three models of acquisition).

A model developed by Brown (2000) which is based on the internal structure of the phoneme, can offer some insight. Brown argues that it is not the phonetic properties of the L1 which rigidify perceptual ability, but the structure of the phonemic system. Therefore, the phonemic properties of the L1 system determine how the L2 sound system will be perceived. According to Brown, children learning their L1 acquire knowledge of phonemic representations as well as the features that comprise those representations. Using the hierarchical feature geometry of the L1 phonological system, she can explain which features of the L2 phoneme will be noticed and subsequently related to a specific phone in the L1. As an example, she examines the case of English speakers learning Hindi. She explains that English does not subdivide the feature coronal into finer articulation space as does Hindi, which distinguishes retroflex from non-retroflex sounds within the coronal space. Therefore, English speakers will perceive all coronal sounds made within the coronal space as the same phoneme. She uses this feature geometry to explain why some L1s choose to substitute stop consonants for the English interdental fricatives while others use continuants. However, her explanation falls short. Within feature geometry English does further divide the coronal space, doing so in two different ways: +/–distributed which distinguishes /θ/ from /s/, and +/–anterior which distinguishes /ʃ/ from /s/. What English does not do within the coronal space is distinguish retroflex from non-retroflex sounds. Thus, the feature geometry as Brown defines it will not account for the complete range of substitutions the learner makes.

Evaluating the models of accent acquisition

Though the models discussed above have given us insight into many aspects of phonological age effects, questions linger. One unanswered question concerns the identification of those properties in the L1 that cause learners to produce particular L2 phonetic substitutions. For example, a feature geometry model such as Brown's (2000) model can explain on the basis of contrasts in the L1 some of the L2 substitutions that might occur. But, as noted above, this model cannot account for all substitutions. For example, it is very curious that two dialects of the same language with very similar phonemic inventories will make different substitutions. Thus, Canadian and continental French learners will consistently use different substitutions for the English interdental fricatives. The continental French employ an /s, z/ substitution, while the French Canadians use /t, d/. As far as one can tell, the feature geometries of the relevant sounds are the same in the two languages. A similar phenomenon is seen in two dialects of Arabic: Egyptian and Palestinian. Here again both seemingly have the same inventory of coronal sounds but Egyptian substitutes an /s, z/ for the interdental fricatives while Palestinian uses an /t, d/ substitution.

Another interesting case is that of Korean learners of English who produce an /s, d/ substitution for English interdental sounds. This unusual continuant-stop substitution pair is usually explained by the fact that Korean does not have a /z/ phoneme. The feature geometry approach, as well as the other models mentioned, cannot explain why the Korean learners use a continuant at all. It would seem more systematic to use stop substitutions for both the voiced and voiceless members of the pair.[2]

Individual differences

There is one remaining phenomenon to discuss. We find a great deal of individual variation in L2 phonological acquisition. Some adults are judged closer to native norms than others, while, on the other hand, some children are rated as less than native-like. In addition, individual adults can outperform some children. Moreover, we find adult-onset learners who can pass for native speakers. What factors allow some learners to excel at accent acquisition while others have great difficulty? Many age-related studies of accent correlate learner variables other than age to the

2. Solutions to the differential substitution problem have been suggested by Weinberger (1997) and Lombardi (2003). Weinberger, after demonstrating that the solutions proposed by various theories of L2 phonology are lacking in specific ways, offers instead an account based on underspecification theory. Lombardi, building on Weinberger's work, addresses the problem using optimality theory.

degree of native-like pronunciation and find that they have an impact (cf. Chapter 9 by Hansen Edwards on social factors and variation).

Moyer (2004) discusses this research in her overview of the individual variables that affect the quality of an L2 accent (see also Piske, MacKay, & Flege 2001). The variables, in addition to age of onset, that show a correlation are phonological structure of the L1, amount of L1 use, amount of L2 use, amount of native-speaker input, instruction/training, length of residence, aptitude, motivation, and cognitive variables.[3] One of the first studies to do a comprehensive investigation of the various factors that influence degree of accent was Purcell and Suter (1980). The study measured 12 different predictors, ranking them in order of importance. The highest ranked factors were the L1, aptitude for oral mimicry, residency, and attitude.

Variation in accents can also be a function of the native language of the subjects. How similar the phonological system is to the L2 system can influence the degree to which the learner manifests a phonological accent; learners speaking a L1 with rules and sounds that show more correspondence to the L2 will acquire accents that are more native-like (Flege, Bohn, & Jang 1997; Purcell & Suter 1980). In fact, Purcell and Suter (1980) found this to be the most important predictor of degree of accent.

Continued use of the L1 has been investigated in many studies conducted by Flege and his associates (this is also discussed in Hansen Edwards, Chapter 9, this volume). Results indicate that variation in the amount of L1 use does not influence L2 pronunciation ability in late learners, but does affect the degree of foreign accent, if any, in early learners (Bohn & Flege 1992; Flege, Frieda & Nozawa 1997; Flege & MacKay 2004; Flege, MacKay, & Meador 1999; Flege, MacKay, & Piske 2002; Flege, Schirru, & MacKay 2003; Guion, Flege, & Loftin 2000). With early learners, results are conflicting. Flege, Frieda, and Nozawa (1997) found that all early learners had some degree of detectable accent, with stronger accents in those who used the L1 often. On the other hand, Flege, MacKay, and Meador (1999) ascertained that neither low nor high use early learners differed significantly from native speakers in both production and perception of speech sounds. In contrast, Flege and MacKay (2004) found that early learners who seldom used the L1 did not differ from native speakers, whereas those who used it often did. Flege, MacKay, and Piske (2002) studied early learners and found that dominance in L1 or L2 correlated with the amount of L1 usage. Not surprisingly, the L1 dominant subjects used their native language significantly more than those who were L2 dominant. Moreover, only the L2 dominant early learners were native-like.

3. As Hansen Edwards discusses in Chapter 9, this volume, social identity is also an important factor in L2 accent.

The most interesting findings on the relationship of L1 use to pronunciation ability are reported in Flege, Schirru, and MacKay's (2003) study of vowel production by Italian learners of English. Again, only the early learners exhibited a difference in accent ability according to L1 use. However, here those with less L1 use produced L2 English vowels that were exaggerated to allow them to become more dissimilar to the equivalent Italian vowels than the actual English pronunciation. It was those early learners who used the L1 more that produced native-like L2 vowels. Much of the discrepancy in results from these studies can be attributed to differences in test design and test stimuli (see also the discussion in Chapter 9 of this issue).

Other identified variables that influenced pronunciation ability include *amount of L2 use* (Flege, Yeni-Komshian, & Liu 1999), *length of residence* (LOR) in the L2 environment (Flege, Bohn, & Jang 1997; Purcell & Suter 1980; however, Flege and Fletcher (1992) determined that LOR was significant only with those in the early stages of acquiring an L2), *target language input* (Flege & Liu 2001), *instruction or training* (Bongaerts, Planken, & Schils 1995; Elliott 1995b; Moyer 1999), *attitude* (Moyer 1999, 2004; Purcell & Suter 1980), *the cognitive variables of field independence and right hemispheric specialization* (Elliott 1995a, b), and *social identity* (see Chapter 9 by Hansen Edwards, this volume).

Overall, it appears that one of the most important individual variables in adult L2 is the learners' aptitude for accurately producing the phonology of another language (Ioup, Boustagui, El Tigi, & Moselle 1994; Novoa, Fein, & Obler 1988; Schneiderman & Desmarais 1988). Purcell and Suter (1980) list aptitude for oral mimicry as the second most important variable (after L1) in predicting pronunciation accuracy. There appears to be a perceptual ability in talented learners that differentiates them from the normal adult population. Kuhl (2000) suggests that talented adult learners may be able to circumvent the interference effects of the L1 phonological system by perceiving novel speech sounds in the same manner as infants do.

This brings us to an important concern. Some researchers reject a critical period for L2 acquisition, arguing that non-biological individual variables account for the age differences in L2 acquisition. Evidence for their position comes from the studies that correlate age differences with individual variables. However, this is not a basis for rejecting the biological explanation. The studies measuring which age-of-onset group best approximates native models do not provide information on whether there is a critical period for phonology. The fact that some adults perfect their accent better than others or that some children in the studies do not perform as well as selected high-performing adults only argues that non-biological factors can influence the degree to which an accent approaches a native-speaker norm. The true test for a critical period in L2 phonological acquisition is whether **any** late learner can be native-like in his or her accent. Once we locate seeming exceptions, we cannot rely on superficial rater judgments to see how native-like they are. The fact that they can pass for native speakers to the untrained ear is not evidence that they have acquired an accent-free L2 phonology. Only if they are tested using acoustic measurements to uncover any differences the ear cannot detect, can we be certain that there exist adult learners who are capable of acquiring a native-like phonological system in an L2. Only then have we ascertained whether there is truly an age-dependant critical period for phonology.

Methodological issues

In this section, the methodological paradigms that are employed to assess age difference in phonological acquisition are discussed. Both production and perception tasks will be evaluated. In terms of production tasks, too many of the conclusions concerning age are drawn from large scale studies employing native-speaker raters to evaluate degree of accent. These studies have two crucial weaknesses. One, it has been shown that native-speaker ratings are not reliable. For example, in many studies, some of the controls, monolingual native speakers, were judged to have a nonnative-like accent. In fact, in a study by Guanzon (2003) only seven of the 30 monolingual controls born in the U.S. were judged to be native speakers. This is undoubtedly due in part to the fact that all speakers have an accent; the raters in these studies may have been unable to clearly distinguish between native and nonnative accents. Second, the type of rater selected makes a difference. Thompson (1991) ascertained that raters who were linguistically experienced were more reliable overall than inexperienced raters but were also more lenient in their evaluations, assigning the non-native speakers higher scores than they received from the linguistically unsophisticated raters.

Additionally, cross-study comparison of results is difficult because the methodologies vary considerably. Some studies use large passages of speech, some use short utterances, while others use single phonemes. Moreover, the types of ratings vary from study to study. Some studies ask raters to use a scale of accentedness with a point system that may vary from 5 to 7 values. Others rely on a description of the values, such as "no accent," "slight accent," or "marked accent." Since studies employ different methodologies, their results may not be comparable, which can account for some of the disparate findings studies exhibit (see Munro, Chapter 7 of this volume for a detailed discussion of accent and intelligibility ratings; see also Piske, MacKay, & Flege 2001 for a more complete discussion of methodological issues).

A further methodological problem was noted by Bialystok and Hakuta (1994), who demonstrated that by varying the way subjects are arranged into groups, one can obtain a different statistical outcome. By reinterpreting data, they altered the age effects found in an existing study, indicating that the results of data-oriented age investigations are impacted by the type of statistical analysis chosen. Of course, this is a consideration in assessing all data-oriented research.

Yet, studies using native-speaking judges cannot be abandoned altogether in favor of studies relying on acoustic measurements. At the current time only a small number of acoustic properties of the speech output can be measured. Accents can be compared acoustically in terms of vowel quality and length, voice onset time, and single versus geminate consonant duration. But these may not be the features that characterize an accent as non-native. The relevant phonetic properties can be elusive and difficult to measure. Therefore, in some way, the use of native raters has an advantage, for the human "ear" is good at perceiving non-measurable subtle differences which may be what characterize one accent as different from another.

Another problem that can impact the validity of studies measuring accent is discussed by Grosjean (1997, 1998) (this issue is also discussed in Zampini, Chapter 8). He notes that bilinguals process their languages in a continuum of modes ranging from a completely monolingual mode, used when interacting with monolingual speakers of one of the languages, to a completely bilingual mode, activated when engaged with other bilingual speakers of the same languages. The level of activation may occur at any of the intermediary points on the continuum according to how much one of the languages takes precedence over the other. Studies have shown that when both languages are activated within a bilingual context, language mixing is common. Such mixing can impact production as well as perception. A problem arises, then, in the studies that rely on acoustic measurements. Many of these studies have found subjects responding with values for both L1 and L2 sounds that are intermediate to those produced by monolingual speakers of either language. Since subjects in these studies are generally aware that their bilingualism is the focus of the research they are engaged in, they may be functioning in a bilingual mode that does not reflect the phonetic ability that would be elicited in a more monolingual mode. As Munro points out in Chapter 7, this volume, a further problem with accent ratings is that rating scales and raters may also confuse accent and intelligibility, which are two separate constructs.

Directions for future research

This brings us to gaps in the research paradigms on age in L2 phonology. Almost all the age studies examine adult subjects, dividing them into groups according to the age at which they began learning the L2. There are no large studies comparing child learners when they are children with adult learners. However, there are two small studies that compare children in the process of L2 learning with adults who are learning the same language. Ioup and Tansomboon (1987) compared the acquisition of Thai by bilingual children between 2 and 4 whose dominant language was English with two L1 English groups of adult-onset learners, one consisting of absolute beginners in their second semester of leaning Thai at a university and the other of extremely fluent adult-onset learners who had been conversing in Thai for over 40 years. The study found that early stage adults mastered novel consonants more easily than novel vowel sounds, while the children were just the opposite: the vowels were easier for them than the consonants. In addition, the first aspects of phonology the children acquired were the phonetic properties of Thai tone, a feature of the phonology that even the very proficient adult learners in the study had not been able to master.

A more recent small-scale study by Baker, Trofimovich, Mack, and Flege (2002) compared native-speaking Korean children and adults in the beginning stages of learning English. Subjects had about one year of residence in the U.S. The children ranged in age from 7–9 while the adults had an age range of 20-23. Subjects were asked to identify which Korean vowels were closest to the English vowels they heard, as well as to rate each for the degree of similarity to its corresponding Korean vowel. When matches were made, both groups selected the same Korean vowels as equivalent; however, the children were significantly less likely than the adults to make such identifications, indicating that the children's L1 sound system was not yet as firmly established as the adult systems. The research gives support to Flege's (1995) Speech Learning Model, which accounts for the comparative ease with which children learn an L2 phonology by theorizing that categories in the L1 sound system strengthen as the speaker matures. The children in this study would seem to have less firmly established perceptual categories than the adults. These two studies are suggestive of the types of contrasts child and adult learners exhibit. Additional research comparing child L2 learners with adult learners would be helpful to learn more about how the actual *processes* of L2 phonological acquisition differ in children and adults.

There is an additional gap in the research paradigms. We have no longitudinal studies that follow the development of a phonological system from the initial or very early stages to the more developed stages. Such studies would give much more information on what influences the change from one stage to another. They would shed light on the processes the learners use, allowing one to better compare individual learners to one another, especially learners of different ages. The information could help us confirm hypotheses that have been suggested as the basis of the observed differences in child and adult L2 phonology.

Conclusion

There has been a great deal of recent research on the nature and degree of accent in the pronunciation of a non-primary language. The conclusion reached is that by and large both the nature of the accent and the degree to which it is manifest is influenced primarily by the age at which language acquisition began. Whether this is the result of biological factors or non age-related individual variables is still the subject of much debate; however, the conclusion reached in this review is that age is the crucial variable. Therefore, neurocognitive and/or abstract linguistic properties of maturation are the main source of the age differences that have been discussed above.

Suggested readings

Best, C. 1995. A direct realist view of cross-language speech perception. In W. Strange (ed.), *Speech perception and linguistic experience*, Baltimore: York Press, 171–206.

Brown, C. 2000. The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (ed.), *Second language acquisition and linguistic theory*, Oxford: Blackwell, 4–63.

DeKeyser, R., & Larson-Hall, J. 2005. What does the critical period really mean? In J. Kroll & A. de Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, Oxford: Oxford University Press, .

Hansen Edwards, J. G. 2008. Social factors and variation in production. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins, .

Kuhl, P. 2000. A new view of language acquisition. *Proceedings of the National Academy of Science* 97: 11850–11857.

Mack, M. 2003. The phonetic system of bilinguals. In M. Banich & M. Mack (eds.), *Mind, brain, and language*, Mahwah, NJ: Lawrence Erlbaum, 309–349.

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

Piske, T., MacKay, I., & Flege, J. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215.

References

Altenberg, E. 2005. The judgment, perception, and production of consonant clusters in a second language. *International Review of Applied Lingusitics* 43: 53–80.

Asher, J., & Garcia, R. 1969. The optimal age to learn a foreign language. *Modern Language Journal* 53: 1219–1227.

Baker, W., Trofimovich, P., Mack M., & Flege, J. 2002. The effect of perceived phonetic similarity on non-native sound learning by children and adults. In B. Skarabela, S. Fish, & A. Do (eds.), *Proceedings of the 26th Annual Boston University Conference on Language Development*, Somerville, MA: Cascadilla Press, 36–47.

Best, C. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In C. Goodman & H. Nusbaum (eds.), *The Development of speech perception: The transition from speech sounds to spoken words*, Cambridge, MA: MIT Press, 167–224.

Bialystok, E., & Hakuta, K. 1994. *In other words: The science and psychology of second-language acquisition*. New York: Basic Books.

Bohn, O., & Flege, J. 1992. The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition* 14: 131–158.

Bongaerts, T., Planken, B., & Schils, E. 1995. Can late learners attain a native accent in a foreign language? A test of the critical period hypothesis. In D. Singleton & Z. Lengyl (eds.), *The age factor in second language acquisition*, Clevedon, UK: Multilingual Matters, 30–50.

Bradlow, A. R. 2008. Training non-native language sound patterns: Lessons from training Japanese adults on the English /ɹ/ – /l/ contrast. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Bradlow, A., Akahane-Yamada, R., Pisoni, D., & Tohkura, Y. 1999. Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in speech perception and production. *Perception & Psychophysics* 61: 977–985.

Bradlow, A., & Pisoni, D. 1999. Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America* 54: 421–428.

Bradlow, A., Pisoni, D., Yamada, R., & Tohkura, Y. 1997. Training the Japanese listener 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.

Brown, C. 2000. The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (ed.), *Second language acquisition and linguistic theory*, Oxford: Blackwell, 4–63.

Broselow, E. 1984. An investigation of transfer in second language phonology. *International Review of Applied Linguistics* 22: 253–269.

Caramazza, A., Yeni-Komshian, G., Zurif, E., & Carbone, E. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America* 54: 421–428.

Chomsky, N. 1981. *Lectures on government and binding*. Dordrecht, The Netherlands: Foris.

Elliott, A. 1995a. Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation Accuracy in Spanish as a foreign language. *Modern Language Journal* 79: 365–371.

Elliott, A. 1995b. Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *Modern Language Journal* 79: 530–542.

Flege, J. 1980. Phonetic approximation in second language acquisition. *Language Learning* 30: 117–134.

Flege, J. 1991. Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America* 89: 395–411.

Flege, J. 1995. Second language speech learning theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, Baltimore, MD: York Press, 233–277.

Flege, J., Bohn, O.-S., & Jang, S. 1997. The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics* 25: 169–186.

Flege, J., & Eefting, W. 1987. Production and perception of English stops by native Spanish speakers. *Journal of Phonetics* 15: 67–83.

Flege, J., & Fletcher, K. 1992. Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America* 91: 370–389.

Flege, J., Frieda, E., & Nozawa, T. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics* 25: 169–186.

Flege, J., & Hillenbrand, J. 1987. Limits on phonetic accuracy in foreign language speech production. In G. Ioup & S. Weinberger (eds.), *Interlanguage phonology: The acquisition of a second language sound system*, Cambridge, MA: Newbury House, 176–203.

Flege J., & Liu, S. 2001. The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition* 23: 527–552.

Flege, J., & MacKay, I. 2004. Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26: 1–34.

Flege, J., MacKay, I., & Meador, D. 1999. Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America* 106: 2973–2988.

Flege, J., MacKay, I., & Piske, T. 2002. Assessing bilingual dominance. *Applied Psycholinguistics* 23: 567–598.

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

Flege, J., Munro, M., & Skelton L. 1992. Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish. *Journal of the Acoustical Society of America* 92: 128–143.

Flege, J., Schirru, C., & MacKay, I. 2003. Interaction between the native and second language phonetic subsystems. *Speech Communication* 40: 467–491.

Flege, J., Takagi, N., & Mann, V. 1996. Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *Journal of the Acoustical Society of America* 99: 1161–1173.

Flege, J., Yeni-Komshian, G., & Liu, S. 1999. Age constraints on second-language acquisition. *Journal of Phonetics* 25: 169–186.

Grosjean, F. 1997. Processing mixed language: Issues, findings, and models. In A. de Groot & J. Kroll (eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, Mahwah, NJ: Lawrence Erlbaum, 225–254.

Grosjean, F. 1998. Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1: 131–149.

Guanzon, C. 2003. Re-examining the Critical Period Hypothesis in L2 phonological fluency. Paper presented at the annual conference of the American Association of Applied Linguistics, Arlington, VA.

Guion, S., Flege, J., & Loftin, J. 2000. The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. *Journal of Phonetics* 28: 27–42.

Hansen Edwards, J. G. 2008. Social factors and variation in production. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Ioup, G. 1984. Is there a structural foreign accent? A comparison of syntactic and phonological errors in second language acquisition. *Language Learning* 34: 1–17.

Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. 1994. Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition* 16: 73–98.

Ioup, G., & Tansomboon, A. 1987. The acquisition of tone: A maturational perspective. In G. Ioup & S. Weinberger (eds.), *Interlanguage phonology: The acquisition of a second language sound system*, Cambridge, MA: Newbury House, 333–349.

Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87: B47–B57.

Kim, R. 1995. The effect of age-of-L2 onset on ultimate L2 production: The English /i-I/ distinction made by Korean speakers. *English Teaching* 50: 257–279.

Kuhl, P. 1992. Psychoacoustics and speech perception: Internal standards, perceptual anchors, and prototypes. In L. Werner & E. Rubel (eds.), *Developmental psychoacoustics*, Washington, DC: American Psychological Association, 293–332.

Kuhl, P. 2000. A new view of language acquisition. *Proceedings of the National Academy of Science* 97: 11850–11857.

Lively, S., Logan, J., & Pisoni, D. 1993. Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* 94: 1242–1255.

Lively, S., Pisoni, D., Yamada, R., Tohkura, Y., & Yamada, T. 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.

Lenneberg, E. 1967. *Biological foundations of language*. New York: John Wiley & Sons.

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

Lombardi, L. 2003. Second language data and constraints on manner: Explaining substitutions for English interdentals. *Second Language Research* 19: 225–250.

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

Mack, M. 1982. Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *Journal of the Acoustical Society of America* 21: 173–178.

Mack, M. 1987. The perception of natural and computer-generated speech by English monolinguals and German-English bilinguals. In W. Bahner, J. Schildt, & D. Viehweger (eds.), *Proceedings of the Fourteenth International Congress of Linguistics*, Berlin: Akademie-Verlag, 470–473.

Mack, M. 2003. The phonetic systems of bilinguals. In M. Banich & M. Mack (eds.), *Mind, brain, and language*, Mahwah, NJ: Lawrence Erlbaum Associates, 309–349.

Mack, M., Bott, S., & Boronat, C. 1995. Mother, I'd rather do it myself, maybe: An analysis of voice-onset time produced by early French-English bilinguals. *Idea* 8: 23–55.

MacKain, K., Best, C., & Strange, W. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics* 2: 369–390.

McCandliss, B., Fiez, J., Protopapas, A., Conway, M., & McClelland, J. 2002. Success and failure in teaching the [r]–[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective & Behavioral Neuroscience* 2: 89–108.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., & Fujimura, O. 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics* 18: 331–340.

Mochizuki, M. 1981. The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics* 9: 283–303.

Moyer, A. 1999. Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition* 21: 81–108.

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

Munro, M. 1993. Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech* 36: 39–66.

Munro, M. J. 2008. Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Munro, M., Flege, J., & MacKay, I. 1996. The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics* 17: 313–334.

Nathan, G., Anderson, W., & Budsayamongkon, B. 1987. On the acquisition of aspiration. In G. Ioup & S. Weinberger (eds.), *Interlanguage phonology: The acquisition of a second language sound system*, Cambridge, MA: Newbury House, 204–212.

Novoa, L., Fein, D., & Obler, L. 1988. Talent in foreign languages: A case study. In L. Obler & D. Fein (eds.), *The exceptional brain: The neurophychology of talent and special abilities*, New York: Guilford Press, 294–302.

Ohala, D. K. 2008. Phonological acquisition in a first language. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Oyama, S. 1976. A sensitive period of the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research* 5: 261–283.

Oyama, S. 1978. The sensitive period and comprehension of speech. *Working Papers on Bilingualism* 16: 1–17.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. 1997. A limit on behavioral plasticity in speech perception. *Cognition* 64: B9–B17.

Patowski, M. 1990. Age and accent in a second language. A reply to James Emil Flege. *Applied Linguistics* 11: 73–89.

Piske, T., MacKay, I., & Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215.

Purcell, E., & Suter, R. 1980. Predictors of pronunciation accuracy: A reexamination. *Language Learning* 30: 271–287.

Rochet, B. 1995. Perception and production of second-language speech sounds by adults. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language speech research*, Timonium, MD: York Press, 379–410.

Schneiderman, E., & Desmarais, C. 1988. The talented learner: Some preliminary findings. *Second Language Research* 4: 91–109.

Scovel, T. 1988. *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. Cambridge, MA: Newbury House.

Sheldon, A., & Strange, W. 1982. The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3: 243–261.

Strange, W., & Dittmann, S. 1984. Effects of discrimination training on the perception of /r–l/ by Japanese adults learning English. *Perception & Psychophysics* 36: 131–145.

Strange, W. & Shafer, V. L. 2008. Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

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

Thornburgh, D., & Ryalls, J. 1998. Voice onset time in Spanish-English bilinguals: Early versus late learners of English. *Journal of Communication Disorders* 31: 215–229.

Weinberger, S. 1997. Minimal segments in second language phonology. In A. James & J. Leather (eds.), *Second language speech: Structure and process*, Berlin: Mouton, 263–312.

Zampini, M. L. 2008. L2 speech production research: Findings, issues, and advances. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

CHAPTER 3

Transfer in second language phonology

A review

Roy C. Major
Arizona State University

Introduction

This chapter critically examines research on transfer in second language (L2) phonology over the past half century. Because of the common threads L2 phonology has shared with learning theory and general issues in second language acquisition (SLA), the chapter begins with these two areas. These sections then serve to place L2 phonology within a historical perspective within general issues in SLA and learning theory. I have surveyed a large number of studies but have not detailed many of them. My purpose here is not only to characterize the major findings and issues in transfer research but also to present the readers with a large number of references that are directly or indirectly relevant to research on transfer.

Transfer in learning: Early research

For at least the past 80 years psychologists have been investigating transfer in learning (Ausubel 1963, 1967; Ausubel & Robinson 1969; Ausubel, Novak, & Hanesian 1978; Bruce 1933; Bugelski 1942; Cheng 1929; Duncan 1958; Gagné 1977; McGeogh 1942; Osgood 1946, 1949; Schultz 1960; Travers 1977). Some of the early work within the behaviorist perspective includes the notions of *proactive* and *retroactive interference* (Osgood 1946). Ausubel (1963) went so far as to say that all learning involves some kind of transfer. He also hinted that there are necessary conditions for transfer to take place; the phenomenon has to have "relevant aspects" and be "organically relatable" to the new experience (p. 28). Earlier, Osgood had used the terms "meaningful similarity" (1946) for this same issue. Gagné (1965) also described some conditions of transfer to the new activity "which incorporates these previously acquired capabilities." (p. 129). This idea is also seen in mathematical learning theory (Atkinson, Bower, & Crothers 1965), which dealt with functional connections and paired associations in Markov chains (which are precursors to connectionism. Also see Ellis 1996). Even recent work using cognitive theory and L2 instruction has discussed connections between old and new information (e.g., Neuner 2002). In sum, much of this early research attempted to specify the conditions necessary for transfer to occur. Later, this issue was to become important in SLA research, for example, in Andersen's Transfer to Somewhere Principle (1983) and Kellerman's Transfer to Nowhere Principle (1995), discussed below.

Transfer in second language acquisition

Early research on transfer

The investigation of transfer in L2 acquisition began long before Contrastive Analysis, which started in earnest in the 1950s. Odlin (1989) noted that controversies concerning the role of transfer in historical change, including diffusion and pidgins and creoles, dates back to the nineteenth century (see Müller 1861/1965). Trubetzkoy (1939/1958) claimed that the "sieve" of the L1 "filters" one's perception in an L2. Weinreich (1953), in his classic work on languages in contact, used the older term interference to detail the different linguistic levels of transfer, including phonological, morphological, syntactic, and lexical.

Contrastive analysis

Although there was much diachronic work in the nineteenth century on transfer, by the mid twentieth century expositions on SLA transfer using synchronic data usually focused on Contrastive Analysis (CA), which reached its heyday in the 1960s. CA is most often associated with language teaching, for example the work of Fries (1945) and Lado's landmark work (1957), which were typical of the belief that transfer accounted for and predicted all errors. This perspective is exemplified by the well-known works of Stockwell and Bowen (1965) and Stockwell, Bowen, and Martin (1965), contrastive analyses of Spanish and English, which also included elaborate hierarchies of difficulty.

The fundamental claims of CA are that transfer explains all errors and on this basis it is possible to predict all errors. Soon after CA became popular the predictive power of CA was criticized when it was pointed out that many learners did not make the predicted errors (e.g., some German learners of English have no difficulty with the /r/). In order to address this shortcoming, Wardhaugh (1970) introduced the *strong* versus the *weak* versions of CA: The strong version predicted errors; the weak version explained errors after the fact. The definition of prediction is very relevant in discounting the strong version. If prediction is defined as an absolute occurrence or nonoccurrence of phenomena for all individuals, then CA is can easily be falsified. However, if prediction is defined in the probabilistic sense then CA is alive and well today. A hypothetical example should be sufficient: If one has the opportunity to teach English to native speakers of Japanese who have never studied English, it can be predicted that many of them (probably most) will experience difficulties with English liquids and syllable structures. For example, because Japanese has only one liquid and no closed syllables (except word internal geminates) MacDonald's restaurant is pronounced [makudonaɾodo] (for further details, see Zampini, Chapter 8, in this volume).

Although CA seemed to be able to explain errors after the fact and is predictive in the probabilistic sense, it did not predict which phenomena should be more difficult than others. However, Oller and Ziahosseiny's (1970) *moderate version* of CA claimed that similar structures in L1 and L2 cause more difficulty than dissimilar structures. Although their study was based on spelling errors (speakers whose native languages did not use the Roman alphabet made fewer spelling mistakes than speakers whose languages did), their hypothesis is generalizable to other phenomena: Similar phenomena are more difficult to learn than dissimilar phenomena.

However, even with the weak and moderate versions, CA was not spared of the criticism of the claim that all errors were due to transfer. Selinker's seminal work on interlanguage and fossilization (1972) pointed out that an interlanguage system was the result of many factors, transfer merely being one of them. With Selinker's and other's work, there became a growing awareness of the existence of non-transfer errors – errors which were due to universals or developmental factors and which were similar or identical to those occurring in L1 acquisition. During the 1970s there was a diminished interest in transfer because of the well-known shortcomings of CA. However, from the 1980s and continuing into the present there has been a resurgent interest and acknowledgment of the importance of transfer in SLA (Gass & Selinker 1983, 1992; Han 2004; Han & Odlin 2006; Kellerman & Sharwood Smith 1986; Odlin 1989, 2003).

Conditions on transfer

Early psychologists noted that there were conditions that have to be present in order for transfer to occur. Gagné (1965) also described conditions of transfer to the new activity "which incorporates these previously acquired capabilities." (p. 129). For Ausubel (1963) the phenomenon had to have "relevant aspects" and to be "organically relatable" to the new experience; even earlier Osgood had discussed "meaningful similarity" (1946). Recall the earlier discussion of Oller and Ziahosseiny's (1970) *moderate version* of CA: Similar structures in L1 and L2 cause more difficulty than dissimilar ones. The implication of Oller and Ziahosseiny's hypothesis is that if patterns are very dissimilar confusion would not result and transfer would not occur, though they did not mention this specifically. Even though the *moderate version* seemed to address one aspect of the problem of predictability, it did not address the issue of what in the L1 would be transferred to the L2 and what would not.

If there are no similar or corresponding structures in L1 and L2 or if the L1 and L2 forms are different, then the most general version of CA predicts that L2 learners will have difficulty with these L2 forms (e.g., the L1 has no tense and the L2 has tense or L1 has tense but different tenses from L2). However, Andersen's Transfer to Somewhere Principle (1983) specified conditions under which transfer can and cannot operate. He claimed there have to be corresponding structures in the L1 and L2 in order for transfer to operate and that "a grammatical form or structure will occur consistently and to a significant extent in the interlanguage as a result of transfer if and only if there already exists within the L2 input the potential for (mis-) generalization from the input to produce the same form or structure" (p. 178. See Gass, p. 125). In contrast, Kellerman took issue with Andersen with his Transfer to Nowhere Principle, stating that "there can be transfer which is not licensed by similarity to the L2 and the way the L2 works may very largely go unheeded." (1995: 137; cited in Odlin 2003: 456). In phonology, Andersen's Principle would predict that English speakers would not transfer English sounds when they tried to imitate clicks; however, Kellerman' Principle would predict they might.

Closely related to the issue of conditions for transfer is the UG (Universal Grammar) accessibility issue. If UG is not accessible this implies that L1 transfer would be the main or perhaps the only factor involved in SLA. If UG is accessible then transfer may or may not operate, depending how accessible UG is. The Fundamental Difference Hypothesis of Bley-Vroman (1989, 1990) claims that L2 learners do not have access to UG; they only know about UG through their NL. In contrast, over the years Lydia White (1989, 2003a, 2003b) has claimed some degree of UG accessibility.

Researchers have also disagreed as to what constitutes the initial state for L2 acquisition, which in turn relates to the UG accessibility issue. Hypotheses that claim the initial state is L1 (as most researchers do) – that transfer necessarily occurs at the beginning stages, are the Full Transfer Full Access Hypothesis of Schwartz and Sprouse (1994, 1996; Sprouse & Schwartz 1998), the Valueless Features Hypothesis of Eubank (1993/1994, 1994, 1996), and the Minimal Trees Hypothesis (Vainikka & Young-Scholten 1994, 1996a, 1996b). In contrast, there are hypotheses that claim SLA proceeds from UG, similar to L1 acquisition, and SLA can occur without transfer: the Full Access Hypothesis (Epstein, Flynn, & Martohardjono

1996, 1998). A modification of the Full Access Hypothesis is the Failed Functional Features Hypothesis (Hawkins & Chan 1997), claiming that L2 learners are limited to L2 features occurring in their L1; if they do not then these structures cannot be learned. For studies supporting and refuting these various positions see Bhatt and Hancin-Bhatt (2002), Herschensohn and Stevenson (2003), Kozlowska-Macgregor, and Leung (2003), Leung (2001), Schwartz & Eubank (1996), White (2002), White, Valenzuela, Kozlowska-Macgregor, and Leung (2003), Whong-Barr and Schwartz (2002).

Although there has been much research on the relationship between transfer and UG accessibility there are still many problematic areas. White herself admits "…we do not know why in some cases the effects of the L1 are so fleeting as to be barely noticeable even in early stages" (2003a: 269). The problematic issue is shared by researchers on pidgins and creoles. According to Siegel (2003: 199) the substrate hypothesis cannot explain "why some substrate features end up in a pidgin or creole, whereas others do not." Thus, many of the conditions on transfer remain a mystery.

Transfer in second language phonology

Early research on transfer in L2 phonology predates Contrastive Analysis. Weinreich (1953) described various types of sound transfer, which included: *sound substitution* (a learner uses the nearest L1 equivalent in the L2, e.g., English [ɹ] for Spanish [r] in L1 English/L2 Spanish), *phonological processes* (a learner uses the L1 allophonic variant that does not occur in the same environment in the L2, e.g., clear [l] in coda position for velarized [ɫ] in L1 French/L2 English), *underdifferentiation* (the L2 has distinctions that the L1 does not; e.g., when two sounds are allophones in the L1 but separate phonemes in the L2, as in [d] and [ð] in L1 Spanish/L2 English), *overdifferentiation* (the L1 has distinctions that the L2 does not, e.g., two sounds are separate phonemes in the L1 but are allophones in the L2, as in [d] and [ð] in L1 English/L2 Spanish), *reinterpretation of distinctions* (reinterpreting secondary or concomitant features as primary or distinctive features, e.g., in L1 German/L2 English a learner interpreting English tense/lax distinctions as long and short distinctions), *phonotactic interference* (making the syllable structure in the L2 conform to the L1 syllable structure, e.g., *pic*[i] *nic*[i] in L1 Portuguese/L2 English), and *prosodic interference* (e.g., producing falling intonation in utterance final words in L1 English/L2 Mandarin, regardless of the tone in Mandarin). Haugen's (1956) work on bilingualism in the Americas employed different terms for some of Weinreich's categories: *simple identification* replaced *sound substitution*, *divergent* replaced *underdifferentiation* and *convergent* replaced *overdifferentiation*.

Although CA has long since been (1962) postulated a taxonomy of error types based on contrasting German and English: (a) phonemic errors, (b) phonetic errors, (c) allophonic errors, and (d) distributional errors (phonotactics). Note that these categories are similar to the classifications proposed earlier by Weinreich (1953) and Haugen (1956). Brière, in oft-cited works (1966, 1968), hypothesized a model of *proactive interference* from a behaviorist perspective based on L1 habits, claiming stimuli identical in L1 and L2 yield correct responses stimuli which in L1 and L2 are different cause learning difficulties. Other works on CA include Redard (1973), who contrasted Italian with eleven other languages, Paik (1977) on L1 Korean/L2 English, Soudak (1977) on L1 Czech/L2 English, Tomaszczyk (1980) on the Polish of Polish Americans, and Anan (1981) on L1 Japanese/L2 French.

Although CA has long since been abandoned as a main framework for research, transfer (often with its interaction with universals) continues to be the focus of much work, ranging from the segmental to prosodic levels and employing a variety of theoretical frameworks. This copious research includes segmentals (Hancin-Bhatt 1994; Hung & Man on Hong Kong English 2002; J-E. Kim and Silva on Korean English 2003; Marghany on Egyptian English 2002; Wang & Geva on Cantonese English 2003; Zampini on Spanish English 1996), syllable structure (Basson 1986; Broselow 1984; Eckman & Iverson 1994; Flores & Rodrigues 1994; Seubsunk 2000), metrical structure (Archibald 1992), rhythm (Sajavaara & Dufva 2001; Wenk 1986; Zsiga 2003), connectionism (Ellis 1996 [for a response to Ellis, see Ioup 1996; Major 1996;]; Shirai 1992), and dialects (Munro, Derwing, & Flege 1999; Wolfram, Childs, & Torbert 2000). Even studies from an OT (Optimality Theory) framework acknowledge the importance of transfer, which in OT terms can be stated as L1 rankings (Broselow, Chen, & Wang 1998, Hancin-Bhatt & Bhatt 1997; Hancin-Bhatt 2000; H-K. Kim 2001; Lombardi 2003).

Loan phonology phenomena are mostly attributable to transfer (Broselow 2000). Loan phonology can be considered as a culturally induced transfer or forced transfer, since native speakers of the borrowing language usually pronounce loan words as if they were native words in their L1, lest they sound too snobbish and affected. Though an announcer on a classical music station may say *Ba*[x] for *Bach* on the radio, when not on the radio he or she would probably not say *Ma*[x]*1* for *Mach1* or *ta*[x]*ometer* for *tachometer*.

Surface versus abstract transfer

It seems that phoneticians and phonologists forever have been at odds with each other over what is a legitimate level of investigation. Phoneticians generally favor more surface phenomena, while phonologists generally favor more abstract phenomena. These differences in points of views have carried over to research in SLA. What is transferred? Is it surface phenomena or abstract features and principles?

Is stylistic variation in the L1 part of abstract competence or is it merely a matter of performance? Is L1 variation transferred to the L2?

There are those who conveniently ignore stylistic variation, claiming it is part of performance, not competence, and base their analyses on abstract general descriptions of the L1s (which usually means a formal style) and supposedly on this basis they can determine what constitutes L1 transfer. However, in doing so, they potentially miss transfer phenomena in other speech styles. If the L2 speech sample includes anything other than citation forms, such as reading word lists, ignorance of L1 running speech characteristics can lead to some erroneous conclusions about transfer. The oft-cited vowel epenthesis rule in Japanese serves an example. A citation reading of *spy* may result in [supay] but in running speech it will most likely be [spay] because of vowel devoicing and deletion between voiceless obstruents, a process that occurs in running speech in Japanese. The two different forms (one nonnative and one native) are both the result of transfer. In a similar fashion, a Brazilian speaker of English may say *tops* as [tapis] in citation but [taps] in casual speech. These examples illustrate that in some cases it appears that surface properties and processes can be more important than abstract phonological characteristics.

Generally Flege's work shows a preference toward surface phonetic categories rather than abstract phonological categories. This point of view is quite clear in H1 of his Speech Learning Model (1995: 239): "Sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level." Halle, Best, and Levitt (1999) also argued for the importance of surface characteristics in perception. They examined identification and discrimination of American English continua of /r/–/l/, /w/–/r/, and /w/–/y/ (yod) by French listeners. Although French has an /r/–/l/ contrast, the French listeners had difficulty with /r/, which they tended to assimilate as "/w/-like." The researchers concluded that the detailed phonetic properties of the L1 and L2 accounted for results rather than solely the abstract phonological properties.

Arguments for a more abstract approach are numerous. Because orthography itself is an abstract representation, errors due to orthographic influence tend to favor a more abstract approach. Although English has an allophonic flap [ɾ], similar to the Spanish single *r*, English speakers typically pronounce the Spanish *r* as an English *r*, when in fact if these learners thought of it as an intervocalic *t* or *d* (e.g., *pot o' gold*), they would be produce it more accurately. (For studies on orthography see Wang & Geva 2003; Seubsunk 2000). Perception is also important in predicting what substitutions occur. In general, an L2 learner uses an L1 sound that is perceptually close in the L2 sound. However, in some cases the occurrence of one L1 substitution over another cannot be explained on the basis of surface perceptual and acoustic characteristics. Hindi speakers of English tend to use their retroflex [ʈ] and [ɖ] rather than the dental [t̪] and [d̪]. To English speakers, the dentals seem closer perceptually to English alveolar /t/ and /θ/, but apparently not to Hindi speakers. Perhaps the explanation can be found by investigating the whole abstract systems. This is precisely what Lombardi (2003) did in attempting to explain why Russians substitute [t] for English [θ], while Japanese substitute [s] for [θ].

When discussing general issues of levels of representation, Gass and Selinker (2001) suggested that "...one can imagine that transfer could occur not just on the basis of surface facts, but also on the basis of underlying structures" (p. 186). Models and theories involving UG are abstract, for example, the Full Transfer Full Access Hypothesis of Schwartz and Sprouse (1994, 1996; Sprouse & Schwartz 1998. Also see Steele (2001) and the Failed Functional Features Hypothesis (Hawkins & Chan 1997), discussed earlier in this chapter. Features have been prominent in phonology since the Prague school (Trubetzkoy (1939/1958; Jakobson 1941). Since features are abstractions of one or more surface phonetic characteristics, most analyses involving features can be said to favor the abstract approach. The Feature Competition Model of Hancin-Bhatt (1994, Hancin-Bhatt & Govindjee 1999) claimed that features that are used more frequently in the L1 figure more prominently in L2 perception. Hale and Reiss (2003) argued that the Subset Principle should be reconceptualized in features rather than segments. Underspecification Theory and Feature Geometry have been used by Brown (1998, 2000), who claims that the categories of the L1 constrain which L2 sounds can be accurately perceived and produced. Although this is a highly abstract model, the general claims recall Flege's Speech Learning Model (1995): New sounds are more easily learned than similar sounds. However, if the new sounds contain features that the L1 does not, then Brown's claims would seem to make opposite predictions to Flege's.

In a study of perception, Curtin, Goad, and Pater (1998) found English speakers acquired Thai voicing before aspiration. They attributed their results to the fact that aspiration in English is a surface, rather than lexical representation, as voicing is. However, in a follow up study Pater (2003) found the opposite results: Aspiration discrimination was more accurate than voicing. Thus, the contradictory results of these two studies leave the abstract/surface issue unresolved.

At times, debates over surface/abstract issues have seemed to disintegrate into dogmatic argumentation. A strongly stated position against the relevance of surface phenomena for phonology (although they do not deal specifically with L2 phonology) is Hale and Reiss (2000) who claim "phonological theory, as part of cognitive science, must be divorced from issues of phonetic substance to be able to categorize the computationally possible phonologies in universal grammar. . .generalizations based on phonetic patterns are irrelevant to phonology" (p. 157). If phonology is the study of sound patterns then it is curious that Hale and Reiss want to divorce the substance of sounds from it. In summary, the abstract/surface issue remains unresolved, and because this issue has been debated for decades, it is unlikely that any reconciliation will occur in the near future.

Age and experience, L1 and L2 use[1]

It is widely observed that the native languages of adult immigrants can often be easily identified when they speak their L2s, an obvious instance of transfer (e.g., a German or French accent can persist even years after immigration). In contrast, young immigrant children quickly acquire native accents of the dominant language of their newly adopted country. This observation has taken the form of the Critical Period Hypothesis, which claims a person must be exposed to a language (L1 or L2) during a specific period of time in order to acquire it natively (see Birdsong 1999; Long 1990; Moyer 2004; Scovel 1969, 1988, 2000). In support of the Critical Period (sometimes called the sensitive period), Oyama's now classic study (1976) found that in 60 Italian-born immigrants to the United States that age of arrival was a strong predictor of foreign accent but length of residence was not.

In addition to age, experience has been extensively investigated recently, that is, the length of exposure to the L2, in addition to L1 and L2 use. As one would expect, the more L2 use and the less L1 use, the less foreign accent; conversely, the less L2 use and the more L1 use the more foreign accent (Atkey 2002; Flege, Schirru, C., & MacKay 2003; Flege & MacKay 2004; Guion, Flege, & Loftin 2000; Moyer 2004; Piske, MacKay, & Flege 2001, but see Oyama 1976 discussed above). Even slips of the tongue can depend on experience. Poulisse (1999) found overall that 30 percent of the slips attributable to the L1 but the percent was related to proficiency.

Similarity

It is conceivable that all types of transfer in L2 phonology are correlated with age and experience. The previous studies cited above dealt with quantitative differences, taking into account age and experience; however, within a particular group of speakers (e.g., same age of exposure and same experience) these studies did not address the question of which phenomena are more susceptible to transfer and which are not. Research on similarities between L1 and L2 has addressed this issue (for a review L2 phonetic research on similarity, see Bohn 2002).

CA's moderate version, the claim that similar phenomena are harder to learn than dissimilar phenomena, spurred widespread research in L2 phonology, perhaps because in phonology (unlike some other linguistic levels, e.g., syntax and

1. See Ioup (this volume) for an extensive discussion of the Critical Period Hypothesis.

discourse), similarities are often easier to define. Although there are problematic cases, there are many obvious undebatable instances. For example, it would be hard to argue against the notion that French /ɛ/ is more similar to English /ɛ/ than it is to /æ/ or that even on the basis of a classical structuralist description of phonemic inventories that Spanish /s/ is more similar to English /s/ than it is to English /ʃ/.

A great deal of research has demonstrated that similar sounds tend to be more difficult than dissimilar sounds. The reason seems to be that because that the larger the differences are, the more easily they tend to be noticed; therefore, learning is more likely to take place. In contrast, minimal differences often go unnoticed, resulting in non-learning, that is, transfer persists. Thus, when speaking French, an English speaker may use the English alveolar aspirated [tʰ] rather than the French unaspirated dental [t̪] because these differences are minimal and are not noticed by the learner. Yet, the same speaker more likely will notice that the French and English *r*s are different and may immediately start making non-English substitutions for French *r*. In these two examples, it would seem that although transfer may occur for both *r* and *t*, transfer would be more likely to occur and persist for *t*.

Psychologists have shown that transfer operates when there are relevant phenomena to transfer. Ausubel, Novak, and Hanesian (1978, discussed above) claimed that past experience has an "impact on relevant properties of cognitive structure" (p. 165), or in other words, transfer operates. The "relevant properties" are crucial for transfer to occur because when the two phenomena are very different, there is very little that can be transferred. This observation was captured by Andersen's (1983) Transfer to Somewhere Principle, cited previously. In contrast, Kellerman's Transfer to Nowhere Principle (also previously cited) claims transfer can occur "which is not licensed by similarity to the L2..." (1995: 137). Crucial to the nowhere/somewhere controversy is the definition of similarity, which in many cases is very hard to define and very seldom defined or even mentioned by those involved in the debate. Transferring English /s/ to Spanish /s/ seems more likely than transferring English /t/ or /k/ to a Zulu voiceless palato-alveolar click. However, there are many less obvious cases. For example, there is probably not an obvious L1 English consonant that would be substituted for the L2 Arabic voiced pharyngeal fricative /ʕ/.

The transfer to somewhere/nowhere controversy is closely related to the similarity issue because "somewhere" implies somewhere similar and "nowhere" implies there is nowhere similar enough for transfer to occur. Further complicating these issues is the observation that different experimental conditions for the same phenomena can produce different results. In teaching phonetics classes over the years I have produced various clicks and have asked students to reproduce them. Hardly ever have they produced an English consonant; instead they have produced some click like sound, though not necessarily the one I had produced. Thus, the Transfer to Somewhere Principle seemed to operate, since there was no obvious similar English consonant. Yet, recently I asked a class of native speakers of English to write the closest English consonant when I produced nonce words with a voiceless alveolar lateral click and a voiceless palatal click. The results were varied. The majority wrote the English consonant *t* or *d* but some were unable to do so, instead writing a question mark. In a similar vain, when most English speakers pronounce the first sound in the word *Xhosa* they use an English consonant instead of a click (in native Xhosa it is a voiceless alveolar lateral click). These results suggest that, depending on the conditions, both Andersen and Kellerman can be correct. Thus, this issue remains unresolved.

Wode's extensive early work on similarity (1977, 1978, 1983a, 1983b) claimed that phonetic similarity was "the basic issue" (1977: 214). He also claimed (1983a) that L1 transfer operates only when "crucial similarity measures" (p. 180) occur between L1 and L2 and that they meet "specifiable similarity requirements" (p. 185). He claimed L2 phenomena not meeting similarity requirements are acquired with patterns that characterize L1 acquisition, for example, German speakers of English using [w] for English /r/, rather than the L1 substitution of German /R/.

Two studies on Swabian German and English also show the importance of similarity. Young-Scholten (1985) found German second graders (whose teacher spoke Swabian German) made errors in phonology and morphology that depended on similarity; she also claimed that errors due to transfer would persist if they are due to similarity. James (1983) found that similarity was correlated with the differences in the amount of Swabian versus standard German in the speakers' English.

Flege probably has done more research on similarity than any other researcher. His "equivalence classification" is central to his Speech Learning Model (1992, 1995). He claimed "equivalent" or similar sounds are difficult to acquire because a speaker perceives them as equivalent to those in the L1; however, "new" (dissimilar or different) sounds are easier to acquire because there are salient differences. What his model implies (although not explicitly stated) is that transfer persists more for similar sounds than for dissimilar sounds. Even before he formalized The Speech Learning Model, his work supported what would eventually be one of the major claims of the model. Flege (1987a) found that advanced learners of French (L1 English) produced /ü/ authentically (the "new" sound) but produced /u/ unauthentically. Other research supports his claims (Flege 1987b, 1990, 1993). Bohn and Flege (1992) found that advanced German learners of English did not produce the similar English sounds /i ɪ ɛ/ authentically because of "equivalence classification." In contrast, they produced the dissimilar sound /æ/ authentically.

Although most research shows similar phenomena are more difficult to learn than dissimilar phenomena, there are exceptions. Bohn and Flege's (1992) study of German speakers of English revealed that some speakers performed better with the similar sound. Major (1987b) found that as speakers' overall native-like accent improved their production of the dissimilar sound improved but their production of the similar sound became progressively worse. Major and Kim (1996) claimed that the concept of "difficulty" is the wrong concept; rather rate is the more relevant concept. The Similarity Differential Rate Hypothesis (SDRH) claims that dissimilar phenomena are acquired at faster rates than similar phenomena. In support of the SDRH their study of Korean learners of English showed that the similar sound /ǰ/ was produced more accurately by both beginning and advanced students than the dissimilar sound /z/, but the rate of acquisition for the dissimilar sound was faster than for the similar sound (which actually was worse for advanced learners compared to beginning learners). Major (1997) analyzed data from five other studies (that were not originally designed to test the SDRH); the data supported the SDRH.

The Ontogeny Phylogeny Model (OPM, Major 2001; a revision of the OM) claims that chronologically L2 acquisition increases, L1 transfer decreases, and universals increase and then decrease (in this model interlanguage is composed of elements of L1 [transfer], L2 [acquired], and universals [including UG]). The Similarity Corollary of the OPM further captures the generalizations of a number of studies and observations concerning similarity. It claims for phenomena that are similar in L1 and L2 that L2 acquisition proceeds slowly, transfer persists; consequently the role of universals is relatively small, compared to "normal" (phenomena that are neither similar nor marked). This is because the components of interlanguage, L1 transfer, L2 forms, and U (universals), have to add up to 100 percent.

The research on similarity all seems to point to the same conclusion: The more similar the phenomena the more likely transfer will operate; however, what constitutes similar is not always clear-cut. Back in 1981 Wode pointed out the lack of a good definition of similarity requirements; twenty-five years later the situation does not seem to have changed. For example, is the Korean liquid more similar to English /r/ or /l/? The evidence is mixed depending on which criterion one deems as most important. Criteria can include acoustic, articulatory, perceptual factors, as well as NS and NNS intuitions, and even orthographic evidence. However, even though the importance of various criteria differ for different researchers, in order to evaluate and compare different studies, a more rigorous and universally agreed upon definition of similarity would seem necessary. For further discussion of the various definitions of similarity see Strange and Shafer (Chapter 6 this volume).

Perception[2]

Implicit in much of the discussion on similarity are assumptions on perception. One of the most basic assumptions is that if L2 learners perceive the L2 in terms of their L1s and thus cannot perceive L2 differences that are not made in their L1, then they will be unable to produce them (cf. Trubetzkoy 1939/1958, cited above). However, there are isolated cases where L2 learners can produce differences correctly even though they cannot hear these differences. Sheldon and Strange (1982) found that for /r/ and /l/, Japanese learners of English performed better in production than in perception. There are other isolated cases where production is better than perception but the usual scenario is that perception is better than production (and in all normal L1 acquisition). Perhaps the reason for the rare instance when production is better than perception is that most L2 learners who have been participants in research are literate and have had instruction in producing contrasts that they may not have been able to perceive (cf. deaf speakers can produce contrasts they obviously cannot hear). Consequently orthographic cues may have aided production. One of the reasons that L2 learners cannot perceive of differences in the L2 is that their perception is governed by their L1s; in other words, they transfer their L1 perceptual systems when hearing the L2s. However, before L1 acquisition is complete, humans have much more acute perceptual abilities in discriminating human speech sounds.

Two well-known models of speech perception are Flege's (1995) Speech Learning Model (SLM) and Best's (1995) Perceptual Assimilation Model (PAM), both of which are discussed in detail by Strange and Shafer (this volume). Among its claims, the SLM proposes that similar sounds will be difficult to perceive and produce because of *equivalence classification*, whereas the "new" sounds will easier to perceive and to produce. Best's PAM is similar to Flege's SLM (for a critique of these two models, see Markham 1997.) Best claims that non-native segments tend to be perceived according to their similarities to native segmental "constellations" (p. 193) that are close phonologically. Accordingly, sounds may (a) be assimilated to a native category, (b) be assimilated as an uncategorizable speech sound, giving rise to a new category, or (c) not be assimilated to speech and not heard as a speech sound.

Interaction of transfer and universals

General universal principles

Although the focus of this chapter is transfer, the role of universals is also relevant to this discussion because universals interact with transfer. When L2 acquisition

2. See Strange and Shafer (this volume) for an extensive discussion on perception.

does not result in native-like mastery, nonnative substitutions are necessarily due to transfer or universals, the proportion of which varies from phenomenon to phenomenon and from learner to learner. Thus, if transfer does not operate, universals must necessarily operate and vice versa. An understanding of one informs our understanding of the other.

Among the universals of language (in L1 and L2) are UG, learning principles (see the *projection problem*, Baker 1979), markedness considerations (Greenberg 1966, 1978), rules, processes, constraints (Prince & Smolensky 1993, 1997, 2004), and stylistic universals (Bayley & Preston 1996). These universals are illustrated in the following examples: L2 learners acquires voiced obstruents in initial position before final position because of markedness, L2 learners may exaggerate the pronunciation of American English /r/ because of hypercorrection, an L2 learner whose L1 does not contrast /f/ and /p/ will tend to produce the contrast more accurately in a word list than in conversation. These examples, also occurring in L1 acquisition, are the result of universals, not L1 transfer.

Over the past 35 years L2 studies have demonstrated the presence of L2 substitutions that could not be attributed to the L1 (Benson 1988; Dreasher & Anderson-Hsieh 1990; James 1986, 1988, 1989, 1996; Leather 1983, 1987; Musau 1993; Nemser 1971; Pennington 1992; Piper 1984; Williams 1979; Wode 1980). These substitutions were said to be the result of developmental processes (which today are simply referred to as substitutions due to universals). An oft-cited example is final obstruent devoicing (Altenberg & Vago 1983; Edge 1991; Flege & Davidian 1984; Hodne 1985; Riney 1989), which also can depend on the height of the preceding vowel (Yavaş 1997). Hecht and Mulford (1982) claimed that when both transfer and developmental factors produce the same substitution (e.g., obstruent devoicing is a developmental process and an L1 process in German), the substitution will persist longer. Their claims, however, do not predict which factor, developmental or transfer, would be more influencial for other instances because the conditions for transfer were not part of their predictions.

Prior to 1987, even though researchers had acknowledged the existence of both transfer and developmental factors for nearly 20 years, there had been no model or theory explicitly describing the interaction. The Ontogeny Model (OM) is explicit about this interaction (Major 1987a). The OM claims that over time transfer processes decrease; concurrently, developmental processes are infrequent at first, later they increase, and then still later they decrease. Although this model claimed an interaction, like Hecht and Mulford (1982), it did not postulate which factor, developmental or transfer, would be more influential.

More recent research has also shown an interaction of universal principles and transfer. Waniek-Kimczak (2002) found Polish learners of English assigned stress using complex strategies that could not be simply attributed to transfer (penultimate stress is the default setting in Polish). The assignment of stress seemed to be quantity sensitive: Speakers tend to stress the long vowels and diphthongs (which is a universal tendency in languages). She concluded that the learners computed stress rather than storing it lexically. Another study of stress involving Spanish and Nigerian speakers of English (Peng & Ann 2002) also found speakers to be quantity sensitive: Both sets of speakers tended to stress diphthongs and long vowels in words where the stress in NSs of English falls elsewhere (e.g., *supervisor*, *educate*).

Zampini (1996) studied voiced stop spirantization of /b d g ð/ in the English of NSs of Spanish and found that /d/ was spirantized the least often, which she attributed to the phonemic value of English /ð/ ([d] and [ð] are allophones in Spanish). She noted that an explanation based purely on L1 transfer would not predict these results. Zampini (1997) further used prosodic principles to argue that the spirantization in Spanish is in the prosodic domain of the intonational phrase. She found her participants (Spanish-speaking learners of English) spirantized most frequently word-internally and that most of the word initial spirantizations occurred in clitic groups. She argued that parameters are reset in stages, beginning with the most restrictive setting. Barlow (2002), using data from a larger study of bilingual Spanish-English children (ages 2–4) argued that the underlying phonemes in Spanish are voiced spirants but also that they are approximants and that the stops are derived by fortition. From this she made the prediction that Spanish learners of English would first distinguish English /d/ from /ð/ in intervocalic position because the Spanish transfer rule of fortition does not apply. Although this is prediction was based on transfer, she based her argument on abstract theoretical principles, rather than surface transfer facts.

Weinreich (1953) categorized one type of sound interference as *underdifferentiation* (sounds are allophones in L1 but separate phonemes in L2), and it has commonly been observed that this type of acquisition is one of the most difficult. Calling it the *allophonic split*, Eckman, Elreyes, and Iverson (2001, 2003) revisited this issue, employing general principles of mainstream phonology in their study involving NSs of Korean and Spanish learning English (2003). The contrasts they studied (which are allophones in the NLs of the participants) were /d/ and /ð/ for NSs of Spanish and /s/ and/ʃ/ for NSs of Korean. Although their article dealt with several phenomena (including deflected contrast and hypercontrast), their innovative contribution employed the Derived Environment Constraint from lexical phonology ("structure-preserving rule applications are restricted to derived environments." p. 176.). Thus, in English, *nightingale* is pronounced [naytəngel] not *[nɪtəngel] because it is monomorphemic, whereas the alternation between [ay] and [ɪ] exists in *divine/divinity*. They hypothesized that this general principle would be reflected in SLA: Learners should perform better (i.e. make the contrasts) in the basic, compared to derived environments. Thus, a Korean learner of English would tend to palatalize the /s/ less frequently in *Jessie* than in *misinterpret*. Using nonce words, they found support for this hypothesis. The authors did not give combined percentages of all participants and conditions, but I calculated overall averages: Participants performed 13.8 percent less accurately in the derived environments than in the basic environments. Their study demonstrated that general linguistic principles can predict results and can delineate some of the conditions on transfer. They further pointed out that traditional Optimality Theory has no way of explaining their results, although they suggest that the work of Lubowicz (2002) perhaps offers a possibility.

Markedness[3]

Markedness universals deal with occurrences and likelihood of occurrences of phenomena. Markedness is defined in various ways (Carr 1993; Chomsky & Halle 1968; Greenberg 1966, 1978; Hawkins 1984; Hyman 1975; Lass 1984). One definition employs implicational hierarchies: *x* is more marked than *y* if the presence of *x* implies the presence of *y* but not vice versa. These hierarchical relationships can be seen in the following examples: (1) final voiced obstruents imply voiced obstruents in initial and medial position but not vice versa (see Eckman 1977, 1984, 1985; Eckman & Iverson 1994), (2) onsets of length *n* imply onsets of length *n–1* (Greenberg 1978), except when *n* = 1, since all languages have syllables with at least one onset), and (3) codas of length *n* imply codas of length *n–1*, but here *n* can be 0, since some languages have no syllables with codas. Markedness can also refer to statistical frequencies, for example, the *r* ([ɹ]) of American English is more marked than /l/ (in the languages of the world, [ɹ] represents only 5.6% of the liquids but [l] 42.6%, Maddieson 1984), and pharyngeal fricatives are more marked than /p/ (although Arabic has pharyngeal fricatives but no /p/). Markedness also pertains to L1 acquisition: Less marked phenomena are acquired before more marked phenomena.

Eckman's (1977) Markedness Differential Hypothesis (MDH,) brought markedness to the fore for SLA. The MDH predicts that in SLA unmarked phenomena are acquired before marked phenomena. Numerous predictions of the MDH have been found to be true, for example, in studies of voicing contrasts (Major & Faudree 1996; Yavaş 1994), epenthesis in initial consonant clusters in Egyptian learners of English (Broselow 1983), fossilization in Brunei English (Mossop 1996), coda cluster deletion in a Vietnamese speaker of English (Osburne 1996), and speech pathology (Edwards & Shriberg 1983; Gierut 1986; Hodson & Edwards 1997).

Interaction of markedness, similarity, and transfer

The SLA research I have discussed has demonstrated that markedness and universals interact with each other but it has not been shown to what degree they

3. See Eckman (this volume) for an extensive discussion on markedness.

affect transfer, specifically, is transfer more or less likely with a more marked phenomenon or a less marked phenomenon? The OPM (Major 2001) is explicit about the relationship between transfer, universals (including markedness), and similarity. The Markedness Corollary of the OPM claims that for marked phenomena L2 acquisition proceeds slowly, transfer decreases and then decreases more slowly, universals increase quickly and decrease slowly. Thus, for marked phenomena, the role of universals is much greater compared to the role of L1 transfer (when compared to less marked phenomena). In contrast, the Similarity Corollary of the OPM claims that for phenomena that are similar in L1 and L2 that L2 acquisition proceeds slowly, transfer persists; consequently the role of universals is relatively small, compared to normal or marked phenomena.

The corollaries of the OPM involving markedness and similarity both make predictions about the role of transfer. A wealth of research has shown that similarity and markedness slow acquisition more than phenomena that are neither marked nor similar. According to the OPM, similar and marked phenomena differ from normal phenomena in the relative importance of transfer versus universals, following the initial stages. For all phenomena at the beginning stages the transfer component is large and the universals component is small. Later the patterns diverge: In similar phenomena transfer persists; thus, the ratio of universals to transfer becomes relatively small throughout the subsequent stages. However, in marked phenomena universals increase rapidly and persist, resulting in a large ratio of universals to transfer in subsequent stages. Thus, for marked and similar phenomena the relative importance of transfer and universals become reversed. That is, transfer is much more important for similar phenomena than for marked phenomena, but universals are much more important for marked phenomena than for similar phenomena.

Optimality theory[4]

In classical generative phonology, SLA is framed in the following manner: Acquisition proceeds from the native language rule, changes to a developmental rule (universals), and then to an L2 rule. Optimality Theory (OT) frames SLA in a similar way but replaces rules with rankings of constraints. Thus, acquisition occurs in the following manner: native language rankings change to non L1 or non L2 rankings, and then to L2 rankings. OT (Prince & Smolensky 1993, 1997, 2004) conceives of phonological systems as a result of rankings of universal constraints. (For edited volumes on OT see Archangeli & Langendoen 1997; Lombardi 2001.) OT departs from the time-honored notion of rule, replacing it with a set of universal constraints.

4. See Hancin-Bhatt (this volume) for an extensive discussion on OT.

As pointed out by Major (2001), OT shares a remarkable similarity to Natural Phonology of nearly 30 years ago (Stampe 1969, 1979),[5] although is not generally acknowledged in the OT literature. OT constraints are Natural Phonology processes (e.g., NoCoda = coda deletion), and OT rankings in Natural Phonology are termed ordering and suppressing of processes.

In OT, markedness is nicely handled in terms of constraint rankings. A very marked phenomenon is characterized by rankings that are very rare in the languages of the world, and an unmarked phenomenon is characterized by rankings that are very common in the languages of the world. In an OT framework, L2 acquisition first starts from the L1 rankings and then proceeds from the least marked (emergence of the unmarked) to the most marked rankings. The L1 rankings are obviously transfer and the intermediate stages that are neither L1 nor L2 rankings are the result of universals (or universal constraints).

OT research in L2 phonology has been increasing, although there is still very little compared to other earlier frameworks. Studies within this framework include: Yip (1996) on loan phonology study, Hancin-Bhatt and Bhatt (1997) on syllable structures, Broselow, Chen, and Wang (1998) on coda obstruents, Hancin-Bhatt (2000) on codas in Thai ESL speakers, H-K, Kim (2001) on Korean English, Davidson (2002) on consonant clusters in English learners of Polish consonant clusters, Hancin-Bhatt (2003) on general issues of OT in L2 phonology, and Pater (2003) on the acquisition of voicing and aspiration of English-speaking Canadian learners of Thai. Bunta and Major's (2004) segmental analysis of Hungarian learners of English is perhaps the first to analyze similarity/dissimilarity within an OT framework. Using the constraints of vowel length, height, and backness, they studied the acquisition of /ɛ/, the similar segment, and /æ/, the dissimilar segment, by showing different stages involving different rankings.

I have claimed (2001) that for marked phenomena universals are more important than for similar phenomena and for similar phenomena transfer is more important than universals. Placing these claims in an OT framework we would expect to find that for similar phenomena, besides the L1 ranking persisting, that the various re-rankings proceeding to the final L2 rankings would be infrequent and/or non persistent, in comparison to marked phenomena. On the other hand, in marked phenomena L1 rankings would not persist as long and the various re-rankings on the way to L2 rankings (which would be the most marked) would be frequent and persistent, since acquisition proceeds from the least marked to the most marked. In sum, my claims of similarity and markedness can be easily stated

5. A further reminder that a theory that is a supposedly new may not really be new at all is that Stampe himself acknowledged that after he developed his theory of Natural Phonology, he later discovered the remarkable similarity of his theory to some of the claims of Baudoin de Courtenay (1895).

within an OT framework. Consider two examples: (1) an L1 French/L2 English acquisition of voiceless aspirated alveolar stop [t^h] versus the unaspirated voiceless dental stop [t̪] as an onset (e.g., *too*) and (2) an L1 Japanese/L2 English acquiring three member voiced codas (e.g., *holds*). The first example involves similarity and the second markedness. Without going into detail it can easily be seen that number of re-rankings for the French speaker would perhaps involve only two non L1/non L2 rankings – that is, whatever constraints would be involved in producing an aspirated dental stop [t̪ʰ] or unaspirated alveolar stop [t]. On the other hand, it is likely that the Japanese speaker would go through many stages: perhaps epenthesis (transfer), deletion, devoicing, and combinations of the two, all of which would be due to universal rankings, which are less marked than the rankings that would produce the three member voiced coda.

Methodological issues

A fundamental issue for research on transfer is how do we know what constitutes transfer and what does not? Previous investigations have generally assumed that identifying transfer is quite transparent, given a basic phonological description of the language, (e.g., terminal obstruent devoicing in German speakers of English is commonly known). However, determining what is transfer is not always easy, given the extensive variability of speech. Recognizing transfer presupposes a thorough phonological description of the L1 of the participants, including dialectal and stylistic variation. However, such a phonological description may be difficult or impossible for the researcher, given the resources. Nevertheless, without an L1 description it is obvious that identification of transfer is impossible. Some examples from L1 Brazilian Portuguese – L2 English illustrate the importance of a thorough knowledge of the L1. A Brazilian from the state of Minas Gerais (or the interior of the state of São Paolo) produces native-like coda /r/ in English *port*. One might think that this person has acquired English /r/, presumably having passed through various other stages, but in fact this is a case of positive transfer: a very American sounding coda /r/ occurs in these dialects, e.g., [pɔrtabɛrta] *porta aberta* "open door." However, when transfer operates in two other dialects, the dialects of Rio de Janeiro and the city of São Paolo, the results are respectively a uvular fricative and alveolar trill, clearly nonnative English sounds. Although producing a native-like English /r/ the speaker from Minas Gerais may pronounce English *bell* as [bɛr]. One may be tempted to think that this is a case of substitution due to universal factors (similar to L1 acquisition) or L2 overgeneralization of /r/ and /l/; however, in fact in this Brazilian dialect a coda /l/ is indeed pronounced [r] (e.g., [farta] *falta* "lacks") – transfer again. Another example concerns consonant clusters. Portuguese has no consonant clusters involving /s/ (also see the Japanese example previously mentioned). When a Brazilian pronounces *laps* as [læps] in running speech one may be tempted to think that he or she has acquired the /ps/ cluster. However, it positive transfer: in running speech in Portuguese (most dialects) surface consonant clusters occur due to vowel devoicing and deletion: [lapis] → [laps] *lápis* "pencil."

The above examples illustrate how crucial it is to possess a phonological description of the L1 of the learners. Ideally, then, the SLA researcher should be a native or near-native speaker of the L1 of the participants, but more importantly have a conscious knowledge of that language so that transfer can be more easily spotted (e.g. many native English speakers are unaware that slight palatalization of /s/ commonly occurs in the environment of /r/, as in *street*). However, this condition is often unrealistic and very restricting to research if strictly followed. If one does not have a thorough description of the L1 phonology then the research should be limited to the domains that pertain to that description of the L1 that one does have. For example, having an L1 description of a formal style (as most phonologies are) but no description of running speech would preclude identifying transfer in this context (see Portuguese example above). In an even worse case scenario one may have no L1 description at all. If this is the case then researchers should record samples of the L1 in the same style as the sample of the L2. After analyzing the L1 speech sample one might expect there would be a greater possibility of spotting what is transfer in the L2.

Future directions

Transfer should not be studied as an isolated phenomenon, but rather only in relationship to other factors, including markedness, similarity/dissimilarity, and other universals. Ever since the early 1970s (e.g., Nemser 1971) there has been an awareness that not only does transfer alone give an incomplete picture of interlanguage but also that transfer interacts with other factors. One goal of SLA research is to characterize interlanguage, what factors govern it as a synchronic and diachronic system, and why particular factors gain precedence over others at any given stage. Assuming we can identify what is transfer and what is not, the crux of research on transfer seems to be what are the conditions on transfer? When does it operate and when does it not? These questions will remain central issues for future transfer research.

Increasingly we have become knowledgeable of more and more details and specifics of factors other than transfer that are involved in interlanguage. Accordingly, in order to investigate the conditions of transfer we should conduct research by holding these other factors constant, which is in the spirit of scientific investigation in general. For example, in order to know how similarity affects transfer we should compare two phenomena with different degrees of similarity but with the same degree of markedness; in order to know how markedness affects transfer we should compare two phenomena with different degrees of markedness but with similarity held constant. I have made claims regarding these relationships in the OPM (Major 2001), but these claims need to be thoroughly tested. There are many opposing views on the role of transfer. To name a few: the Full Transfer Full Access Hypothesis of Schwartz and Sprouse (1994, 1996; Sprouse & Schwartz 1998); the Full Access Hypothesis (Epstein, Flynn, & Martohardjono 1996, 1998); the Failed Functional Features Hypothesis (Hawkins & Chan 1997).[6] Perhaps many of the controversies and opposing hypotheses could be resolved if researchers conducted studies following a basic scientific principle: Control for as many factors as possible, that is, keep all factors constant except the variable under consideration.

It is also necessary to be explicit about the style of speech we are investigating, as well as to be explicit whether we are making claims pertaining to a highly abstract interlanguage system or pertaining to actual utterances by the speakers that may differ from this abstract characterization.[7] OT rankings vary depending on style. Therefore, if one is trying to make general claims about the role of L1 rankings (i.e., transfer) in relationship to L2 rankings and non L1 and non L2 rankings (formerly called developmental factors) then one needs to control for the style of the L2 data, as well as the styles to which one is referring in native L1 and native L2, including the corresponding native rankings.

Conclusion

Even though most L2 phonologists do not necessarily claim or admit that transfer is the focus of their work, transfer is implicated in almost every instance. Over 10 years ago, Sharwood Smith (1996) claimed that "In the mid-seventies, given the disapproval lavished on it by creative constructionists, it seemed that no more serious words would ever be written about language transfer. In fact, the real story has only just begun" (p. 81). Perhaps the real story has more than just begun. It continues to unfold because it is widely known that the past affects one's present and future behavior.

6. Other views are discussed in this chapter in the section on Conditions on Transfer.

7. For example, as pointed out previously, consonant clusters that do not occur underlying in Japanese and Portuguese do occur in running speech due to vowel devoicing and deletion between voiceless obstruents.

Suggested readings

Andersen, R. 1983. Transfer to somewhere. In S. M. Gass & L. Selinker (eds.), *Language transfer in language learning*, Rowley, MA: Newbury House, 177–201.

Ausubel, D. P. 1963. *The psychology of meaningful verbal learning*. New York: Grune & Stratten.

Kellerman, E. 1995. Crosslinguistic influence: Transfer to nowhere? *Annual Review of Applied Linguistics* 15: 125–150.

Odlin, T. 2003. Cross-linguistic influence. In C. J. Doughty & M. H. Long (eds.), *The handbook of second language acquisition*, Malden, MA: Blackwell, 436–486.

Oller, J. W., & Ziahosseiny, S. M. 1970. The contrastive analysis hypothesis and spelling errors. *Language Learning* 20: 183–189.

Wardhaugh, R. 1970. The contrastive analysis hypothesis. *TESOL Quarterly* 4: 123–130.

Weinreich, U. 1953. *Languages in contact.* New York: Linguistic Circle of New York.

References

Altenberg, E., & Vago, R. 1983. Theoretical implications of an error analysis of second language phonology production. *Language Learning* 33: 427–447.

Anan, F. 1981. A study of intonation interference (translation from Japanese). *The Bulletin, The Phonetic Society of Japan* 167: 9–13.

Andersen, R. 1983. Transfer to somewhere. In S. M. Gass & L. Selinker (eds.), *Language transfer in language learning*, Rowley, MA: Newbury House, 177–201.

Archangeli, D., & Langendoen, D. T. (eds.). 1997. *Optimality Theory: An overview*. Malden, MA/Oxford, England: Blackwell.

Archibald, J. 1992. Transfer of L1 parameter settings: Some empirical evidence from Polish metrics. *Canadian Journal of Linguistics / Revue canadienne de Linguistique* 37: 301–339.

Atkey, S. 2002. The acquisition of non-native segmental contrasts: A look at English speakers' acquisition of Czech palatal stops. In A. James & J. Leather (eds.), *New Sounds 2000: Proceedings of the Fourth International Symposium on the Acquisition of Second-Language Speech*, Klagenfurt, Austria: University of Klagenfurt, 22–31.

Atkinson, R. C., Bower, G. H., & Crothers, E. J. 1965. *An introduction to mathematical learning theory*. New York: John Wiley & Sons.

Ausubel, D. P. 1963. *The psychology of meaningful verbal learning*. New York: Grune & Stratten.

Ausubel, D. P. 1967. *Learning theory and classroom practice*. Toronto: The Ontario Institute for Studies in Education.

Ausubel, D. P., & Robinson, F. G. 1969. *School learning: An introduction to educational psychology*. New York: Holt, Rinehart, & Winston.

Ausubel, D. P., Novak, J. D., & Hanesian, H. 1978. *Educational psychology: A cognitive view* (2nd ed.). New York: Holt, Rinehart, & Winston.

Barlow, J. A. 2002. Error patterns and transfer in Spanish-English bilingual phonological development. *Proceedings of the Annual Boston University Conference on Language Development* 26: 60–71.

Basson, S. H. 1986. *Patterns of pronunciation errors in English by native Japanese and Hebrew speakers: Interference and simplification processes. Dissertation Abstracts International,* A: The Humanities and Social Sciences, 47, 4, Cct, 1113-A.

Bayley, R., Preston, D. R. (eds.). 1996. *Second language acquisition and linguistic variation*. Amsterdam: John Benjamins.

Benson, B. 1988. Universal preference for the open syllable as an independent process in interlanguage phonology. *Language Learning* 38: 221–242.

Best, C. T. 1995. A direct-realist view of cross-language speech perception. In W. Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues*, Timonium, MD: York Press, 171–204.

Bhatt, R. M, & Hancin-Bhatt, B. 2002. Structural minimality, CP and the initial state in second language acquisition. *Second Language Research* 18: 348–392.

Birdsong, D. (ed.). 1999. *Second language acquisition and the critical period hypothesis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bley-Vroman, R. 1989. What is the logical problem of foreign language learning? In S. Gass & J. Schachter (eds.), *Linguistic perspectives on second language acquisition*, Cambridge: Cambridge University Press, 4–68.

Bley-Vroman, R. 1990. The logical problem of foreign language learning. *Linguistic Analysis* 20: 3–49.

Bohn, O.-S., & Flege, J. E. 1992. The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition* 14: 131–158.

Bohn, O-S. 2002. On phonetic similarity. In P. Burmeister, T. Piske, & A. Rohde (eds.), *An integrated view of language development*, Trier: WVT, Wissenschaftlicher Verlag Trier 191–216.

Brière, E. 1966. An investigation of phonological interference. *Language* 42: 768–796.

Brière, E. 1968. *A psycholinguistic study of phonological interference.* The Hague: Mouton.

Broselow, E. 1983. Non-obvious transfer: On predicting epenthesis errors. In S. M. Gass & L. Selinker (eds.), *Language transfer in language learning*, Rowley, MA: Newbury House, 269–280.

Broselow, E. 1984. An investigation of transfer in second language phonology. *International Review of Applied Linguistics* 22: 253–269.

Broselow, E. 2000. Stress, epenthesis, and segment transformation in Selayarese loans. In S. S. Chang, L. Liaw, & J. Ruppenhofer (eds.), *Proceedings of the twenty-fifth annual meeting of the Berkeley Linguistics Society 1999: General session and parasession on loan word phenomena*, Berkeley, CA : Berkeley Linguistics Society, 311–325.

Broselow, E., Chen, S-I., & Wang, C. 1998. The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition* 20: 261–280.

Brown, C. 1998. The role of the L1 grammar in the acquisition of segmental structure. *Second Language Research* 14: 139–93.

Brown, C. 2000. The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (ed.), *Second language acquisition and linguistic theory*, Malden, MA: Blackwell, 4–63.

Bruce, R. W. 1933. Conditions of transfer of training. *Journal of Experimental Psychology*, 16: 343–361.

Bugelski, B. R. 1942. Interferences with recall of original responses after learning new responses to old stimuli. *Journal of Experimental Psychology* 30: 368–379.

Bunta, F., & Major, R. C. 2004. An optimality theoretic account of Hungarian ESL learners Acquisition of /ɛ/ and /æ/. *International Review of Applied Linguistics* 42: 277–298.

Carr, P. 1993. *Phonology*. New York: St. Martin's Press.

Cheng, N. Y. 1929. Retroactive effect and degree of similarity. *Journal of Experimental Psychology* 12: 444–449.

Chomsky, N., & Halle, M. 1968. *The sound pattern of English*. New York: Harper & Row.

Curtin, S, Goad, H., & Pater, J. V. 1998. Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language* Research 14: 389–405.

Davidson, L. 2002. The effects of hidden rankings on the acquisition of consonant clusters. In A. James & J. Leather (eds.), *New Sounds 2000: Proceedings of the Fourth International Symposium on the Acquisition of Second-Language Speech*, Klagenfurt, Austria: University of Klagenfurt, 87–96.

Dreasher, L. M., & Anderson-Hsieh, J. 1990. Universals in interlanguage phonology: The case of Brazilian ESL learners. *Papers and Studies in Contrastive* Linguistics 26: 69–92.

Duncan, C. P. 1958. Transfer after training with single versus multiple tasks. *Journal of Experimental Psychology* 55: 63–72.

Eckman, F. R. 1977. Markedness and the contrastive analysis hypothesis. *Language Learning* 27: 315–330.

Eckman, F. R. 1984. Universals, typologies and interlanguage. In W. Rutherford (ed.), *Language universals and second language acquisition*, Amsterdam: John Benjamins, 79–105.

Eckman, F. R. 1985. Some theoretical and pedagogical implications of the Markedness Differential Hypothesis. *Studies in Second Language* Acquisition 7: 289–307.

Eckman, F. R. 2008. Typological markedness and second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Eckman, F. R., & Iverson, G. K. 1993. Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research* 9: 234–252.

Eckman, F. R., & Iverson, G. K. 1994. Pronunciation difficulties in ESL: Coda consonants in English interlanguage. In M. Yavaş (ed.), *First and second language phonology*, San Diego, CA: Singular, 251–265.

Eckman, F. R., Elreyes, A., & Iverson, G. K. 2003. Some principles of second language phonology. *Second Language Research* 19: 169–208.

Edge, B. A. 1991. The production of word-final voiced obstruents in English by L1 speakers of Japanese and Cantonese. *Studies in Second Language Acquisition* 13: 377–393.

Edwards, M. L., & Shriberg, L. D. 1983. *Phonology: Applications in communicative disorders*. San Diego: College Hill Press.

Ellis, N. C. 1996. Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition* 18: 91–126.

Epstein, S., Flynn, S., & Martohardjono, G. 1996. Second language acquisition: Theoretical and experimental issues in contemporary research. *Brain and Behavioral Sciences* 19: 677–758.

Epstein, S., Flynn, S., & Martohardjono, G. 1998. The strong continuity hypothesis: Some evidence concerning functional categories in adult L2 acquisition. In S. Flynn & G. Martohardjono (eds.), *The generative study of second language acquisition*, Mahwah, NJ: Lawrence Erlbaum Associates, 61–71.

Eubank, L. 1993/1994. On the transfer of parametric values in L2 development. *Language Acquisition* 3: 183–208.

Eubank, L. 1994. Optionality and the initial state in L2 development. In T. Hoekstra & B. D. Schwartz (eds.), *Language acquisition studies in generative grammar: Papers in honor of Kenneth Wexler from the 1991 GLOW Workshops*, Amsterdam: John Benjamins, 369–388.

Eubank, L. 1996. Negation in early German-English interlanguage: More valueless features in the L2 initial state. *Second Language Research* 12: 73–106.

Flege, J. E., & MacKay, I. R. A. 2004. Perceiving vowels in a foreign language. *Studies in Second Language Acquisition* 26: 1–34.

Flege, J. E., Schirru, C., & MacKay, I. R. A. 2003. Interaction between the native and second language phonetic subsystems. *Speech Communication* 40: 467–491.

Flege, J. E. 1987a. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–65.

Flege, J. E. 1987b. Effects of equivalence classification on the production of foreign language speech sounds. In A. James & J. Leather (eds.), *Sound patterns in second language acquisition*, Dordrecht, The Netherlands: Foris, 9–39.

Flege, J. E. 1990. English vowel production by Dutch talkers: More evidence for the "similar" vs. "new" distinction. In J. Leather & A. James (eds.), *New Sounds 90: Proceedings of the Amsterdam Symposium on the Acquisition of Second-Language Speech*, Amsterdam: University of Amsterdam, 255–293.

Flege, J. E. 1992. The intelligibility of English vowels spoken by British and Dutch talkers. In R. D. Kent (ed.), *Intelligibility in speech disorders: Theory, measurement, and management* (Vol. 1), Philadelphia: John Benjamins, 157–232.

Flege, J. E. 1993. Production and perception of a novel, second language phonetic contrast. *Journal of the Acoustical Society of America* 93: 1589–1608.

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-linguistic research*, Timonium, MD: York Press, 233–277.

Flege, J. E., & Davidian, R. 1984. Transfer and developmental processes in adult foreign language speech production. *Applied Psycholinguistics* 5: 323–347.

Flege, J. E., Bohn, O.-S., & Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25: 437–470.

Flores, B., & Rodrigues, X. 1994. The influence of language transfer on consonant cluster production. *Revista de Filologia y Linguistica de la Universidad de Costa Rica* 20: 99–112.

Fries, C. C. 1945. *Teaching and learning English as a foreign language*. Ann Arbor: University of Michigan Press.

Gagné, R. M. 1965. The acquisition of knowledge. In R. C. Anderson & D. P. Ausubel (eds.), *Readings in the psychology of cognition*, New York: Holt, Rinehart, & Winston, 116–132.

Gagné, R. M. 1977. *The conditions of learning* (3rd ed.). New York: Holt, Rinehart, & Winston.

Gass, S. M., & Selinker, L. 2001. *Second language acquisition: An introductory course* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S. M., & Selinker, L. (eds.). 1983. *Language transfer in language learning*. Rowley, MA: Newbury House.

Gass, S. M., & Selinker, L. (eds.). 1992. *Language transfer in language learning*. Amsterdam: John Benjamins.

Gierut, J. A. 1986. *On the relationship between phonological knowledge and generalization learning in misarticulating children*. Bloomington: Indiana University Linguistics Club.

Greenberg, J. 1966. *Language universals*. The Hague, The Netherlands: Mouton.

Greenberg, J. 1978. Some generalizations concerning initial and final consonant clusters. In J. Greenberg, C. Ferguson, & E. Moravcsik (eds.), *Universals of human language* (Vol. 2), Stanford, CA: Stanford University Press, 243–279.

Guion, S., Flege, J. E., & Loftin, J. D. 2000. The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. *Journal of Phonetics* 28: 27–42.

Han, Z-H. 2004. *Fossilization in adult second language acquisition.* Clevedon/Buffalo/Toronto/Sydney: Multilingual Matters.

Han, Z-H., & Odlin, T. (eds.). 2006. *Studies of fossilization in second language acquisition*. Clevedon/Buffalo/Toronto: Multilingual Matters.

Hale, M., & Reiss, C. 2000. "Substance abuse" and "dysfunctionalism": Current trends in phonology. *Linguistic Inquiry* 31: 157–169.

Hale, M., & Reiss, C. 2003. The Subset Principle in phonology: Why the tabula can't be rasa. *Journal of Linguistics* 39: 219–244.

Halle, P. A., Best, C. T., & Levitt, A. 1999. Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics* 27: 281–306.

Hancin-Bhatt, B. 1994. Segment transfer: A consequence of a dynamic system. *Second Language Research* 10: 241–269.

Hancin-Bhatt, B. 2000. Optimality in second language phonology: Codas in Thai ESL. *Second Language Research* 16: 201–232.

Hancin-Bhatt, B. 2003. Second language phonology in optimality theory. Unpublished paper present at the Second Language Research Forum, Tucson, AZ.

Hancin-Bhatt, B. 2008. Second language phonology in optimality theory. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Hancin-Bhatt, B. & Bhatt, R. M. 1997. Optimal L2 syllables: Interactions of transfer and developmental effects. *Studies in Second Language Acquisition* 19: 331–378.

Hancin-Bhatt, B., & Govindjee, A. 1999. A Computational Model of Feature Competition in L2 Phonology. In P. Broeder & J. Murre, (eds.), *Language and thought in development: Cross-linguistic studies*, Tubingen, Germany: Narr, 145–162.

Haugen, E. 1956. *Bilingualism in the Americas: A bibliography and research guide.* Baltimore: American Dialect Society.

Hawkins, R., & Chan, Y-H. C. (1997). The partial availability of Universal Grammar in second language acquisition: The "Failed Functional Features Hypothesis." *Second Language Research* 13: 187–226.

Hecht, B. F., & Mulford, R. 1982. The acquisition of a second language phonology: interaction of transfer and developmental factors. *Applied Psycholinguistics* 3: 313–328.

Herschensohn, J., & Stevenson, J. 2003. Failed features or missing inflection? Child L2A of Spanish morphology. *Proceedings of the Annual Boston University Conference on Language Development* 27: 299–310.

Hodne, B. 1985. Yet another look at interlanguage phonology: The modification of English syllable structure by native speakers of Polish. *Language Learning* 35: 405–422.

Hodson, B. W., & Edwards, M. L. 1997. *Perspectives in applied phonology*. Gaithersburg, MD: Aspen.

Hung, T. N., & Man, V. 2002. Investigating interlanguage phonology: The case of Hong Kong English. In A. James & J. Leather & (eds.), *New Sounds 2000: Proceedings of the Fourth International Symposium on the Acquisition of Second-Language Speech*, Klagenfurt, Austria: University of Klagenfurt, 174–183.

Hyman, L. 1975. *Phonology: Theory and analysis*. New York: Holt, Rinehart, & Winston.

Ioup, G. 1996. Grammatical knowledge and memorized chunks: A response to Ellis. *Studies in Second Language Acquisition* 18: 355–360.

Ioup, G. 2008. Exploring the role of age in the acquisition of a second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Jakobson, R. 1941. *Kindersprache, Aphasie, und Allgemeine Lautgestze*. Translated 1968 as *Child language, aphasia, and phonological universals* (A. R. Keiler, Trans.). The Hague: Mouton.

James, A. R. 1983. Transferability and dialect phonology: Swabian: Swabian English. In A. James & B. Kettemann (eds), *Dialektphonologie und Fremdsprachenerwerb* [*Dialect phonology and foreign language acquisition*], Tübingen, Germany: Narr, 162–188.

James, A. R. 1986. *Suprasegmental phonology and segmental forms*. Tübingen, Germany: M. Niemeyer.

James, A. R. 1988. *The acquisition of a second language phonology: A linguistic theory of developing sound structures*. Tübingen, Germany: Gunter Narr Verlag Tübingen.

James, A. R. 1989. Linguistic theory and second language phonological learning: A perspective and some proposals. *Applied Linguistics* 10: 367–381.

James, A. R. 1996. Second language phonology. In P. Jordens & J. Lalleman (eds.), *Investigating second language acquisition*, Berlin: Mouton de Gruyter, 293–320.

Kellerman, E. 1995. Crosslinguistic influence: Transfer to nowhere? *Annual Review of Applied Linguistics* 15: 125–150.

Kellerman, E., & Sharwood Smith, M. (eds.). (1986). *Cross-linguistic influence in second language acquisition*. New York: Pergamon Press.

Kim, H-K. 2001. The interlanguage phonology of Korean learners of English: A computational implementation of Optimality Theoretic constraints. Unpublished doctoral dissertation, Georgetown University.

Kim, J-E. , & Silva, D. J. 2003. An acoustic study of the American English pronunciation of recently arrived Korean adult immigrants. *Ohak Yonku/Language Research* 39: 613–637.

Lado, R. 1957. *Linguistics across cultures*. Ann Arbor: University of Michigan Press.

Lass, R. 1984. *Phonology: An introduction to basic concepts*. Cambridge: Cambridge University Press.

Leather, J. 1983. Speaker normalization in perception of lexical tone. *Journal of Phonetics* 11: 373–382.

Leather, J. 1987. F0 pattern inference in the perceptual acquisition of second language tone. In A. James & J. Leather (eds), *Sound patterns in second language acquisition*, Dordrecht, The Netherlands: Foris, 59–80.

Leung, Y-K. I. 2001. The initial state of L3A: Full transfer and failed Features? *Selected Proceedings of the Second Language Research Forum,* 55–75.

Lombardi, L. 2003. Second language data and constraints on manner: Explaining substitutions for the English interdentals. *Second Language Research* 19: 225–250.

Lombardi, L. (ed.). (2001). *Segmental phonology in optimality theory: Constraints and representations*. Cambridge: Cambridge University Press.

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

Lubowicz, A. 2002. Derived environment effects in Optimality Theory. *Lingua* 112: 243–280.

MacKain, K. S., Best, C. T., & Strange, W. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics* 2: 368–390.

Maddieson, I. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Major, R. C. 1985. Stress and rhythm in Brazilian Portuguese. *Language* 61: 259–282.

Major, R. C. 1987a. A model for interlanguage phonology. In G. Ioup & S. H. Weinberger (eds.), *Interlanguage phonology: The acquisition of a second language sound system*, New York: Newbury House/Harper & Row, 101–125.

Major, R. C. 1987b. Phonological similarity, markedness, and rate of L2 acquisition. *Studies in Second Language Acquisition* 9: 63–82.

Major, R. C. 1996. Chunking and phonological memory: A response to Ellis. *Studies in Second Language Acquisition* 18: 351–354.

Major, R. C. 1997. Further evidence for the Similarity Dissimilarity Rate Hypothesis. In J. Leather & A. James (eds.), *New Sounds 97: Proceedings of the Third International Symposium on the Acquisition of Second-Language Speech*, Klagenfurt, Austria: University of Klagenfurt, 215–222.

Major, R. C. 2001. *Foreign accent: The ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Major, R. C., & Faudree, M. C. 1996. Markedness universals and the acquisition of voicing contrasts in Korean speakers of English. *Studies in Second Language Acquisition* 18: 69–90.

Major, R. C., & Kim, E. 1996. The Similarity Differential Rate Hypothesis. *Language Learning* 46: 465–496.

Marghany, M. M. 2002. Some phonetic features of Egyptian English: A descriptive and pedagogical perspective. In A. James & J. Leather & (eds.), *New Sounds 2000: Proceedings of the Fourth International Symposium on the Acquisition of Second-Language Speech*, Klagenfurt, Austria: University of Klagenfurt, 231–235.

Markham, D. 1997. *Phonetic imitation, accent, and the learner*. Lund: Lund University Press; Kent, England: Chartwell-Bratt.

McGeogh, J. A. 1942. *The psychology of human learning*. New York: Longmans, Green.

Mossop, J. W. 1996. Markedness and fossilization in the interlanguage phonology of Brunei English. *World Englishes* 15: 171–182.

Moyer, A. 2004. *Age, Accent and experience in second language acquisition*. Clevedon: Multilinguial Matters.

Müller, M. 1861/1965. *Lectures on the science of language*. New Delhi: Munshi Ram Manohar Lal.

Munro, M., Derwing, T., Flege, J. E. 1999. Canadians in Alabama: A Perceptual Study of Dialect Acquisition in Adults. *Journal of Phonetics* 27: 385–403.

Musau, P. M. 1993. *Aspects of interphonology: The study of Kenyan learners of Swahili*. Bayreuth, Germany: Bayreuth University.

Nemser, W. 1971. *An experimental study of phonological interference in the English of Hungarians*. Bloomington: Indiana University Press.

Neuner, S. 2002. Learning as information processing: Reflections on cognitive theory in second language instruction (original title: Lernen als Verarbeitung von Informationen. Kognitionstheoretische Uberlegungen zum Fremdsprachenunterricht). *Babylonia* 10: 35–39.

Odlin, T. 1989. *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Odlin, T. 2003. Cross-linguistic influence. In C. J. Doughty & M. H. Long (eds.), *The handbook of second language acquisition*, Malden, MA: Blackwell, 436–486.

Oller, J. W., & Ziahosseiny, S. M. 1970. The contrastive analysis hypothesis and spelling errors. *Language Learning* 20: 183–189.

Osburne, A. G. 1996. Final consonant cluster reduction in English L2 speech: A case study of a Vietnamese speaker. *Applied Linguistics* 178: 164–181.

Osgood, C. A. 1946. Meaningful similarity and interference in learning. *Journal of Experimental Psychology* 36: 277–301.

Osgood, C. A. 1949. The similarity paradox in human learning: A resolution. *Psychological Review* 56: 132–143.

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

Paik, Y.-J. 1977. Pronunciation difficulties of Korean students learning English as a second language. *Dhak Yonku* 13: 177–189.

Pater, J. 2003. The perceptual acquisition of Thai phonology by English speakers: Task and stimulus effects. *Second Language Research* 19: 209–223.

Peng, L., & Ann, J. 2002. An alternative to stress placement: Unity in stress placement in English as second language. In A. James & J. Leather (Eds.), *New Sounds 2000: Proceedings of the Fourth International Symposium on the Acquisition of Second-Language Speech*, Klagenfurt, Austria: University of Klagenfurt, 271–279.

Pennington, M. C. 1992. Discourse factors in second language phonology: An exploratory study. In J. Leather & A. James (eds.), *New Sounds 92: Proceedings of the Amsterdam Symposium on the Acquisition of Second-Language Speech*, Amsterdam: University of Amsterdam, 137–155.

Piper, T. 1984. Observations on the second-language acquisition of the English sound system. *Canadian Modern Language Review* 40: 542–551.

Piske, T, MacKay, I. R. A., Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215.

Poulisse, N. 1999. *Slips of the tongue: Speech errors in first and second language production*. Amsterdam: John Benjamins.

Prince, A., & Smolensky, P. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished manuscript, Rutgers University, New Brunswick, NJ and University of Colorado, Boulder, and Rutgers University Optimality Archive.

Prince, A., & Smolensky, P. 1997. Optimality: From neural networks to Universal Grammar. *Science* 275: 1604–1610.

Prince, A., & Smolensky, P. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA/Oxford, England: Blackwell.

Redard, F. 1973. *Contrastive phonemics: Notes and exercises*. Neuchâtel, Switzerland: Centre de Linguistique, Universite de Neuchâtel.

Riney, T. J. 1989. Syllable structure and interlanguage phonology. *Kansas Working Papers in Linguistics* 14. Lawrence: University of Kansas.

Sajavaara, K., & Dufva, H. 2001. Finnish-English Phonetics and Phonology. *International Journal of English Studies* 1: 241–256.

Schultz, R. W. 1960. Problem solving behavior and transfer. *Harvard Educational Review* 30: 61–77.

Schwartz B. D, & Eubank, L. 1996. What is the 'L2 initial state'? *Second Language Research* 12: 1–5.

Schwartz, B. D., & Sprouse, R. A. 1994. Word order and nominative case in nominative case in nonnative language acquisition: A longitudinal study of (L1 Turkish) German interlanguage. In T. Hoekstra & B. D. Schwartz (eds.), *Language acquisition studies in generative grammar: Papers in honor of Kenneth Wexler from the 1991 GLOW Workshops*, Amsterdam: John Benjamins, 317–368.

Schwartz, B. D., & Sprouse, R. A. 1996. L2 cognitive states and the full transfer/full access model. *Second Language Research* 12: 40–72.

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

Scovel, T. 1988. *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. New York: Newbury House/Harper & Row.

Scovel, T. 2000. A critical review of the Critical Period research. *Annual Review of Applied Linguistics* 20: 213–223.

Selinker, L. 1972. Interlanguage. *International Review of Applied Linguistics* 10: 209–231.

Seubsunk, S. 2000. The pronunciation of English stop-liquid and stop-glide clusters in word-initial position by native speakers of two different dialects of Thai. Unpublished doctoral dissertation, University of Delaware.

Sharwood Smith, M. 1996. Crosslinguistic influence with special reference to the acquisition of grammar. In P. Jordens & J. Lalleman (eds.), *Investigating second language acquisition*, Berlin: Mouton de Gruyter, 71–83.

Sheldon, A., & Strange, W. 1982. The Acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3: 243–261.

Shirai, Y. 1992. Conditions on transfer: A connectionist approach. *Issues in Applied Linguistics* 3: 91–120.

Siegel. J. 2003. Substrate influence in creoles and the role of transfer in second language acquisition. *Studies in Second Language Acquisition* 25: 185–209

Soudak, L. I. 1977. Error analysis and the system of English consonants. *ELT Journal* 31: 125–130.

Sprouse, R. A., & Schwartz, B. D. 1998. In defense of full transfer in German-English and French-English interlanguage: Comparative L2 acquisition research. *Proceedings of the Annual Boston University Conference on Language Development* 22: 726–736.

Stampe, D. 1969. The acquisition of phonetic representation. *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society,* 443–453.

Stampe, D. 1979. *A dissertation in Natural Phonology*. New York: Garland.

Steele, J. 2001. Phonetic cues to phonological acquisition: Evidence from L2 syllabification. *Proceedings of the Annual Boston University Conference on Language Development* 25: 732–743.

Stockwell, R., & Bowen, J. 1965. *The sounds of English and Spanish*. Chicago: University of Chicago Press.

Stockwell, R., & Bowen, J., & Martin, J. W. 1965. *The grammatical structures of English and Spanish*. Chicago: University of Chicago Press.

Strange, W., & Shafer, V. L. 2008. Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Tomaszczyk, J. 1980. On accented speech: The Polish of Polish Americans. *Studia Anglica Posnaniensia*, 12: 121–137.

Travers, R. M. W. 1977. *Essentials of learning*, fourth edition. New York: Macmillan.

Trubetzkoy, N. 1958. Grundzüge der Phonologie [Fundamentals of Phonology] (Travaux du cercle linguistique de Prague 7). Göttingen: Vandenhoek & Ruprecht. (Original work published 1939).

Vainikka, A., & Young-Scholten, M. 1994. Direct Access to X'-theory: Evidence from Korean and Turkish adults learning German. In Hoekstra & B. D. Schwartz (eds.), *Language acquisition studies in generative grammar: Papers in honor of Kenneth Wexler from the 1991 GLOW Workshops*, Amsterdam: John Benjamins, 265–316.

Vainikka, A., & Young-Scholten, M. 1996a. The early stages of adult L2 syntax: Additional evidence from Romance speakers. *Second Language Research* 12: 140–176.

Vainikka, A., & Young-Scholten, M. 1996b. Gradual development of L2 phrase structure. *Second Language Research* 12: 7–39.

Wang, M., & Geva, E. 2003. Spelling acquisition of novel English phonemes in Chinese children. *Reading and Writing* 4: 325–348.

Waniek-Kimczak, E. 2002. How to predict the unpredictable – English word stress from a Polish perspective. In E. Waniek-Kimczak & P. J. Melia (eds.), *Accents and speech in teaching English phonetics and phonology*, Frankfurt: Peter Lang, 221–241.

Wardhaugh, R. 1970. The contrastive analysis hypothesis. *TESOL Quarterly* 4: 123–130.

Weinreich, U. 1953. *Languages in contact.* New York: Linguistic Circle of New York.

Wenk, B. J. 1986. Crosslinguistic influence in second language phonology: Speech rhythms. In E. Kellerman & M. Sharwood Smith (eds.), *Crosslinguistic influence in second language acquisition*, Elmsford, NY: Pergamon Press, Inc., 120–133.

White, L, Valenzuela, E., Kozlowska-Macgregor, M. & Leung, Y-K. I. 2003. Gender agreement in L2 Spanish: Evidence against failed features. *McGill Working Papers in Linguistics/Cahiers linguistiques de McGill* 17: 153–182.

White, L. 1989. *Universal Grammar and second language acquisition*. London: Hodder & Stoughton.

White, L. 2002. Morphological variability in endstate L2 grammars: The question of L1 influence. *Proceedings of the Annual Boston University Conference on Language Development* 26: 758–768.

White, L. 2003a. S*econd language acquisition and Universal Grammar*. London: Hodder & Stoughton. Cambridge: Cambridge University Press.

White, L. 2003b. On the nature of interlanguage representation: Universal Grammar in the second language. In C. J. Doughty & M. H. Long (eds.), *The handbook of second language acquisition*, Malden, MA: Blackwell 19–42.

Whong-Barr, M., & Schwartz, B. D. 2002. Morphological and syntactic transfer in child L2 acquisition of the English dative alternation. *Studies in Second Language Acquisition* 24: 579–616.

Williams, L. 1979. The modification of speech perception and production in second language learning. *Perception and Psychophysics* 26: 95–104.

Wode, H. 1977. The L2 acquisition of /r/. *Phonetica* 34: 200–217.

Wode, H. 1978. The beginning of non-school room L2 phonological acquisition. *International Review of Applied Linguistics in Language Learning* 16: 109–125.

Wode, H. 1980. Phonology in L2 acquisition. In S. Felix (ed.), *Second language development*, Tübingen, Germany: Narr, 123–136.

Wode, H. 1983a. Phonology in L2 acquisition. In H. Wode (ed.), *Papers on language acquisition, language learning and language teaching*, Heidelberg, Germany: Groos, 175–187.

Wode, H. 1983b. Contrastive analysis and language learning. In H. Wode (ed.), *Papers on language acquisition, language learning and language teaching*, Heidelberg, Germany: Groos, 202–212.

Wolfram, W., Childs, B., & Torbert, B. 2000. Tracing language history through consonant cluster reduction: Comparative evidence from isolated dialects. *Southern Journal of Linguistics* 24: 17–40.

Yavaş, M. 1994. Final stop devoicing in interlanguage. In M. Yavaş (ed.), *First and second language phonology*, San Diego, CA: Singular, 267–282.

Yavaş, M. 1997. The effects of vowel height and place of articulation in interlanguage final stop devoicing. *International Review of Applied Linguistics* 35: 115–125.

Yip, Y. M. 1996. Cantonese loan phonology and Optimality Theory. *Journal of East Asian Linguistics* 2: 261–291.

Young-Scholten, M. 1985. Interference reconsidered: The role of similarity in second language acquisition. *Selecta* 6: 6–12.

Zampini, M. L. 1996. Voiced stop spirantization in the ESL speech of native speakers. *Applied Psycholinguistics* 17: 335–354.

Zampini, M. L. 1997. L2 Spanish spirantization, prosodic domains, and interlanguage rules. In S. J. Hannahs & M. Young-Scholten (eds.), *Focus on phonological acquisition*, Amsterdam: John Benjamins, 209–234.

Zampini, M. L. 2008. L2 speech production research: Findings, issues, and advances. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Zsiga, E. C. 2003. Articulatory timing in a second language. *Studies in Second Language Acquisition*, 25: 399–432.

CHAPTER 4

Typological markedness and second language phonology

Fred R. Eckman
University of Maryland Center for the Advanced Study of Language and University of Wisconsin-Milwaukee

Introduction

The purpose of this chapter is twofold:[1] The primary aim is to give an overview of the role of typological markedness in the explanation of facts about second language (L2) phonology. A secondary goal is to explore some of the implications of using such markedness principles to explain facts about L2 phonology. This discussion leads naturally to a consideration of some of the major issues and counter claims surrounding the use of markedness as an explanatory principle in second language acquisition (SLA) in general, and L2 phonology in particular.

The remainder of this chapter is structured as follows. The background section sketches out a brief history of typological markedness, with the following sections discussing the two major hypotheses in SLA that have been formulated around this concept. The treatment of each hypothesis includes a presentation of the kind of evidence that has been adduced in favor of the hypothesis, as well as an evaluation of what the field has gained from the hypothesis and a critical look at what remains to be learned. The discussion then turns to what appears to be a viable future direction for a research program in L2 phonology that incorporates markedness. The final section concludes the chapter.

1. An earlier version of this paper was presented at the 2003 Second Language Research Forum, October 18, 2003 at the University of Arizona. I would like to thank members of the audience for their questions, comments and general feedback. I also wish to express my appreciation to the editors of this volume, Jette Hansen Edwards and Mary Zampini for their comments and suggestions. As always, any remaining errors or inconsistencies are my own. This work was supported in part by a grant from the National Institutes of Health 1 R01 HD046908-01-A2.

I begin with a brief discussion of the origin of markedness in linguistic theory and its use in SLA.

Background

Markedness

The principle of markedness was pioneered by the Prague School of Linguistics in the theories of Nikolai Trubetzkoy (1939) and Roman Jakobson (1941). The idea behind this concept was that binary oppositions between certain linguistic representations (e.g. voiced and voiceless obstruents, nasalized and oral vowels, open and closed syllables) were not taken to be simply polar opposites. Rather, one member of the opposition was assumed to be privileged in that it had a wider distribution, both within a given language and across languages. Imposing a markedness value on this opposition was one way of characterizing this special status: the member of the opposition that was more widely distributed than the other was designated as unmarked, indicating that it was, in some definable way, simpler, more basic and more natural than the other member of the opposition, which was in turn defined as the marked member. In the examples cited above, voiceless obstruents, oral vowels and open syllables are all unmarked relative to, respectively, voiced obstruents, nasalized vowels and closed syllables.

As Battistella (1990) points out, there have been over the years a number of different approaches to, and definitions of, markedness (see Moravcsik & Wirth 1986 for some examples), including the presence or absence of overt marking, occurrence in the environment in which neutralization occurs, amount of evidence required for acquisition by child-learners, and the frequency of occurrence across the world's languages. The last notion, distribution among the languages of the world, where there is an implicational relationship between the occurrence of the members of the opposition, is known as typological markedness, and was developed extensively in the work of Greenberg (1976) and can be defined as in (1).[2]

(1) A structure X is typologically marked relative to another structure, Y, (and Y is typologically unmarked relative to X) if every language that has X also has Y, but every language that has Y does not necessarily have X.
(Gundel, Houlihan & Sanders 1986: 108)

2. Other definitions of markedness have been used in the literature on L2 phonology, including the conceptualization of markedness in terms of parametric variation within Universal Grammar. This work includes studies such as Broselow and Finer (1991) and Archibald (1998).

Under this view, typological markedness is an asymmetric, irreflexive and transitive relationship between linguistic representations across the world's languages, such that the presence of one structure in a language implies the presence of another structure, but not vice versa.

Greenberg (1976) noted that, in attempting to formulate universal generalizations about human languages, linguists have often found the most insightful statements to be implicational; that is, the most enlightening universals are formulated in terms of typological markedness. To take a concrete example, not all languages have a contrast in voice, and furthermore, if a language exhibits a voice contrast in some environments, it may not exhibit this contrast in all environments. Nevertheless, it is possible to state a universal generalization about the occurrence of a voice contrast in a language if one states this generalization implicationally, as in (2) below.

(2) If a language has a voice contrast in syllable coda position, it necessarily has this contrast in syllable onset position, but not vice versa.

Thus, a language may not evince a voice contrast in any of its utterances; but if a language does have a voice contrast anywhere, it will have it in syllable onset position. In addition to onset position, a language may also have a voice contrast in coda position; but if a language has a voice contrast in codas, it will necessarily have the contrast in onsets. The claim underlying the idea of markedness, then, is that there is something "basic", "natural" or "common" about a language having a voice contrast in onsets but not in codas, or a language having only oral vowels but not nasalized vowels, or a language having open syllables, but not closed syllables. It is this type of thinking that is embodied in the idea of typological markedness.

Finally, an important aspect of typological markedness that has made it a particularly useful theoretical tool is that linguists have been able to apply this construct to virtually all kinds of linguistic expressions, including, besides the above phonological examples, lexical, morphological, and syntactic structures, in a number of sub-domains of linguistics. In the next section I will focus on the role of markedness in L2 phonology, more specifically, I will discuss the claim that marked structures are more difficult than the corresponding unmarked structures.

Markedness in second language phonology

The markedness differential hypothesis

There are two hypotheses relevant to L2 phonology that have been formulated using the construct of typological markedness, the Markedness Differential Hypothesis (MDH) (Eckman 1977), and the Structural Conformity Hypothesis (SCH) (Eckman 1991). I consider each in turn.

The MDH, stated in (3), claimed that typological markedness must be incorporated into the classic Contrastive Analysis Hypothesis (CAH) (Lado 1957; Stockwell & Bowen 1965) as a measure of relative difficulty in SLA.

(3) The Markedness Differential Hypothesis (Eckman 1977: 321)
The areas of difficulty that a language learner will have can be predicted such that

a. Those areas of the target language which differ from the native language and are more marked that the native language will be difficult;
b. The relative degree of difficulty of the areas of difference of the target language which are more marked that the native language will correspond to the relative degree of markedness;
c. Those areas of the target language which are different from the native language, but are not more marked than the native language will not be difficult.

Whereas the CAH attempted to explain L2 learning difficulty only on the basis of differences between the native language (NL) and target-language (TL), the MDH claimed that NL-TL differences were necessary for such an explanation, but they were not sufficient; rather, one needed to incorporate typological markedness into the explanation. The hypothesis asserts that, within the areas of difference between the NL and TL, marked structures are more difficult than the corresponding unmarked structures.

What follows immediately from this hypothesis is that not all NL-TL differences will cause equal difficulty. TL structures that are different from the NL but are not related by markedness principles to any other structures are predicted to cause no difficulty, while TL constructions which are related to other representations by markedness principles are predicted to cause learning problems. The degree of difficulty involved is predicted to correspond directly to the relative degree of markedness.

Evidence for the Markedness Differential Hypothesis

The kind of evidence adduced in support of the MDH involved showing that learner errors could not be accounted for on the basis of NL-TL differences alone, but that typological markedness was necessary to explain the difficulty that learners encountered. One such type of evidenced can be termed "directionality of difficulty", and results when speakers from two different NL backgrounds attempt to learn the other's language, and one learner encounters more difficulty than the other. Another type of evidence involves markedness being invoked to explain the different degrees of difficulty associated with learners from different NL backgrounds all acquiring the same TL. The third type of evidence in support of the MDH indicates that markedness can predict the relative degree of difficulty associated with the learning of various TL structures. I will discuss examples of each in turn.

An example of the first type of evidence, directionality of difficulty, was reported in Moulton (1962), in which the author stated that the difference between German and English involving voice contrasts in syllable codas caused more difficulty for German speakers learning English than it did for English speakers learning German. The phonological facts are that English has a voice contrast in obstruents word-initially, -medially and -finally, whereas German exhibits this contrast only word-initially and word-medially. In word-final position in German, this contrast is neutralized in favor of voiceless obstruents. Moulton (1962) stated that, for Germans learning English, acquiring a voice contrast in word-final position was very difficult, whereas for English speakers learning German, the lack of a voice contrast word-finally was not difficult to learn. This example was discussed within the context of the MDH in Eckman (1977), where it was argued that this asymmetry in learning resulted from the German speakers' having to acquire a relatively more marked structure, a voice contrast in codas, compared to what the English-speaking learners of German had to acquire.

An example of the second kind of evidence, that deriving from different amounts of difficulty associated with learners from diverse NL backgrounds learning a given TL, comes from, among others, Anderson (1987). This study analyzed the learning of onset and coda clusters in English for subjects from three NL backgrounds, Egyptian Arabic, Mandarin Chinese and Amoy Chinese. The markedness principle in question in Anderson's study concerned consonantal sequences in syllable-onset and syllable-coda positions. Specifically, the existence of an onset cluster of length N in a language implies the occurrence of onset clusters of length N-1 in that language, where N is an integer. For example, a language that allows three consonants to cluster in onsets will necessarily allow two-consonant cluster, but not vice versa, and a language that allows bi-consonantal onsets will also permit singleton consonants in onsets, but not vice versa. The same principle holds for codas: the presence of a coda cluster of length N in a language implies the occurrence of coda clusters of length N-1. In sum, longer clusters in onsets and codas are more marked relative to, respectively, shorter clusters in onsets and codas. The results of Anderson's study supported the MDH in that the performance of the Chinese-speaking subjects was less target-like than that of the Arabic-speaking subjects on coda clusters, and the difference in performance correlated with degree of markedness and with the amount of NL-TL difference.

Additional examples of this kind of support for the MDH were reported in Eckman (1981a, 1981b), in which it was argued that speakers from different NL backgrounds performed differently on voiced obstruents in codas. The determining factor seemed to be whether or not the subject's NL allowed any obstruents in codas; Japanese and Mandarin do not allow coda obstruents, and consequently subjects from these NL backgrounds were more likely to add a vowel at the end of the TL word, whereas Cantonese and Spanish do allow some obstruents in codas, and subjects from these NL backgrounds devoiced the final consonant in the TL word.

The third kind of evidence, data showing that learners' performance on different TL structures can be explained only by invoking the markedness relationships that exist among the structures in question, is exemplified in Carlisle (1991). In this study the author analyzed the production of complex onsets in English by native speakers of Spanish, using a reading task. Because the elicitation involved the subjects' producing an oral text, the number of different environments for inserting an epenthetic vowel to break up a consonant cluster was increased by taking into account the final segments in the preceding word. The findings showed that the subjects modified the complex onsets by inserting an epenthetic vowel, and that the likelihood of a given onset type being modified was a function of the relative degree of markedness of two factors: the cluster in question and the preceding sounds.

Another example of this kind of evidence for the MDH comes from a study by Benson (1988), in which she tested the performance of Vietnamese speakers on a number of onset and coda clusters in English. The data were elicited using a reading task in which the subjects produced single words, and the results were in conformity with the predictions of the MDH. The subjects' performance on the syllable-final clusters was in accord with the hypothesis, though the scores on the syllable-onset clusters exhibited ceiling effects due to the relatively high proficiency of the subjects.

Having discussed some of the evidence in the literature for the MDH, I now consider some of the methodological issues confronting the claims of the hypothesis.

Issues surrounding the Markedness Differential Hypothesis

The methodological issues that have confronted the MDH in the literature on L2 phonology stem from the fact that the MDH is completely programmatic with the Contrastive Analysis Hypothesis (CAH) in two important respects. First, both the MDH and the CAH make claims about L2 learning difficulty, and second, both hypotheses base their claims about such difficulty, at least in part, on the areas of difference between the NL and TL. I take up each of these issues in turn.

The fundamental prediction of the MDH is that linguistic representations in the TL that are both different and more marked than corresponding structures in the NL will cause learning difficulty. The obvious question then becomes how one measures learning difficulty. The vast majority of work in L2 phonology has calibrated difficulty in terms of learner errors: other things being equal, the more errors made on a structure, the more difficult that structure is interpreted to be. However, it has been recognized since the early days of Error Analysis (Schachter 1974) that learner errors are not the only measure of difficulty, and at times may not even be the most reliable measure. One hypothesis that has attempted to address this question is the Similarity Differential Rate Hypothesis (Major & Kim 1996). The central claim of this hypothesis is that rate of acquisition, rather than difficulty, is a more insightful measure of learning. As the Similarity Differential Rate Hypothesis is covered by Major (this volume), the reader is referred to the chapter by Major, and nothing more will be said about the SDRH here.

The second methodological issue confronting the MDH concerns the claim that NL-TL differences are crucial to the predictions of the MDH. To be sure, the CAH and the MDH differ in the amount of importance they place on NL-TL differences: for the former, such distinctions are paramount in that they are both necessary and sufficient to predict learning difficulty; for the latter, these differences are necessary but not sufficient. In addition to there being NL-TL differences, the claim is that typological markedness must be incorporated into the hypothesis as a measure of difficulty.

The problem with the predictions of the MDH being based on differences between NL-TL is that some reported error patterns corresponded directly to markedness principles, but the errors did not occur in an area of difference between the NL and TL. In this situation, the spirit of the MDH seemed to be invoked, in that more marked structures caused more errors than the corresponding less marked structures; however, the letter of the MDH prevented the hypothesis from making any predictions. That is to say, as stated, the MDH made predictions only when the marked and unmarked structures in question occurred in an area of difference between the NL and TL. If the structures in questions were found in both the NL and TL, then, as stated, the hypothesis made no prediction at all.

This type of pattern with respect to final obstruent devoicing was reported independently by Altenberg and Vago (1983) for Hungarian-speaking learners of English, and by Eckman (1984) for native speakers of Farsi learning English. In both studies it was shown that these L2 learners of English regularly devoiced word-final obstruents, an error pattern which involved a marked position of contrast, but which occurred in an area in which the NL and TL do not differ. English contains many words which exhibit a voice contrast in word-final obstruents, and both Hungarian and Farsi also have a word-final voice contrast in obstruents. In such cases, the MDH would predict that the L2 learners in question would be able to produce TL voice contrasts successfully by virtue of the similarity of such contrasts in the NL; however, this was not the case. Such data are, therefore, exceptional to the MDH in that it is reasonable for the hypothesis to account for these L2 utterances, but instead the errors lie outside the domain of the hypothesis.

The second type of fact that the MDH could not address was that it gave no prediction as to the kind of strategy the learner would employ when encountering a particular TL difficulty. The hypothesis could not explain why, in other words, L2 learners altered or simplified the marked structures in the way that they did, rather than in some other way. This point will be addressed below in the section on Future Directions.

To summarize this subsection briefly, L2 phonological research bearing on the MDH has concentrated largely on NL-TL differences in allowable coda consonants, and on distinctions in consonant clusters in onsets and codas. The reason for this focus may well lie in the fact that the typological generalizations that have been formulated about onsets and codas are relatively robust. As far as the results from these studies are concerned, none of them has uncovered evidence directly falsifying the claim that learners experience more difficulty with more marked structures than they do with corresponding less marked structures, though it is clear that evidence exists that runs counter to the spirit of the MDH, if not the letter. It is this type of evidence that constituted at least part of the motivation for the formulation of an alternative hypothesis, to which the focus now turns.

The Structural Conformity Hypothesis

The other hypothesis which invoked typological markedness, or at least the generalizations underlying markedness principles, is the Structural Conformity Hypothesis (SCH) (Eckman 1991), stated as in (4).

(4) The Structural Conformity Hypothesis (Eckman 1991:24)
The universal generalizations that hold for primary languages hold also for interlanguages.

The primary motivation for the SCH, as argued in Eckman (1996), is an L2 pattern, perhaps, but not necessarily, an error pattern, in which the L2 structures adhere to markedness principles, but the constructions in question are not an area of difference between the NL and TL. Since the pattern does not arise in an area of NL-TL difference, it is not explained by the MDH. One way to address this shortcoming was to eliminate NL-TL differences as a criterion for invoking markedness to explain the L2 learning facts. Essentially, then, the SCH is the result of stripping NL-TL differences from the statement of the MDH. If it is reasonable to assume that a learner will perform better on less marked structures relative to more marked structures, then the MDH can be seen as a special case of the SCH, namely, the case in which universal generalizations hold for the interlanguage (IL) in question, and the structures for which the generalizations hold are ones in which the NL and TL differ,

As stated in (4), the SCH is not formulated within a particular school of thought on language universals, and therefore would be programmatic with any research program invoking linguistic universals. The hypothesis simply asserts that interlanguages and primary languages are similar in at least one important respect: they both obey the same set of universal generalizations.[3]

The strongest kind of evidence that has been adduced in support of the SCH is an interlanguage pattern that is neither NL-like nor TL-like, but nevertheless obeys the kinds of universal patterns found in some of the world's languages. Examples of this kind of evidence have been reported in Eckman (1991), in Carlisle (1997, 1998) and in Eckman and Iverson (1994). Each of these studies considered the case of consonant clusters in onsets or codas, where the TL allowed both a greater number of clusters, as well as more marked clusters, than did the NL.

In Eckman (1991) the data were obtained using several elicitation tasks, including a free-conversation interview, from eleven ESL learners, four speakers each of Japanese and Korean and three speakers of Cantonese. The speakers' performance was analyzed using an 80%-threshold criterion to determine whether a given cluster type was part of a subject's IL grammar. This determination was then used to test the SCH using several universal generalizations about the co-occurrence of consonant cluster types in a language. Out of over five hundred individual tests, the hypothesis was shown to hold in all but five cases.

The studies by Carlisle (1997, 1998) also tested the occurrence of consonant clusters, but in the interlanguage grammars of Spanish-speaking learners of English. The specific hypotheses tested by Carlisle predicted that more marked clusters would be modified by the learners more frequently than related clusters that were less marked. The results supported the hypotheses in each case. Thus, Carlisle's studies supported the findings of Eckman (1991), but had the additional advantage of showing the operation of the SCH without imposing a criterial threshold on the data.

Eckman and Iverson (1994) analyzed English complex codas as produced in free conversation by native speakers of Japanese, Korean and Cantonese, languages which do not allow complex codas. The findings showed that the learners made more errors on the more marked codas than they did on the less marked ones. As a consequence, the respective IL grammars had the more marked cluster type only if they also exhibited the less marked type.

3. As stated, the SCH is neutral as to whether those universal generalizations fall within the context of Universal Grammar (UG), or are stated as typological generalizations. As a matter of practice, however, the SCH has been tested and invoked as an explanatory principle only within the context of typological universals.

What is common among studies reporting this kind of evidence in support of the SCH is that in each instance the IL grammars contained cluster types that were more complex than those allowed by the NL, but not as complex as those required by the TL. In this respect, the IL grammars fell between the NL and TL, but always did so in a way that was in conformity with the applicable universal generalizations.[4]

Having presented the kind of evidence adduced in support of the SCH, I now turn to the major methodological issue that has been raised with respect to this hypothesis, namely, whether it constitutes an explanation for the facts in question.

Issues surrounding the Structural Conformity Hypothesis

Despite the accounts given in the previous section claiming that the SCH has provided an explanation for a number of different facts about L2 phonology, it seems that some researchers in SLA have taken the position that markedness, in general, and the SCH, in particular, are not viable explanatory principles. There appear to be at least two arguments given for this position. The first is that markedness itself is simply a fact to be explained, and as such does not offer an explanation. This position is taken by Archibald (1998: 150) and is represented in (5).

(5) "My general assessment of this sort of typological universals approach to second language acquisition is that it provides an interesting *description* of the phenomena to be explained. I'm less sure of their (sic) status as an explanation of the observed facts. All in all, I prefer to assume some sort of structural explanation ..." (emphasis added)

The second counter-claim against invoking typological universals as explanatory principles asserts that invoking such generalizations raises more questions than it answers. This position was taken Gass and Selinker (2001: 154), and is quoted in (6).

(6) "For implicational universals to have any importance in the study of second language acquisition, two factors must be taken into consideration. First, one must understand why a universal is a universal. It is not sufficient to state that second languages obey natural language constraints because that is the way languages are. This only pushes the problem of explanation back one step."

I will consider each of these claims.

The important point that both of these criticisms miss is this: there are *levels* of scientific explanations, where the levels correspond to the generality of the laws

4. This same point is made in the study by Broselow and Finer (1991), but within the framework of parametric variation as allowed by UG.

invoked. To debate whether a generalization is a description or an explanation is to debate the level of explanation, not whether an explanation has been given. And to reject a hypothesis because it pushes the problem of explanation back one step misses the point that *all* hypotheses push the problem of explanation back one step–indeed, such "pushing back" is necessary if one is to proceed to higher level explanations.

To address these claims, let us take a brief look at the nature of scientific explanation. The goal will be to show that the accounts offered by markedness principles and the SCH for facts about L2 phonology do in fact constitute explanations.[5]

Scientists explain facts about the world by subsuming them under general laws. The fact to be explained is shown to be a specific instance of a more general phenomenon (Hempel & Oppenheim 1948). To take a concrete example, how do scientists explain that a rod or stick looks bent when it is partially submerged in a container of water? Or to consider a linguistics example, how do phonologists explain the fact that [t] alternates with [s] in the Finnish words in (7a & b) whereas the [t] in (7c) and (7d) does not?

(7)	a.	[haluta]	to want
	b.	[halusi]	wanted
	c.	[tila] *[sila]	room
	d.	[aiti] *[aisi]	mother

In the first example, scientists make reference to the laws regarding the velocity of light through different media, noting that light travels faster through air than it does through water, thus causing the partially submerged stick to appear bent. The appearance of the partially-submerged stick, therefore, is shown to be a particular case of a more general phenomenon, namely, the fact that light travels at different velocities in different media.

The explanation for the for the alternation in (7a & b) follows the same general pattern, except that it uses laws that refer to sound segments and phonological environments. Specifically, phonologists explain the alternations in question by appealing to a universal principle known as the Derived Environment Constraint (DEC) (Kiparsky (1982). The representations in (7a & b) motivate a rule (or some other construct) for Finnish that will account for the fact that [t] alternates with [s] before the vowel [i]. The DEC restricts this principle (and other similar rules or constructs) to apply only in what is called a derived environment, one in which a morpheme boundary separates the relevant segments, in this case, the [t] or [s] and the [i]. Thus, the alternation is licensed in (7a & b), but not in (7c) or (7d).

5. A more detailed account of explanation in SLA is given in Eckman (2004).

The same kind of explanation was given to explain why, for example, the L2 learners studied in Eckman (1991), cited above, evinced patterns of consonant clusters in onset and coda positions, where these clusters were not TL-like, nor were they licensed in the learners' NL. The SCH was invoked as a covering law in this case, claiming that the observed IL patterns that adhered to markedness generalizations about consonant clusters were a particular instance of a more general phenomenon, namely, IL grammars obeying universal generalizations.

The facts in the above examples are explained, then, by showing that they occurred in accordance with general laws. Now, there is an important point that was first raised by Hempel and Oppenheim (1948), and that bears crucially on the above statements in (5) and (6) by Archibald and by Gass and Selinker: the question of "why" can also be raised with respect to the general laws that are invoked as explanations. These laws, in other words, can come to be regarded as facts to be explained, and would be explained if one could subsume them under generalizations which are more comprehensive; that is to say, they would be explained if it were possible to deduce them from some more-encompassing laws or principles.

Given this background, it is important to recognize the following point: any proposed explanation of some phenomenon always engenders additional questions, because the generalizations serving as explanatory principles can also become the target of explanation. Thus, it is always the case that any explanation is adequate only to the extent of the current state of knowledge and understanding of the phenomenon under investigation.

It follows from this that there are levels of explanation, where "level" can be defined as the relative generality of the principles used in the explanation (Sanders 1974). In the context of the examples presented so far, any generalization from which it would be possible to derive the velocity of light in different media, or from which one could deduce the Derived Environment Constraint, or which would subsume the SCH, would constitute a higher-level explanation for those generalizations. It follows further that all empirical generalizations and hypotheses are, at the same time, a means for explaining lower-level generalizations, and the object of explanation for higher-level generalizations (Sanders 1974).

Some linguists may refer to the Derived Environment Constraint, the SCH and principles of markedness as descriptions of the facts rather than as an explanation. And based on the above discussion, these linguists would be partly correct and partly incorrect. They would be right in saying that these principles constitute a description of the facts in the sense that lower-level generalizations become facts for higher-level generalizations to explain. But these linguists would be incorrect in asserting that these principles are not explanations, because they are law-like statements which make testable predictions. Thus, a linguist who can propose a generalization from which the DEC follows, or who is able to formulate higher-order principles from which markedness generalizations are derivable, is justified in referring to the DEC or to markedness principles as facts, and not as explanations. However, it is sound scientific reasoning to reject the DEC and markedness generalizations as explanations *only* if one can then propose higher-level generalizations under which these principles can be subsumed. In the absence of more general principles, it is scientifically imprudent to reject the DEC and markedness generalizations as merely descriptions, because in so doing one would be left with no explanation at all.

This discussion now sets the context in which to reconsider the claims in (5) and (6) above about whether hypotheses that invoke markedness principles, such as the MDH and the SCH, are viable explanations.

Considering first the quotation in (5), I suggest that what Archibald has missed here is that typological universals are laws that subsume phenomena under a generalization, make predictions, and thus constitute an explanation. As was pointed out above, markedness principles, as is the case with all generalizations, can themselves be the target of explanation. To the extent that one can invoke a higher-order generalization under which to subsume the principle in question, it is possible to refer to the generalization as a fact; if one cannot offer a more-encompassing principle, then it is not scientifically sound to refer the law as a fact instead of an explanation.

A similar point can be made with respect to the claim in (6). If one does not accept a universal generalization as an explanation for L2 facts because such a generalization "pushes the problem of explanation back one step", one would never be able to accept **any** generalization as an explanation, because **all** generalizations, **all** hypotheses push the question back another step by raising further questions. Indeed, such questioning is necessary if our level of understanding is to deepen.

To summarize this subsection, the Structural Conformity Hypothesis addresses the shortcomings of the Markedness Differential Hypothesis, first, by making predictions about the nature of interlanguage grammars rather than about learning difficulty, and second, by expanding the domain of the hypothesis beyond only areas of difference between the NL and TL. The SCH asserts simply that ILs will obey the same universal generalizations as primary languages. The thrust of the explanation is that interlanguages are the way they because they are specific instances of a more general phenomenon, namely, human languages. Finally, although the universal generalizations and markedness relations, which serve as explanatory principles under the SCH, may themselves also be the target of explanation, this does not vitiate their standing as explanatory principles, because all scientific laws are, at the same time, explanations as well as facts to be explained.[6]

6. It is certainly true that the Hempel-Oppenheim model of explanation assumed for this discussion is not without its problems. However, space limitations prevent pursuing this point further here. For a fuller discussion of this topic, see Eckman (2004).

Given this discussion of the SCH within the broader context of what constitutes an explanation, the question naturally arises as to whether there is an important fact about L2 phonology that the SCH cannot account for, but that can be explained within an alternative framework. And here I reprise once again the case of word-final devoicing reported in the studies by Altenberg and Vago (1983) and Eckman (1984). What is particularly intriguing about this case is that the L2 learners in question evince a pattern that is not found in the NL, nor is it derivable from TL input, but it is nevertheless attested in the grammars of numerous languages of the world. This raises the question as to what the source for such regularities is. The SCH provides no explanation for whence these patterns might originate, as it simply allows whatever kinds of universal constraints and principles are found to govern primary languages to hold also for interlanguages. This question will be addressed below in the section on *Future Directions*.

Before concluding this chapter with a discussion of directions for future work on markedness and L2 phonology, I will consider briefly some of the methodological options and weaknesses involved with this approach, and suggest that the framework of Optimality Theory may well have a natural source for IL patterns that are not explainable on the basis of NL transfer or TL input.

Methodological options and weaknesses

The premise underlying the program of research that attempts to explain facts about second language phonology in terms of implicational universals is that these universal generalizations act as constraints on the L2 learner's internal IL grammar. In this section, I will outline two methodological options for testing the claims of this research program, and will also point out what might be seen as the major methodological weakness involved in conducting research within this paradigm.

The claim that the same universal generalizations that are true for L1 phonologies also hold for L2 phonologies is supported to the extent that it can be shown that L2 sound patterns do not violate the universal that have been formulated for L1 grammars. The most interesting test of this claim is one in which the IL patterns in question are not explainable either through the influence of the learner's NL, because the NL does not attest the pattern, or in terms of TL input, because the pattern is not part of TL. There are two methodological approaches to showing that the research hypothesis is supported. The first simply analyzes the L2 data with respect to the universal to determine whether the data conform to the universal. The second attempts to manipulate the learning of the TL patterns using the generalization as a strategy for intervention. I will consider each in turn.

The first option for testing the claim underlying this research is to select an implicational generalization against which the L2 data are to be analyzed and tested. The NL and TL need to be chosen strategically such that the generalization in question cannot be satisfied by virtue of the learner simply transferring patterns from the NL. This has been the logic of a number of studies over the years, including, for example, Eckman (1991), Eckman and Iverson (1994) and Carlisle (1997), with respect to onset and coda consonant clusters. In this work, the NL contained relatively few consonantal clusters while the TL, English, allowed clusters consisting of up to three or four consonants, depending on whether they occurred in the onset or the coda. The interesting tests for the hypothesis in question arise when the IL grammar satisfies the universal in question in a way that is different from both the NL and TL. This is what arose in the works cited above: the IL grammars in question allowed more in the way of consonantal clusters than did the NL, but at the same time attested fewer cluster patterns than those found in the TL. These differences were found while at the same time the IL grammars under study conformed to the universal generalizations being tested.

The second methodological option for testing the claims underlying this research program is to attempt to intervene in the development of an IL grammar using an implicational generalization as a strategy for intervention. More specifically, the goal is to train the learners in one of two experimental conditions, the first of which teaches the learner an IL grammar type that conforms to the universal in question, and the other condition attempts to induce the learner to acquire an IL grammar that will violate the universal. In the first experimental condition the learner is taught only the implicatum of the universal, and in the second experimental condition the learner is trained on only the implicans. The intent of the intervention is not that the learners trained on only the implicans will actually develop an IL evidencing only the implicans, which would contradict the universal, but rather, the expectation is that these learners will generalize their acquisition and acquire the implicatum also, making the IL conform to the universal in question. On the other hand, those subjects trained on only the implicatum, which targets an IL grammar type that is allowed by the universal, will not necessarily generalize their training to the implicans.

Most of the instructional studies in L2 acquisition that have employed this intervention strategy have been in the area of syntax, and virtually all have dealt with the acquisition of relative clauses (Gass 1982; Eckman, Bell & Nelson 1988; Doughty 1991). The one study of this kind in the area of L2 phonology that I am aware of is that by Eckman, Elreyes & Iverson (2003), in which specific phonemic contrasts were taught to L2 learners under the two experimental conditions outlined above. Learners who acquired the contrast in question that was the implicans of the universal generalized their learning to the include the implicatum, whereas the learners who were taught the contrast that was the implicatum did not necessarily generalize their learning to include the implicans.

Having outlined these two methodological options, I now turn to the methodological weakness surrounding this research program.

The most significant issue surrounding this research paradigm is establishing whether a particular universal is attested in an IL grammar. In order to test whether an implicational generalization claiming that a language having structure A also has structure B, but not vice versa holds also for some IL grammars, it is, of course, necessary to establish that the IL in question attests structures A and B.[7] This is normally done by arguing that the structures in question occur systematically in the utterances of the learner. The criterion for systematicity that is usually invoked is a relatively high frequency of occurrence for the structure in appropriate or obligatory contexts. The weakness of using this method for determining presence or absence of a structure is that the criterion threshold, generally taken to be 80% occurrence, is arbitrary.

Future directions

Having argued in the preceding sections that universal, typological generalizations and markedness relations can be used in the explanation of various facts about L2 phonology, I now face the question of whether markedness principles can be naturally incorporated into a theory of language, a question which has arisen from time to time over the years within the SLA literature (Flynn 1987; White 1987). Until recently, phonological theories have had difficulty incorporating markedness principles and generalizations in any natural way. Although there seems to have been recognition over the decades that markedness generalizations are an important component of phonological theory, markedness principles appear to have been little more than appendages tacked on to the theory, almost as an after-thought. In fact, in one of the major phonological works in the last few decades, (Chomsky & Halle 1968) *The Sound Pattern of English*, markedness is treated in the very last chapter of the book, under the heading of *Epilogue and Prologue*.

To date, the only phonological theory, with the possible exception of Natural Phonology (Stampe 1979), to explicitly and intrinsically incorporate markedness is Optimality Theory (Prince & Smolensky 1993). Because Optimality Theory (OT) is discussed in detail by Hancin-Bhatt (this volume), its treatment here will

7. To be sure, this same question arises in work on universals involving primary languages, but in many cases the problem is diminished by the fact that researchers can determine the existence of a certain pattern in a language by consulting written grammars, though, admittedly, this simply pushes the issue back another level.

necessarily be brief. There are, however, three important features of OT that are worth pointing out within the context of this chapter, viz., that OT grammars all consist of a universal set of constraints, that grammars of particular languages result from different rankings of these universal constraints, and that the constraints are divided into two subgroups, markedness constraints and faithfulness constraints. Each of these will be considered in turn.

First, a way to view the universal constraints is as a set of criteria for well-formedness. From the stipulation that all constraints are universal, and that grammars differ only in the particular ranking of the universal constraints, the theory makes the claim that well-formedness criteria do not differ from language to language; rather, what varies across languages is how these criteria are applied, that is, how they are ranked. OT is thus inherently a theory of typology: any ranking of the universal constraints should yield a grammar of a language, and any grammar of a language should conform to one of the possible rankings of the constraints.

Second, given that the goal of a grammar is to specify all and only the well-formed utterances in the language, or in the case of phonologies, all the well-formed pronunciations, OT grammars and rule-based grammars accomplish this aim differently. Rule-based grammars begin with the lexical representation of an utterance and execute a derivation, applying the appropriate rules to the lexical representation, making the changes specified by the rules, producing intermediate representations to which other rules are applied, and continuing until all of the applicable rules have been brought to bear, and the output is specified. The well-formed utterances of the language are predicted to be all and only those which can be successfully derived using the rules of the grammar. An ill-formed, or ungrammatical, utterance is characterized by showing that its derivation violates one or more of the rules of the grammar. On the other hand, the constraints of an OT grammar are violable; no single utterance can satisfy all of the universal set of criteria for well-formedness. Within OT, therefore, grammaticality is not characterized on the basis of whether or not an utterance violates one or more of the constraints; instead, the grammaticality of an utterance is determined by an optimization procedure whereby well-formed utterances are those that conform to the highest ranked constraints in the grammar.

This leads to the third important feature of OT within the context of this chapter, that the set of universal constraints is divided into two categories, faithfulness constraints and markedness constraints. Interestingly, this division has been cited as corresponding, roughly and respectively, to the notions contrast and articulatory ease (Gundel et al. 1986). The important point, from our perspective, is that, within OT, markedness is incorporated as a basic tenet of the theory.

Incorporating markedness into the general theory through the ranking of the universal set of constraints provides a natural explanation for the kind of L2 phenomenon that has previously eluded explanation, namely, an IL pattern, attributable to neither the NL nor TL, but one that is nonetheless attested in the grammars of at least some of the world's languages. Within an OT framework, this situation follows naturally from the tenet that the set of constraints is universal, and therefore is present in the grammars of all languages. In the case under consideration, the fact that neither the NL nor the TL evidences word-final devoicing is because the constraints that characterize such devoicing within OT grammars, though present in the grammars of both the NL and TL, are ranked low in these grammars. Consequently, the word-final devoicing constraints are not determinant in characterizing the utterances of either the NL or TL.

Now, if it is assumed that interlanguage grammars are also characterized by a ranking of these universal constraints, then a clear source for the observed word-final devoicing pattern emerges. If the constraints in the IL grammar are ranked differently than those in the NL or TL, then the possibility exists that the constraints that characterize devoicing could become determinant, producing the observed pattern.

Two interesting consequences follow from this view of interlanguage grammars within an OT framework. First, one should expect the error patterns observed in L2 utterances to be attested as structures in the grammars of at least some other languages. This is true because IL grammars consist of the same constraints, though perhaps with different rankings of those constraints, as other languages. And second, interlanguage grammars are predicted to differ from primary language grammars in the same way that primary language grammars differ from each other. This follows because IL grammars are characterized using the same constructs, viz., constraints and constraint rankings, as are used in primary language grammars.

Conclusion

This chapter has attempted to argue the following points: that typological markedness has played a significant role in the explanation of facts about L2 phonology; that markedness generalizations are explanatory principles in the sense of being covering laws under which phenomena can be subsumed; and that markedness will continue to play a significant role in L2 phonology within the framework of Optimality Theory.

Suggested readings

Archibald, J. 1998. *Second language phonology*. Amsterdam: Benjamins.

Battistella, E. 1990. *Markedness: The evaluative superstructure of language*. Albany: The State University of New York Press.

Broselow, E., Chen, S., & Wang, C. 1998. The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition* 20: 261–280.

Carlisle, R. 1998. The acquisition of onsets in a markedness relationship: A longitudinal study. *Studies in Second Language Acquisition* 20: 245–260.

Chomsky, N., & Halle, M. 1968. *The sound pattern of English*. New York: Harper and Row.

Eckman, F. 1991. The Structural Conformity Hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition* 13: 23–41.

Hancin- Bhatt, B. 2000. Optimality in second language phonology: Codas in Thai ESL. *Second Language Research* 16: 201–232.

Moravcsik, E., & Wirth, J. 1986. Markedness: An overview. In F. Eckman, E. Moravcsik & J. Wirth (eds.), *Markedness*, New York: Plenum, 1–11.

References

Altenberg, E., & Vago, R. 1983. Theoretical implications of an error analysis of second language phonology production. *Language Learning* 33: 427–447.

Anderson, J. 1987. The markedness differential hypothesis and syllable structure difficulty. In G. Ioup & S. Weinberger (eds.), *Interlanguage Phonology: The Acquisition of a Second Language Sound System*, New York: Newbury House/Harper & Row, 279–291.

Archibald, J. 1998. *Second language phonology.* Philadelphia: John Benjamins.

Battistella, E. 1990. *Markedness: The evaluative superstructure of language.* Albany: The State University of New York Press.

Benson, B. 1988. Universal preference for the open syllable as an independent process in interlanguage phonology. *Language Learning* 38: 221–242.

Broselow, E., & Finer, D. 1991. Parameter setting in second language phonology and syntax. *Second Language Research* 7: 35–59.

Carlisle, R. S. 1991. The influence of environment on vowel epenthesis in Spanish/English interphonology. *Applied Linguistics* 12: 76–95.

Carlisle, R. S. 1997. The modification of onsets in a markedness relationship: Testing the Interlanguage Structural Conformity Hypothesis. *Language Learning* 47: 327–361.

Carlisle, R. S. 1998. The acquisition of onsets in a markedness relationship: A longitudinal study. *Studies in Second Language Acquisition* 20: 245–260.

Chomsky, N., & Halle, M. 1968. *The sound pattern of English.* New York: Harper & Row.

Doughty, C. 1991. Second language instruction does make a difference. *Studies in Second Language Acquisition* 13: 431–469.

Eckman, F. 1977. Markedness and the contrastive analysis hypothesis. *Language Learning* 27: 315–330.

Eckman, F. 1981a. On predicting phonological difficulty in second language acquisition. *Studies in Second Language Acquisition* 4: 18–30.

Eckman, F. 1981b. On the naturalness of interlanguage phonological rules. *Language Learning* 31: 195–216.

Eckman, F. 1984. Universals, typologies and interlanguages. In W. E. Rutherford (ed.), *Language universals and second language acquisition*, Philadelphia: John Benjamins, 79–105.

Eckman, F. 1991. The Structural Conformity Hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition* 13: 23–41.

Eckman, F. 1996. A functional-typological approach to second language acquisition theory. In W. C. Ritchie & T. K. Bhatia (eds.), *Handbook of second language acquisition*, San Diego: Academic Press, 195–211.

Eckman, F. 2004. Universals, innateness and explanation in second language acquisition. *Studies in Language* 28: 682–703.

Eckman, F., Bell, L., & Nelson, D. 1988. On the generalization of relative clause instruction in the acquisition of English as a second language. *Applied Linguistics* 9: 1–20.

Eckman, F., Elreyes, A., & Iverson, G. 2003. Some principles of second language phonology. *Second Language Research* 19: 169–208.

Eckman, F., & Iverson, G. 1994. Pronunciation difficulties in ESL: Coda consonants in English interlanguage. In M. Yavas (ed.), *First and second language phonology*, San Diego: Singular Publishing Group, 251–266.

Edge, B. 1991. The production of word-final voiced obstruents in English by speakers of Japanese and Cantonese. *Studies in Second Language Acquisition* 13: 377–393.

Flynn, S. 1987. Contrast and construction in a parameter-setting model of second language acquisition. *Language Learning* 37: 19–62.

Gass, S. 1982. From theory to practice. In M Hines & W. E. Rutherford (eds.), *On TESOL*, Washington DC: TESOL, 129–139.

Gass, S., & Selinker, L. 2001. *Second language acquisition: An introductory course.* Mahwah, NJ: Lawrence Erlbaum Associates.

Greenberg, J. 1976. *Language universals.* The Hague: Mouton.

Gundel, J., Houlihan, K., & Sanders, G. 1986. Markedness distribution in phonology and syntax. In F. Eckman, E. Moravcsik, & J. Wirth (eds.), *Markedness*, New York: Plenum Press, 107–138.

Hancin-Bhatt, B. 2008. Second language phonology in optimality theory. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Hansen, J. 2001. Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics* 22: 338–365.

Hansen, J. 2004. Developmental sequences in the acquisition of English L2 syllable codas: A preliminary study. *Studies in Second Language Acquisition* 26: 85–124.

Hempel, C., & Oppenheim, J. 1948. Studies in the philosophy of science. *Philosophy of Science* XV: 135–175.

Jakobson, R. 1941. *Child language, aphasia, and phonological universals*. The Hague: Mouton.

Kiparsky, P. 1982. Lexical phonology and morphology. In I. S. Yang (ed.), *Linguistics in the Morning Calm*, Seoul: Hanshin, 3–91.

Lado, R. 1957. *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.

Major, R. C. 2008. Transfer in second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Major, R., & Kim, E. 1996. The similarity differential rate hypothesis. *Language Learning* 46: 465–496.

Moravcsik, E., & Wirth, J. 1986. Markedness: An overview. In F. Eckman, E. Moravcsik, & J. Wirth (eds.), *Markedness*, New York: Plenum, 1–11.

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

Prince, A., & Smolensky, P. 1993. *Optimality theory: Constraint interaction in a generative grammar*. Rutgers University Optimality Archive.

Sanders, G. 1974. Introduction. In D. Cohen (ed.), *Explaining Linguistic Phenomena*, New York: John Wiley & Sons, 3–20.

Sato, C. 1984. Phonological processes in second language acquisition: Another look at interlanguage syllable structure. *Language Learning* 34: 43–57.

Schachter, J. 1974. An error in error analysis. *Language Learning* 24: 205–214.

Stampe, D. 1979. *A dissertation in natural phonology.* New York: Garland.

Stockwell, R., & Bowen, J. 1965. *The sounds of English and Spanish*. Chicago: University of Chicago Press.

Trubetzkoy, N. 1939. *Principles of phonology.* Paris: Klincksieck.

White, L. 1987. Markedness and second language acquisition: The question of transfer. *Studies in Second Language Acquisition* 9:261–286.

CHAPTER 5

Second language phonology in optimality theory

Barbara Hancin-Bhatt
University of Illinois

Introduction

The field of second language (hereafter, L2) phonology dates back at least to Weinreich's (1953) and Lado's (1957) work, which addressed, in part, how L2 sounds are constrained within a first language-based phonology. Despite these 50 years of thinking about L2 sound patterns and substantial research that identifies characteristics of these patterns, there is, to date, no singular model of the L2 sound system that has been widely adopted to understand the myriad results observed in L2 sound pattern research. The foremost challenges for model-theoretic proposals are that of *generalizability*, particularly across the levels of phonology; *accountability* for the range of possible structures that occur in L2 sound patterns; and *predictability*, particularly with regard to how input triggers a restructuring of the grammar. (See Grosjean 1998 and White 2000 for similar claims on model building in L2 studies.)

To overcome these challenges, a richly defined model of the interlanguage (phonological) grammar and how it is accessed and restructured should substantially address, define and explain the following issues:

(1) Model-theoretic Proposals on L2 Sound Patterns

Grammar Representation:	What are the assumed features (at all phonological levels) of the initial and subsequent grammatical states?
Acquisition:	What are the inputs and representations that force a restructuring in the grammar?
Variable Competence:	How does the grammar accommodate multiple grammatical representations, as exhibited in task-based differences, for a given input?

In this chapter, I explore one theory's account of grammatical representations, their restructurings and variations and apply this to some of the research findings in the field of L2 phonology. After briefly reviewing major findings and some of the outstanding problems in the field, I introduce the basic principles of Optimality Theory (OT) and then show how it has been used in recent L2 studies. I conclude this paper with speculations on where future studies in OT can advance our understanding and research in the field of L2 phonology.

Outstanding problems in L2 phonology

L2 sound pattern research has, to a certain extent, advanced proposals on what defines the interlanguage grammar and how that grammar gets accessed during perception and production. The following is a brief review of some rather robust findings and outstanding issues regarding what constitutes the L2 learner's knowledge and how it is accessed and impacted by experience. A model of the interlanguage sound system must be able to address these issues before it can be widely accepted as a framework for researching and understanding learner patterns.

Knowledge representations

Markedness, transfer and their interaction

The standard paradigms of L2 research assume some level of language transfer or influence of the L1 grammar on L2 sound patterns. Essentially, L2 structures that are similar or the same as their counterparts in the L1 can have a generally facilitative effect in learning, while L2 structures that are not present in the L1 grammar provide a substantial challenge.[1] In addition to language transfer, L2 acquisition is significantly influenced by developmental effects that can be captured by markedness generalizations in natural languages. For example, syllables are simplified to CV structures in early language (L1 and L2) acquisition, and this has been linked to the common occurrence of CV syllables crosslinguistically. Eckman (Chapter 4, this volume) discusses the substantial role of markedness in L2 sound pattern research.

Furthermore, both markedness and language transfer effects have been shown to have varying influences on interlanguage grammar throughout the stages of acquisition, such that transfer tends to be more prominent during early stages of acquisition, while developmental effects emerge over time and, in most cases,

1. For the sake of economy, this description of language transfer and markedness effects in learning effect is simplified. See chapters by Eckman and Major, this volume, for a thorough review.

overcome transfer effects (cf. Major 1986, 1994 and Chapter 3, this volume). While empirical evidence for these effects is substantial, linguistic-theoretic analyses providing an account of these effects and how they interact over time are few, thereby limiting these analyses' predictability (cf. Hancin-Bhatt & Bhatt 1997).

Role of prosodic constraints
Another area of inquiry under development is the role of prosodic constraints in L2 sound pattern acquisition. Prosodic constraints are characterized as the influence of one level of phonology on another. For example, certain segments in a language may be restricted to specific syllable positions, such as the /h/ occurring only syllable initially in English, while other segments are less restricted, such as the /r/ occurring in any syllable position in English. This cross-level dependency impacts learning. Young-Scholten and Archibald (2000) suggest that there may be an implicational hierarchy in learning L2 sound patterns such that L2 segments are learned before L2 syllable structure and prosodic licensing effects. The challenge is to encode cross-level dependencies within a model of L2 phonological representations such that any observed implicational hierarchies emerge as a natural consequence of the system during acquisition.

States of interlanguage grammar
Also outstanding in L2 sound pattern research is a model that captures the effects stated above over time: What are the assumed representations of the interlanguage grammar in the initial state, subsequent states, and what triggers a restructuring within these representations, forcing a move from state to state? While there has been some discussion on what types of input trigger acquisition, there has been little said about what defines the actual states (or stages) of a learner's grammar and precisely how those grammars are restructured. Without a specific proposal on how the grammar is represented and what forces its restructuring, a model's predictability is minimal at best.

Knowledge access effects

L2 sound pattern researchers have made specific observations on how the learner's knowledge is accessed during perception and production, noting that task-based effects can substantially complicate our understanding of what the learner knows of the L2. Not only have task-based differences created a notion of learner 'variable competence,' but they also affect how the L2 knowledge develops. This variability within L2 sound patterns has either gone unaccounted for within models of the L2 grammar or has been given special, theory-external status.

Input: Misperceptions impact L2 grammar

L2 perception researchers provide evidence that L2 sound contrasts are difficult to perceive, but not uniformly so, depending on the cues used for contrasts in the first language and general acoustic salience of the cue. These findings are reviewed in the chapters by Strange and Shafer (Chapter 6) and Bradlow (Chapter 10), this volume. What is perceived, then, should in principle affect the L2 knowledge representation. In addition to auditory cues, researchers have also argued for the impact of visual cues on learner knowledge, providing evidence of orthographic miscues, called 'spelling pronunciations,' (Altenburg & Vago 1987; Young-Scholten 1995, 1997) and of facial expression cues (Hardison 1996). This research underscores the need for specificity in model development, particularly with regard to what constitutes the input: what is the range of forms that the input can take and how do these impact the grammar?

Output: Task-based effects

Learners exhibit substantial variability across tasks that constrain the amount of attention that can be given to articulatory control. Evidence shows that reading aloud, where little semantic/syntactic interpretation is required and more attention can be paid to articulatory control, generates fewer errors than spontaneous speech, where demands on semantic/syntactic parsing are high and thus less attention is available for articulatory control. Similarly, formal/careful speech produces fewer errors than casual speech. These task-based effects have been observed throughout the history of L2 research (e.g., Dickerson 1975; Tarone 1978, 1982, 1983; Zampini 1994), and contribute to the notion of learner 'variable competence.' Accounting for this variability within a model of the interlanguage grammar has essentially been left undone.

Perception – production asymmetry

Another longstanding observation has been that L2 perceptual abilities do not match L2 production abilities (cf., Flege 1993 for a review). In perception, listeners attend to acoustic phonetic features of sounds to identify them, while in production, talkers produce specific articulatory configurations to distinguish sounds from each other. Generally, there is evidence that L2 learners can have highly accurate perceptual abilities, but relatively inaccurate production ones. Alternatively, L2 learner production abilities can be more target-like than their perceptual abilities at certain levels of the phonology. Not only do perception and production require different primitives, but they also can have a differential rate of development, as discussed more fully in Bradlow (Chapter 10, this volume). The different cues and skills used in perception and production complicate our access to and understanding of the learner's knowledge, and has provided a formidable challenge for L2 researchers working within specific grammatical frameworks that do not allow for a range of inputs into the interlanguage grammar.

While this discussion of some major findings in L2 phonology has been brief, it provides a basis on which to explore a model's potential for advancing understanding within the field. The remainder of this paper offers Optimality Theory as a productive option within which to address outstanding concerns and to generate new research on the developing bilingual phonological grammar.

Basics of optimality theory

Why optimality theory?

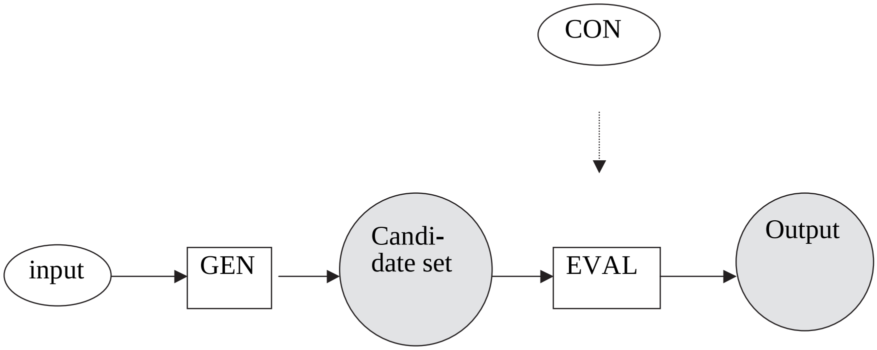
Before exploring any model as a potential account of L2 phenomena, it is prudent to ask why, on a general level, the model should even be considered. To begin, OT is a generalized theory of grammar, with a focus on specifying the interactions of grammatical principles. It is, therefore, not uniquely a theory of phonology, nor of acquisition, but a theory that assumes a language faculty designed to manage and accommodate a full range of linguistic inputs, idealized or not. As a model theoretically applicable to all levels of the grammar, OT does not require different basic assumptions on how the linguistic input types are parsed. The theory assumes a restricted set of generalized operations that affect whatever the linguistic input is, thereby reflecting one basic assumption on economy in cognitive design.

Furthermore, OT encodes Universal Grammar in the range of candidate structural descriptions it admits, as well as in the constraints that evaluate those descriptions, thereby, in principle, delimiting the range of linguistic structures that can occur crosslinguistically. Variation results simply from differences in the prominence of the given constraints in different grammars. An OT grammar, therefore, does not have to assume special rules or representations to accommodate L2 input that cannot be parsed within an L1 grammar. That is, this feature of OT obviates the need to design special rules for L2 learner grammars. Furthermore, it begins to acknowledge and encode how language variation is a natural consequence of a dynamic system, particularly during acquisition. L2 evidence, then, is no longer considered 'extralinguistic phenomona,' rather it is a natural consequence of a dynamic system, a grammar attempting to accommodate a new form.

Grammatical representations

The OT grammar assumes an input-output design in which a linguistic input, commonly a lexical representation, is parsed, not by rules, but by a universal set of constraints on the well-formedness of linguistic structures. The basic design of the OT grammar is as follows.

(2) Diagram of an OT Grammar (Archangeli 1999)



A linguistic input enters the grammar and a set of candidate structural descriptions for that input is automatically generated via a function known as the Generator, or GEN. These universal structural descriptions are then evaluated according to how well they conform to the set of universal constraints (CON) on linguistic structures. The harmony evaluations are determined by the evaluator, or EVAL, which is guided by hierarchical ranking of the universal constraints. The candidate description that is evaluated to be the most harmonic is the optimal output for the given input.

The constraints in EVAL are universal and violable, indicating that all languages are subject to the constraint generalizations, but they are not always respected across all languages. Two families of constraints conspire to determine optimal forms in a grammar: MARKEDNESS and FAITHFULNESS. MARKEDNESS constraints capture generalizations on linguistic structures that commonly occur in natural languages (unmarked) and those that do not (marked). FAITHFULNESS constraints are those that ensure congruence between the structures that form the input into the grammar and those that are in the output. That is, this family of constraints wants every input sound to have a corresponding output sound, which is identical and in the same position. Examples of faithfulness constraints are given in (3).

(3) Two Families of Constraints in OT
Markedness Constraints

A. ONSET. Syllables must have onsets.
B. NOCODA. Syllables must not have codas.
C. *COMPLEX. Only one C or V may associate to any syllable position node.
D. *VOICED-CODA. Obstruents must not be voiced in coda position.
E. *V_{NASAL}. Vowels must not be nasals.

Faithfulness Constraints

A. Max-IO. The output must preserve all segments present in the input. (No deletion)
B. Dep-IO. Output segments must have counterparts in the input. (No epenthesis)
C. Ident-IO(F). Output segments and corresponding input segments must share values for a feature [F]. (No feature-changing/substitutions.)

Together, these constraints conspire to determine which of the candidate descriptions is most optimal, for a given input, meaning which candidate, among the choices, has the least serious constraint violations. The constraints are inherently conflicting, but they are ranked relative to each other, with the principle that the higher the ranking of a constraint, the more serious its violation or, alternatively, the more strongly it holds in a language. Constraint rankings are language-specific, which accounts for the variation in structures observed cross-linguistically.

The core features of constraints are summarized in (4).

(4) Core assumptions on constraints in OT (Kager 1999)
Universality: Constraints are universal.
Violability: Constraints are violable.
Optimality: An output is 'optimal' when it incurs the least serious violations of a set of ranked constraints.
Domination: The higher-ranked of a pair of conflicting constraints takes precedence over the lower-ranked one.

A brief illustration of a tableau representing harmony evaluation is given in (5). The asterisks denote a violation of the constraint and an asterisk with an exclamation point is fatal, signifying that the violation disqualifies the candidate from being the optimal one in the evaluation.

(5) Sample Tableau

/input/	Constraint 1	Constraint 2	Constraint 3
Description A	*!		
Description B		*	*!
☞ Description C		*	

/CVCC/	*Complex	Dep-IO	Max-IO
CVCC	*!		
CVCCV		*!	
☞ CV<C>C			*

Ranking: *Complex >> Dep-IO >> Max-IO

The tableau shows that *Complex dominates Dep-IO, which dominates Max-IO. The input CVCC generates multiple candidate structural descriptions, and, for purposes of illustration, we consider three potential candidates here, namely a syllable with a complex coda (CVCC), and two that simplify the complex coda: one parse that epenthesizes a segment/nucleus to yield two syllables (CVCCV), and a parse that deletes a segment yielding a single syllable structure (CV<C>C). The first candidate, CVCC, fails to be the optimal one because it violates the highest ranking constraint, *Complex. The second candidate fails because it violates a higher ranking constraint than the third. Thus, the optimal candidate is the one that has the least serious constraint violations.

Acquisition algorithms

Researchers attempt to understand language acquisition from different perspectives, the two most common being: 1) eliciting and codifying data from actual acquisition situations and then hypothesizing what acquisition entails, including implications for cognitive design, and 2) focusing on a theory of cognitive design that establishes the learnability of a grammar, gleaning evidence from a variety of linguistic contexts. Studies in OT exemplify the second of these perspectives, as the theory defines the universal and language-specific features of a grammar, and current research examines exactly how the grammar responds to and changes when linguistic inputs are introduced. OT, more so than previous linguistic models, requires the definition of how a grammar responds to new or unexpected, non-idealized input, thereby necessitating a theory of acquisition within the model of linguistic representations. Since, in OT, variation observed across languages is captured by variant constraint rankings, the job of a language learner is to infer the specific constraint ranking of the target language, based on what is perceived.

In OT, learning is a function of an algorithm that forces constraint re-rankings based on mismatches between the input and the optimal output, thereby giving feedback a central role in the developing system.[2] In essence, an OT learning algorithm shows how it is possible for a learner to deduce rankings of constraints based on an output form and the universal constraints. Two algorithms are currently receiving a good deal of attention: The Constraint Demotion Algorithm (CDA) of Tesar & Smolensky (1996, 2000) and the Gradual Learning Algorithm (Boersma 1997; Boersma & Hayes 1999), and the competing proposals are eliciting many new questions in the area of acquisition. Due to space constraints, I only discuss the basic assumptions of the algorithms.

2. For a comparison of OT and a Principles and Parameters learning algorithm (e.g., Gibson & Wexler 1994), see Boersma, Dekkers and van de Weijer (2000).

These learning algorithms have been developed assuming a first language (L1) acquisition context, and posit that the initial state of the learner is a set of unranked, undominated constraints.[3] In the CDA, for example, by comparing the constraint violations of the optimal form (or winner) with those of the suboptimal forms (or losers), the grammar determines which constraints should be demoted. The algorithm, then, is error-driven, such that violations in the optimal output trigger a demotion of the violated constraint. Thus, a constraint demotion provides a new dominance hierarchy that reflects the input more closely, and reflects a new state in the learning of the target grammar. [Other algorithms allow promotion of constraints based on feedback from the optimal output comparison.]

(6) Basic Assumptions of Constraint Demotion Algorithm.
Initial State: Unranked, undominated constraint hierarchy[4]
Stages of acquisition: Constraint re-rankings
Triggers for re-ranking: Positive evidence of a constraint violation in the optimal output. (Constraints violated in output are dominated and need to be re-ranked.)
Robust Interpretive Parsing: Process by which grammar deduces the hierarchical ranking responsible for a non-harmonic input-output pair.
Constraint Demotion: When comparison of optimal to suboptimal candidates requires, constraints are demoted immediately below the constraint that induces its violation in the optimal output.
Target State: Constraint ranking that represents the least serious violations in an optimal output.

To summarize this brief overview, in OT, *knowledge* of a language consists of the universal set of structural descriptions, the universal set of constraints on these descriptions and the language-particular constraint ranking. *Acquisition* is a problem of learning the constraint rankings that hold for the target language.

Research findings in L2 phonology and OT

OT was formally introduced a decade ago with Prince and Smolensky's 1993 manuscript (Prince & Smolensky 2004), and since then text books, edited volumes and books, and numerous articles in journals and on the electronic archive at Rutgers University have been published, all devoted to testing and developing

3. Not all proposals assume an initially unranked set of constraints. See, e.g., Boersma, Escudero, & Hayes 2003 for an alternative proposal.

4. Note that some proposals include rankings in the initial state to reflect cross-linguistic ranking tendencies.

the paradigm. OT has had the greatest impact in phonology, but there is substantial development in the areas of syntax and discourse studies, as well. The theory's substantial impact in such a relatively short time reflects its complexity and range. Unfortunately, however, OT has, to date, had little impact on the field of L2 acquisition. Relatively few studies in L2 phonology have adopted an OT approach to understand observed phenomena in L2 learning, and the majority of these studies focus on L2 syllables. The remainder of this paper is an attempt to highlight how OT has been adopted to account for some of the outstanding issues in L2 phonology raised earlier and demonstrate promising directions for future research.

Markedness effects, transfer and their interaction

The role of markedness and language transfer effects, and their interaction, has received increasing attention in L2 phonology in the past couple of decades, but specific proposals that characterized this interaction were L2-studies-specific and, as such, did not gain widespread visibility and thus have not been tested and refined. By adopting an OT framework, however, the interaction of language transfer and markedness effects can be explained within a theory of general grammar design, with the interaction developing as a consequence of the grammar design, thereby obviating the need for special rules or characterizations for the developing bilingual grammar.

Role of markedness constraints

Eckman (1977) codified the observation that L2 learners are more likely to acquire unmarked structures more easily than marked ones in what he termed the Markedness Differential Hypothesis. Subsequent work has supported the finding that, in L2 learning, marked structures are more likely to be problematic than unmarked ones (cf., Eckman & Iverson 1993). In spite of the fairly robust evidence that markedness conditions exist and constrain learning, rule-based models of the grammar have not captured this effect in an efficient manner, i.e. within the basic mechanisms posited for the grammar (parameters, etc.). Reflecting on this observation, Broselow, Chen and Wang (1998), turn to Optimality Theory to explain why L2 learners produce unmarked structures that are not present in either their L1 or their L2. They argue that OT provides a logical explanation of how an unmarked structure can surface when it is not obviously from the L1 or L2, without having to posit any special rules or conditions to accommodate the L2 learning situation.

Broselow et al. (1998) looked at data from 10 Mandarin-speakers learning English. One of the features they examined was final devoicing. Mandarin does not allow $CVC_{(stop)}$ syllables, while English does. Mandarin speakers learning English resort to epenthesis and deletion to resolve this dissonance between their native and target grammars, but they also have used another strategy, which is to produce the $CVC_{(stop)}$ with a devoiced stop (final devoicing). Devoicing is an unexpected outcome, since neither the native nor the target language has a final consonant devoicing rule.

For their analysis, Broselow et al. adopt the following constraints.

(7) Broselow et al.'s (1998: 267–275) constraints: Emergence of the Unmarked
Markedness
No Voiced Obs Coda: Syllable codas may not contain voiced obstruents.
No Obs Coda: Syllables may not contain obstruents.
Faithfulness
Max i: Maximize the consonants in the input. (Do not delete consonants.)
Dep (v): The vowels in the output should be dependent on the input. (Do not add vowels.)
Ident (voi): An output segment should be identical in voicing to the corresponding input segment.

The following sets of rankings of these constraints exemplify three coda patterns observed cross-linguistically:

(8) Crosslinguistic Variation in Codas
A. No coda obstruents (Mandarin): *vɪg, *vɪk
No Obs Coda, No Voiced Obs Coda >> Max i, Dep (v), Ident (voi)
B. Voiceless coda obstruents only (German): *vɪg, vɪk
No Voiced Obs Coda >> Max i, Dep (v) >> Ident (voi) >> No Obs Coda
C. Voiced and voiceless coda obstruents (English): vɪg, vɪk
Max i, Dep (v), Ident (voi) >> No Obs Coda, No Voiced Obs Coda

The rankings in 8.A. reflect Mandarin's avoidance of coda obstruents. When Mandarin speakers confront English words that contain a coda obstruent, then it is predicted that learners will avoid the obstruent placement through either deletion or epenthesis because the Faithfulness constraints – Max i and Dep (v) – are ranked lower than the markedness constraints. Broselow et al.'s data support this. However, their data also indicate that some of the speakers are devoicing the coda, producing an unmarked form that does not appear in either their Mandarin or English. Broselow et al. explain this by saying that the devoicing speakers have reranked the No Obs Coda constraint relative to the No Voiced Obs Coda. That grammars choose a form that violates the voicing feature correspondence between an output and input segment indicates that Ident (voi) is the minimal violation within the family of Faithfulness constraints. The representational evaluation of /vig/ in Mandarin is represented in (9).[5]

(9) Devoicing in Obstruent Codas: No Voic Obs Coda >> Max, Dep >> Ident (Voi)

/vɪg/	*No Voic Obs Coda	Max, Dep	Ident (Voi)
vɪg	*!		
vɪ<g>		*!	
vɪ.gə		*!	
☞ vɪk			*

To summarize, then, a tendency toward the unmarked is captured within the basic tenets of the OT grammar through the family of constraints capturing markedness generalizations. These constraints are assumed to be part of the universal grammar and, thus, do not have to be "learned". By simply re-ranking given constraints, Broselow et al. show how an unmarked structure, devoiced final obstruent, ***despite its non-obvious existence in either the L1 or the target language***, can surface in interlanguage development without appealing to any extra-grammatical, functions or special rules.

Role of L1 transfer and its interaction with UG

One of the first research interests in the field of L2 phonology was exploring how the L1 affects L2 sound patterns. Language transfer has continued to appear as a topic in L2 phonology, and researchers continue to seek to explain why transfer occurs, how it occurs, and how it is eventually overcome. Models include L1–L2 feature matching algorithms at the segmental level (Best & Strange 1992; Brown 2000; Flege 1987, 1990; Hancin-Bhatt 1994; Hancin-Bhatt & Govindjee 1999), structural transfer beyond the segment, for example, in syllable structure, syllabification and licensing (cf. Broselow 1987; Sato 1987; Broselow & Finer 1991; Young-Scholten & Archibald 2000); at level of stress (Archibald 1993); in tone/intonation (Broselow, Hurtig & Ringen 1987), and proposals on how transfer and markedness (or developmental) effects interact (e.g., Major 1986, 1994).

While transfer and markedness effects are observed in L2 sound patterns at various levels, there have been few attempts that try to capture these generalizations within a single set of assumptions on the design of the interlanguage grammar. The result is that most of the analyses cited above are relevant for a specific level of the sound system and do not generalize easily to other levels of the phonology. For example, the proposed mechanics of feature mapping at the

5. Broselow et al. develop their analysis further to reflect a wider range of Mandarin-ESL phonological patterns, but this example was used simply to illustrate the 'emergence of the unmarked' phenomenon.

segmental level do not generalize clearly to syllable structure matching or other prosodic domains. There is increasing interest in adopting OT to understand L2 data (e.g., Hancin-Bhatt & Bhatt 1997; Broselow, Chen & Wang 1998; Hayes 1999; Hancin-Bhatt 2000; Lombardi 2003), which forces researchers to make explicit assumptions about the learner's initial state and how development of the grammar is constrained by Universal Grammar, coincidentally supporting a Full Transfer/Full Access (Schwartz & Sprouse 1996) model of L2 learning.

Full Access to UG is reflected in the constraints proposed, in particular, the markedness-based ones, and the candidate structural descriptions (potential outputs) available to the learner. Full Transfer is reflected by a full instantiation of the L1 constraint ranking into the initial state of the interlanguage grammar. Stages of interlanguage development are defined by constraint re-rankings from the L1-instantiated hierarchy to be more faithful to the L2 input. Thus, L1 transfer is predicted to be the basis for learner difficulty initially, but these difficulties can be overcome, and the 'repairs' to the unlearned structures become increasingly defined by universal tendencies toward less marked structures.

One of the first studies to use OT to understand transfer and developmental effects in L2 sound patterns was Hancin-Bhatt and Bhatt (1997). Their study focused on L2 syllables with complex onsets and codas, generated by Spanish and Japanese English L2 learners, and showed how error rates and types reflect developmental and transfer effects.

To illustrate one aspect of their analysis, consider the case of complex codas word finally. As can be seen in (10), neither Japanese, nor Spanish allow complex codas word finally, thus these coda types are predicted to be difficult for both sets of learners. Both languages allow only a limited range of simplex codas.

(10) Japanese and Spanish (word-final) coda inventories.

Consonants	Japanese	Spanish
C	/n/	/n/, /l/, /r/, /s/, /d/
CC	*	*

(11) Japanese and Spanish ESL Complex Coda Production

Sample Complex Coda Types in English

liquid + stop	lt, rt, lp, rp, lk, lp, ld, rd, lg, rg, etc.
liquid + fricative	ls, rs, lf, rf, lv, rv, etc.
liquid + nasal	lm, rm, ln, rn

ESL simplification strategies: Deletion

In a task designed to elicit a range of coda types, including the ones given above, intermediate-level Spanish and Japanese ESL learners simplified the difficult coda clusters through deletion, not epenthesis strategies. Thus, in these learners' interlanguage, *Complex ranks higher than the faithfulness constraints, and within

Faith, Dep-IO dominates Max-IO, indicating that deletion of a segment is more harmonious than epenthesis. But, as evidenced in the data and analyzed within the proposed constraint ranking, deletion of a single consonant is a lesser violation than deletion of both consonants, which would entail a double violation of Max.

(12) *Complex >> Faith (Dep-IO >> Max-IO)

CC	*Complex	Dep-IO	Max-IO
CC	*!		
CV.C		*!	
☞ <C>C			*
<C><C>			**!

ESL simplification strategies: Markedness effects

The Japanese and Spanish ESL data also indicate that some complex codas are more difficult than others, regardless of the L1. Complex codas with increasingly smaller sonority distances between the coda consonants had an increasing number of errors/simplifications.

Sonority Distance:	high ______________________________ low
Mean errors:	liquid+stops < liquid+fricatives < liquid+nasals

To account for this effect, Hancin-Bhatt and Bhatt (1997) appealed to a generalized constraint on margin sonority developed in Colina (1995), called M(argin) Son and expanding that to O(nset) Son and C(oda) Son. For purposes of this discussion, we present the definition of C Son.

(13) C Son is a constraint on the minimum distance in sonority that consonants in the same syllable position can have.

By adopting a constraint on sonority distance required for segments within a syllable position, the analysis captured segment-syllable licensing effects within the generalized mechanisms of OT. More recent work in OT, however, has developed an alternative to characterizing licensing effects by postulating a new set of constraints on alignment. Exploring how this development on the family of alignment constraints furthers our understanding of prosodic licensing effects in L2 phonology acquisition is a critical direction for future research.

ESL simplification strategies: Language transfer effects

In addition to finding that certain coda clusters presented more difficulty than others, Hancin-Bhatt and Bhatt found a clear transfer effect in how these learners were resolving this difficulty. Specifically, Spanish-speakers were more likely to delete the second consonant in the cluster (and thus maintain the liquid), while

Japanese-speakers were more likely to delete the first consonant in the cluster (and thus maintain the obstruent).

(14) Spanish ESL Simplifications CC → CØ > ØC
Japanese ESL Simplifications CC → ØC > CØ

The data indicate that complex cluster simplification is language-dependent, suggesting that consonant licensing effects from the L1 are being transferred to the interlanguage grammar. Hancin-Bhatt and Bhatt (1997) show how these licensing effects can be captured in a family of constraints on associational harmony (based on Smolensky & Prince 1993[2004]), which capture generalizations on the harmony of certain segments in specific syllable positions. The basic argument is that the language-dependent error types are a function of the learners' L1 constraint-ranking transferred into the interlanguage grammar. In this case, Spanish speakers have a different set of coda associational harmonies than Japanese speakers and the differences are a result of different constraint rankings.

A second OT-based L2 study on the interaction of L1 and markedness effects is Lombardi (2003), which focuses on segmental substitutions. Her study re-examines the long-standing problem of differential interdental substitutions, whereby the English /θ/ is substituted with either [s] or [t], depending on the first language of the L2 learner. What makes interdental substitutions interesting is that the substitutions are language-dependent, even though all the L1s contain both segments. Lombardi's main claim is that some L1 speakers produce substitutions that have an L1 genesis [e.g., θ → s], but other L1 speakers produce substitutions that are guided by a markedness principle describing languages' tendency to favor stops [e.g., θ → t]. Languages that maintain continuancy, F [continuant] have a dominant constraint Ident-cont that maintains manner faithfulness to the input.

Her analysis proceeds as follows:

(15) Lombardi's (2003) Analysis of Interdental Substitutions

A. Grammars contain a markedness constraint that indicates that fricatives are more marked than stops. *[**continuant**] >> *[**stop**]

B. Markedness constraints conspire against the occurrence of interdentals in segment inventories and this is represented simply as *[**θ**].

C. Languages vary in their ranking of manner faithfulness relative to the markedness constraints. The relevant faithfulness constraint for interdental substitution is **Ident-Manner**, which is defined by the following manner features: [stop], [continuant], [strident].

D. Languages with /θ/ → [t] substitutions merely reflect UG tendencies toward unmarked segments (stops), while languages with /θ/ → [s] substitutions reflect transfer of re-ranked faithfulness constraints over markedness ones.

(16) Interdental Substitutions

/θ/→t: In languages with no stop/fricative distinction (context-dependent), the markedness constraints remain dominant over the faithfulness constraints, as (possibly) given by UG (cf., McCarthy 2002).

/θ/		*θ	*cont	*stop	Ident-Manner
	θ	*!	*		
	s		*!		
☞	t			*	*

/θ/→s: In languages with a stop/fricative distinction, faithfulness constraints are re-ranked above the markedness ones, and this re-ranking is context-dependent. [Possible alternative: exploding Ident-Manner such that the [continuant] feature gains prominence in the grammar: a Ident-Cont >> Ident-Stop.

/θ/		*θ	Ident-Manner	*cont	*stop
	θ	*!		*	
☞	s			*	
	t		*!		*

Lombardi tested her analysis with Japanese and Thai data and developed the syllable-dependency feature of her proposal.

To summarize this section on the interaction of transfer and developmental effects, an OT framework provides explicit assumptions on what the learner begins with, how markedness effects are encoded and can eventually emerge in an interlanguage grammar, and how L1-based constraint rankings make specific predictions on how the learner will resolve unlearned L2 phonological structures. Again, although OT was not proposed to account for L2 acquisition issues, the interaction of transfer and developmental effects finds an elegant explanation within this model of the grammar. There is, of course, more work to be done to further test the predictions of OT, examining a range of L1–L2 learning situations, and this initial work provides solid motivation and directions within which to pursue studies in this area.

Role of prosodic constraints

OT has been particularly useful in showing how levels of phonology interact with each other, and this can be seen by browsing the contents of the increasing number of books on OT. The strength of the theory is that the set of constraints proposed for describing prosodic effects is limited, thereby allowing a relatively small number of basic constraints to handle prosodic effects at all levels of the sound system.

An example from syllable structure is that codas are limited cross-linguistically in the types of segments that may surface. These coda limitations, or conditions, trigger 'repair' rules within rule-based accounts, yet the rules to repair illicit codas may vary (e.g., deletion, epenthesis, substitution), and there is no theory-internal explanation on how these rules are selected. By incorporating Coda Conditions within the set of well-formedness constraints, and assuming specific rankings within Faithfulness constraints, OT captures the common cause, and effect, for the repair strategies, thereby capturing their functional similarity. Kager (1999) described this as follows:

(17) On the functional unity of rules in OT (Kager 1999: 139)
"Any theory that fails to recognize the output as a level at which phonological generalizations hold fails to capture the functional unity of these phonological processes. In contrast, OT captures this functional unity straightforwardly thereby creating unity in typological diversity."

One L2 phonology study which examines how OT handles the interaction of phonological levels within a sound system is Hancin-Bhatt (2000). This study looks at Thai-speakers learning English and focuses on the family of associational constraints on the harmony of segments in specific syllable positions. Only a subset of the Thai segment inventory can occur in syllable coda position, so these learners experience substantial difficulty in learning the range of codas that English has. This provides a clear example of how syllable position limits what segments can surface.

To account for the segment-syllable dependency, Hancin-Bhatt adopts OT's family of associational constraints, represented as Coda-ac, given in (18).[6] This representation shows that constraints cluster together to be more or less prominent in this language's hierarchy, whereby constraints in A dominate B, and those in B dominate C. Because they are higher ranked, the constraints in (A), are the most serious violations in this grammar, while the lowest ranked violations in (C) are the least serious. Thus, the segments that are more likely to occur in Thai coda margins (i.e., are more harmonious in codas) are the nasals, glides and voiceless stops. Less likely to occur in Thai coda margins (i.e., less harmonious) are fricatives and the liquid /l/, while voiced or aspirated stops and the liquid /r/ are the least likely to occur (i.e., least harmonious) in the coda. The target grammar for these learners is English, which does not have as many restrictions on what segments can appear syllable finally as Thai does.

6. For purposes of this paper, Coda-ac is equivalent to CodaCond.

(18) Associational Constraints on Thai Codas.

CODA-AC (Thai)

A. $*M_{cod}$/voiced stops, $*M_{cod}$/aspirated stops, $*M_{cod}$ /r, $*M_{cod}$/h
   >>
B. $*M_{cod}$ /f,s, $*M_{cod}$ /l
   >>
C. $*M_{cod}$ /nasals, $*M_{cod}$ /glides, $*M_{cod}$ /voiceless stops

CODA-AC (English)

A. $*M_{cod}$/h
   >>
B. $*M_{cod}$/voiced & voiceless stops, $*M_{cod}$ /f,s, $*M_{cod}$ /liquids, $*M_{cod}$ /nasals, $*M_{cod}$ /glides

The learner's task, then, is to learn the new rankings that define optimal codas in English.

Hancin-Bhatt (2000) presents data gathered from Thai-speakers' production of monosyllabic English pseudowords. The data considered here are from subjects' productions of the pseudowords with non-Thai-like codas and the results are summarized in (19). Nasal codas had the fewest errors at 95% accuracy, while voiced stops had the most at 66% accuracy. An analysis of errors shows that substitution was the most frequent repair strategy for dissonant codas.

(19) Summary of Results on Unlicensed Codas in Thai ESL. (Hancin-Bhatt 2000)

	Coda Accuracy	Substitution	
Nasals	95%	2%	4% epenthesis
Fricatives (f, s, v, z)	89%	12%	
Voiceless stops	88%	12%	
Liquids	83%	9%	8% deletion
Voiced stops	66%	33%	

These results indicate that, predictably, Thai speakers did not have difficulty with nasals in coda position. Liquids are predictably difficult because a liquid in coda position in Thai is very rare (i.e., no /r/ and rarely /l/). The surprising result here is that learners do not have much difficulty with fricatives, which are predicted to be more difficult than voiceless stops, and this can be interpreted as a constraint demotion of the margin condition on codas that avoids fricatives in that position. This constraint re-ranking is represented in outline form in (20). Based on the data and on what is known of the L1 and target grammars, Hancin-Bhatt proposes how the Thai-ESL interlanguage grammar restructures.

(20) Thai-ESL Interlanguage Grammar: Coda Constraint Rankings and States

Initial State: $^*M_{cod}$/voiced stops, r, h >> **$^*M_{cod}$/f, s** >> $^*M_{cod}$/voiceless stops, nasals

Intermediate State: $^{**}M_{cod}$/voiced stops, r, h >> $^*M_{cod}$/voiceless stops, nasals, $^*\mathbb{M}_{\mathbf{cod}}$/**f**, **s**

...

Target State: $^*M_{cod}$/h >> $^*\mathbb{M}_{\mathbf{cod}}$/**voiced stops**, **r**, **h**, $^*M_{cod}$/voiceless stops, nasals, $^*M_{cod}$/f, s

A second interesting finding from this data is that simple codas are repaired mostly through substitution, not the epenthesis of a vowel or the deletion of the illicit consonant. Within OT, this is explained by appealing to rankings within the family of the Faithfulness Constraints (Faith). That substitution is more likely to occur than epenthesis or deletion reflects Ident-IO's low-ranking relative to Max-IO and Dep-IO. That is, it is more harmonious to violate Ident-IO than it is to insert or delete a segment in an optimal output.

(21) Thai ESL Optimal Outputs on Illicit Codas: Faithfulness Constraint Rankings
Dep-IO, Max-IO >> Ident-IO
Epenthesis and deletion repairs are more serious than segment substitution repairs.

Hancin-Bhatt (2000) develops the analysis of errors and their interpretation within an OT framework further, but the point of this section is to discuss how OT provides a framework within which to discuss the interaction between segment qualities and syllable positions. Associational and correspondence constraints define the universal range of disallowed pairings, but how strongly those constraints hold are language-specific, reflected in how high (or low) they are ranked in a grammar.

The finding highlighted here is that when encountered with segments not licensed for coda position in the coda, these learners were more likely to substitute the segment, rather than to delete or epenthesize a vowel to modify the syllable structure. Other language learners or language learners at other stages of development adopt other strategies to 'repair' unlearned structures. OT accounts for this variation across repairs simply through re-ranking of the universal constraints.

Interlanguage grammar restructuring

In addition to accounting for learner error in L2 phonology, Optimality Theory has also opened up distinct possibilities for a new strain of L2 acquisition studies that explore not only how predicted constraint re-rankings reflect states in the dynamic interlanguage grammar, but also what features of the input-output evaluation force a restructuring in those states.

IL grammar representations

The beginnings of constraint re-rankings as a reflection of interlanguage grammar states have been posited, but are underdeveloped in the L2 phonology literature. Hancin-Bhatt (1998) suggested a specific development in complex coda acquisition based on constraint re-rankings, given in (22). During initial stages of language acquisition, the constraint hierarchy is instantiated onto the L2, with a simultaneous demotion of the Faithfulness constraints. During early stages of complex coda acquisition (for learners whose L1 does not have complex codas), learners are first working on re-ranking constraints on segment correspondence within the Coda Conditions. That is, they are first learning new phonotactic or prosodic licensing patterns of the target language. Subsequently, learners override the constraint on complex segments within a syllable position, thereby allowing clusters. Eventually, in the target grammar, the re-rankings force Faithfulness constraints to dominate, to reflect a target-like constraint ranking.

(22) Stages in Coda Development
State$_{\text{initial}}$: *Complex$_{\text{COD}}$, CodaCond >> Faith
State$_{\text{n}}$: *Complex$_{\text{COD}}$ >> Faith >> **CodaCond**
State$_{\text{target}}$: Faith >> ***Complex$_{\text{cod}}$** >> CodaCond

An interesting variant on perspectives of L2 learning is Hayes (1999), a study that uses OT's Constraint Demotion Algorithm to predict stages in English-speakers' acquisition of Japanese syllable structure. Because Japanese syllable structure is a subset of allowable English syllables, the learner's task is not to learn new structures, but rather to *delimit* the set of possible Japanese syllables. Specifically, Japanese does not allow complex onsets or codas, and, as already discussed earlier, possible codas are restricted to a nasal.

Hayes' mini-longitudinal study asked English-speaking learners of Japanese to do a (phonological) grammaticality judgment task where they listened to pseudowords and had to indicate whether or not the word was a possible Japanese word. They then repeated the pseudoword to make it sound Japanese. The pseudowords included English-like syllable structure, in particular complex onsets and codas and a range of non-nasal simplex codas. Some of the findings are summarized in (23).

(23) English-speakers' Production Errors: Japanese Pseudowords. (Hayes 1999)
RATES: Errors decreased over three sessions: SI-17.6%; S2-11.6%; S3-10.8%
TYPES: The different types of errors had different patterns over the three sessions
*Complex$_{\text{ONS}}$: S1-32%; S2-18%; S3-19%
CodaCond: S1-25%; S2-17%; S3-15%
*Complex$_{\text{COD}}$: S1-4%; S2-5%; S3-5%

As is expected, overall error rates decreased over the three sessions during which data were collected. However, the types of errors learners made had varying patterns. Errors in producing the non-Japanese complex onsets were the greatest and decreased substantially between sessions over the three sessions, while errors in complex codas were minimal and steady, representing only 4–5% of the productions over all three sessions. Errors in Japanese coda conditions followed a similar pattern as that of complex onsets. Hayes suggests that the varying error type patterns reflect different constraint re-rankings within the interlanguage grammar, and those are given in (24). Interestingly because her study looks at how learners *delimit* the range of possible L2 structures, her re-rankings of FAITH indicate an initial high ranking, but a gradual demotion, a pattern opposite to what Hancin-Bhatt (1998) proposed for learners trying to acquire a greater range of linguistic structures. (See also 22 above.)

(24) Delimiting Linguistic Structures: English-speakers' Learning Japanese Syllable Structure

$\text{State}_{\text{initial}}$:	FAITH >> $\text{*COMPLEX}_{\text{COD}}$, CODACOND, $\text{*COMPLEX}_{\text{ONS}}$
State_{n}:	$\text{*COMPLEX}_{\text{COD}}$ >> **FAITH** >> CODACOND, $\text{*COMPLEX}_{\text{ONS}}$
$\text{State}_{\text{target}}$:	$\text{*COMPLEX}_{\text{COD}}$, CODACOND, $\text{*COMPLEX}_{\text{ONS}}$ >> **FAITH**

Forces that restructure grammar: Learning algorithms

As the examples above show, OT provides a framework within which to discuss potential states of the interlanguage grammar and what order they may come in. This work is still very preliminary, but these examples demonstrate the potential for making specific predictions about what learners are working on/acquiring at various stages of their development. The next logical question, of course, is how learners move through the developmental stages. What specifically forces a shift from one interlanguage state to the next.

To answer this question within an OT framework, researchers appeal to learning algorithms which map out how a system learns from the input it receives and how well it conforms to the current grammatical state. There have been numerous proposals on learning algorithms, the first one being Tesar and Smolensky's (2000) RIP/CDA, and another promising algorithm that adopts a functional phonology approach is that of Boersma (2000). At this time, there are no published L2 phonology studies that focus on how a learning algorithm determines exactly why constraints re-rank given a specified input. However, one feature of the constraint demotion algorithm that is particularly interesting for L2 phonology is the assumptions on what the given input into the grammar is. This idea is revisited in the section below on the Perception-Production asymmetry.

To summarize this section on acquisition, one initial observation that can be made is the seemingly unstable nature of the prominence of the family of

FAITHFULNESS constraints. While this family of constraints tends to be high ranking in an L1, thereby mitigating against superfluous deletion, epenthesis and substitution patterns in a language, the set is clearly demoted in early stages of L2 acquisition, as is evidenced by learners' multiple accommodations (or repairs) of unfamiliar L2 linguistic structures. That markedness constraints dominate faithfulness constraints in initial stages of acquisition has also been assumed by some in first language acquisition (cf., McCarthy 2002: 206). This effect coupled with the Full Transfer/Full Access hypothesis on the interlanguage grammar discussed earlier yields a working hypothesis on the generalized set of interlanguage grammar development.

(25) Interlanguage Grammar Development: States of Constraint (Re-)rankings

$\text{State}_{\text{initial}}$:	L1-ranking instantiated, with automatic demotion of FAITH (Markedness >> Faithfulness)
State_{n}:	Re-rankings within MARKEDNESS
$\text{State}_{\text{n+1}}$:	Re-rankings of FAITH above various MARKEDNESS constraints
$\text{State}_{\text{target}}$:	FAITH generally dominating MARKEDNESS

It is also important to point out that, while in acquisition mode, these states are dynamic and, as such, the constraint rankings are not fixed or steady. Rather, constraints may rank and re-rank under certain conditions or even be unranked, and this instability is evidenced in learners' variable outputs. The value of assuming OT is that the variable outputs, erred and error-free, are predictable, given the constraints and assumptions on their re-rankings.

Variable competence

As discussed above, many studies in L2 phonology acknowledge that their learners have target-like performance, as well as flawed performance within the same task. The range of errors on a particular structure can also vary, and some examples of variant productions are given in (26). It should be understood that these productions do not all occur at the same rate within a learner, some are more likely than others to occur, particularly at different stages of acquisition. However, the point here is that the variant productions are realities in an interlanguage grammar and, thus, must be accounted for.

(26) Sample Variant Productions by ESL Learners.

Thai-ESL		Spanish-ESL
/malk/	malk ~ mak ~ mals	malk ~ mal
/krelp/	krelp ~ krep ~ krelf ~ kref	krelp ~ krel

Many second language studies have focused on understanding the errors and trying to determine the acquisitional stage of the learner, given their error rates and types, but have not discussed that learners simultaneously produce accurate target structures, even within the same task. In other words, the learners have achieved some knowledge of the target grammar, and the obvious question then, is how is this target-like competence accommodated within proposed models of the interlanguage that consider mostly errors as the basis for model design.

OT is a theory that assumes linguistic variation is a natural consequence of a system that responds to its input, and the growing number of OT studies on language change, dialectal variation, and loanword phonology attest to the framework's potential. Demuth (1997) highlights evidence of variation within L1 acquisition and how OT is a framework within which to understand a developing grammar's 'multiple optimal outputs'. To illustrate, consider the OT evaluation of the English pseudoword /alk/ by a Spanish-speaking learner, given in (27). The CodaCond adopted here is from Colina (1995), that Spanish allows only coronal consonants in codas. The indeterminacy in the ranking reflected here is between *Complex and Max-IO, and the ambiguous ranking is represented by the dotted line between the two constraints. *Complex is outlined to suggest its re-ranking relative to Max-IO in the target grammar.

(27) Tied Constraints and Variable Outputs

/alk/	*Complex	Max-IO	CodaCond
alk	*!		*
a<l>k		*	*!
☞ al<k>		*	

/alk/	Max-IO	*Complex	CodaCond
☞ alk		*	*
a<l>k	*!		
al<k>	*!		

OT-based (L1) acquisition studies are converging onto a particular generalization on developing grammars: Faithfulness constraints appear to be ranked below the Markedness constraints in early stages of the developing grammar, and the grammar must allow for indeterminacy of specific rankings relative to each other. But the outstanding question here is whether this indeterminacy is due to equal ranking of competing constraints or to variant inputs into the grammar. Studies on development within perception are beginning to shed light on potential variant inputs, but clearly there is a need for more research in this area.

Perception-production asymmetry

That learners have asymmetrical competencies in their perception and production has increasingly received accounts within analyses of the L2 sound patterns.

In fact, because of the importance of /input/ in defining and restructuring the interlanguage grammar, recent years have seen a substantial increase in the number of studies looking at perceptual abilities at various levels of the phonology, a shift from phonological studies focused only on production evidence.

To date, there are few to no published studies that present an OT analysis of L2 perception. A study by Jacobs and Gussenhoven (2000) looks at loan phonology and interprets the accommodations of L2 input within an L1 grammar as a function of transfer in the perceptual domain, not production. Their claim is that the adult's universal parser assigns phonological representations to L2 structures, just as a child's would, with the difference being that the two are using different constraint hierarchies to evaluate the potential outputs, thus different forms are selected as optimal. That is, adults may 'hear' the L2 structure differently than the child due to the different constraint rankings involved in the parse.

There is promising work being done within an OT framework developing the idea that there is a separate perceptual grammar (cf., Boersma 1999), providing yet another avenue within which L2 researchers can re-address and hopefully expand on the yet outstanding explanation of the perception-production asymmetry and its impact on the grammar development.

Summary and future directions

This paper has presented a few exemplars of L2 phenomena analyzed within an Optimality Theoretic framework, but much more research is required to fully understand the theory's potential for informing studies in L2 acquisition. Given this ideas covered briefly here, there are certain implications and possible directions for future studies.

Theoretical implications

The theoretical implications of OT are numerous due to the richness of its development. I will generalize the implications by referring back to the issues discussed at the beginning of this paper on what a model of the sound patterns should address.

- Grammar Representation: Within OT, the interlanguage grammar consists of a universal set of constraints on structural well-formedness and input-output faithfulness and an L1-instantiated constraint ranking in which constraints progressively re-rank to converge onto a target-like ranking. *FT/FA Hypothesis.*
- Acquisition: Under OT, the acquisitional stages are manifest by the learning algorithm and its linguistic consequence – the re-ranking of constraints understood as restructured interlanguage grammar.
- Variable Competence: Under OT, constraint rankings are dynamic and can be unranked relative to each other, particularly during acquisition, prior to the grammar having received a steady state.

As we expand these theoretical assumptions into explicit predictions for specific L1-L2 situations, the field of L2 phonology enters a new era whereby theory-driven research can inform and disambiguate the major findings from the predominantly data-driven past. This is a critical step for the field in that it should produce research that is faithful to theoretical claims on cognitive design, as well as allow us to predict areas of obvious, and not-so-obvious difficulty for learners. Advances at both these levels are over due in the field. Furthermore, because OT is a theory of variation within Universal Grammar, evidence from the developing bilingual's grammar can finally be considered as central to theory construction, which could have the consequence of attracting more researchers to this field of study. Finally, by adopting a rich theory of the grammar that specifies how acquisition takes place, researchers will be forced to carefully control how evidence is gathered to test their predictions.

Methodological implications

Although various methodological controls in the field of L2 phonology have been amply discussed by others (see, e.g., Crookes 1991; Tarone, 1983) and researchers are increasingly attentive to implications of various methodologies, the field is still developing more sophisticated methods for collecting data and controlling variance. By adopting OT and greater precision in predictions for L2 phonology acquisition, researchers must continue to refine methods for data collection. Much can be said regarding this issue, but, due to space constraints, these methodological implications are only briefly presented here.

– Separate perception tasks/evidence from production tasks/evidence: With the increasing evidence that perception and production data cannot be directly compared and the developing proposals that the perception and production processes are mediated differently, it is critical that studies explicitly commit to either perception or production evidence and that generalizations obtained from one domain not be used to test predictions at another level, unless, of course, the predictions specifically refer to similarities/differences in the perception and production processes and not in the underlying grammar.
– Control for prosodic domains that may impact learner performance within phonological levels. Specific domains that are particularly important to attend to are syllable position of segments, word position, and even the utterance level when looking at stress/intonation.
– Observe learners over time (or, less desirably, cross-sectionally). When looking at acquisition of structures and changes in interlanguage grammar, it is particularly critical that learners are observed over time. There is little theory-driven L2 phonology research that provides longitudinal data, but without it, claims about acquisition within an OT model are weak.
– Measure and account for all variants (erred and error-free) within task. Much of the previous research on L2 phonology accounted for only errors made by learners. With OT, the accurate productions are also considered variants in the grammar and, thus, need to be accounted for.
– Control for variant task demands that can elicit a greater (or lesser) number of errors. Again, the concern here is understanding the range of variants that occur and the frequency with which they occur. The challenge is to determine the nature of the variant (transfer, markedness or other?) and how a greater or lesser frequency of the variants reflects specific states in the interlanguage grammar, particularly when that grammar is considered to be dynamic.
– Control the input. In OT, the grammar evaluates the harmony of the input relative to the constraint rankings, and, thus, the form of the input is primary. What are the cues given in the input (written/aural, clear/degraded) and how do the various types of input cues get evaluated in the grammar?

Implications for future directions

This review of the current work in L2 and OT offers specific directions for future research. In general

1. What is the nature of the initial state, especially with respect to the Faith and Markedness constraints and is this dominance universal?
2. What types and how much of input trigger a constraint demotion?

3. More studies are needed on the interaction between phonological levels and how those find an account within OT.
4. Many more L1-L2 learning situations need to be explored within perception, production domains.
5. Should we assume a separate perception and production grammar, and, if so, how do the two interact? Or more broadly, how do we connect the perceptual and production processes with grammatical representations?
6. Do learners have multiple inputs associated with an L2 form? If so, how do we represent that?
7. Related to 5, how do we encode psycholinguistic phenomena that impact how the grammar determines optimal outputs? E.g., training, attention effects, enhanced input, mode (aural v. written) of input on development of grammar all likely have an impact on optimal outputs.

To conclude, the value of OT to the field of L2 acquisition lies in its account of an interlanguage grammar, its restructuring and the variability observed therein. As a generalized model of the grammar, OT assumes that interlanguage grammars are natural, dynamic systems in process of accommodating new inputs and that L1 influence and markedness effects are merely a consequence of the system's design. No special rules or accommodations need be posited for the L2 learning situation. As a generalized model of the grammar that encodes a system of input–output correspondence and feedback on the harmony of that correspondence, OT not only allows for a range of inputs that can be parsed and impact learning, but it also allows, under certain circumstances, for more than one output to be parsed as optimal (or least dissonant), thereby defining the nature of variability. As a generalized model of the grammar, OT provides a set of constructs within which to explore hypotheses on various levels of grammar, including interfaces between levels such as the role of prosodic constraints and the perception – production relationship. With this rich design, L2 researchers can explore developing bilingual grammars and how they are influenced by the input, leading us to new questions in the field of L2 sound patterns and strengthening connections across related disciplines that also explore the developing bilingual brain.

Suggested readings

Archangeli, D. 1999. Introducing Optimality Theory. *Annual Review of Anthropology* 28: 531–552.

Dekkers, J., van der Leeuw, F., & van de Weijer, J. (eds.). 2000. *Optimality theory: Phonology, syntax and acquisition*. New York: Oxford University Press.

Kager, R. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Rutgers Optimality Archive. *http://roa.rutgers.edu/*

Prince, A., & Smolensky, P. 2004. *Optimality Theory: Constraint interaction in Generative Grammar*. Blackwell Publishers.

Tesar, B., & Smolensky, P. 2000. *Learnability in optimality theory*. Cambridge, MA: MIT Press.

References

Altenburg, E., & Vago, R. 1987. In G. Ioup & S. Weinberger (eds.) *Interlanguage phonology: The acquisition of a second language sound system,* Cambridge, MA: Newbury House, 148–164.

Archangeli, D. 1999. Introducing Optimality Theory. *Annual Review of Anthropology* 28: 531–552.

Archibald, J. 1993. *Language learnability and L2 phonology: The acquisition of metrical parameters*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Best, C., & Strange, W. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20(3): 305–330.

Boersma, P. 1997. How we learn variation, optionality, and probability. [ROA ms. number 221–1097. *http://roa.rutgers.edu/*]

Boersma, P. 1999. On the need for a separate perception grammar. [ROA ms. number 358–1099. *http://roa.rutgers.edu/*]

Boersma, P. 2000. Learning a grammar in Functional Phonology. In J. Dekkers, F. van der Leeuw, & J. van de Weijer, (eds.), *Optimality Theory: Phonology, syntax and acquisition*, New York: Oxford University Press, 465–523.

Boersma, P., Escudero, P., & Hayes, R. 2003. Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. In M. J. D. Sole, D. Recansen, & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona.

Boersma, P., Dekkers, J., & van de Weijer, J. 2000. Introduction. In J. Dekkers, F. van der Leeuw, & J. van de Weijer, (eds.), *Optimality Theory: Phonology, syntax and acquisition*, New York: Oxford University Press, 1–44.

Boersma, P., & Hayes, B. 1999. Empirical tests of the Gradual Learning Algorithm. ms. University of Amsterdam and UCLA. [Rutgers Optimality Archive 348, *http://roa.rutgers.edu/*]

Bradlow, A. R. 2008. Training non-native language sound patterns: Lessons from training Japanese adults on the English /ɹ/–/l/ contrast. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Broselow, E. 1987. An investigation of transfer in second language phonology. In G. Ioup & S. Weinberger (eds.), *Interlanguage phonology: The acquisition of second language sound system*, Cambridge, MA: Newbury House, 261–278.

Broselow, E., Chen, S., & Wang, C. 1998. The emergence of the unmarked. *Studies in Second Language Acquisition* 20(2): 261–280.

Broselow, E., & Finer, D. 1991. Parameter setting in second language phonology and syntax. *Second Language Research* 7: 35–39.

Broselow, E., Hurtig, R., & Ringen, C. 1987. The perception of second language prosody. In G. Ioup & S. Weinberger (eds.), *Interlanguage phonology: The acquisition of second language sound system*, Cambridge, MA: Newbury House, 350–361.

Brown, C. 2000. Speech perception and phonological acquisition. In J. Archibald & M. Young-Scholten (eds.), *Second language and linguistic theory*, Malden, MA/Oxford: Blackwell Publishers, 4–63.

Colina, S. 1995. A constraint-based analysis of syllabification in Spanish, Catalan, and Galician. Unpublished Ph.D. dissertation. University of Illinois, Urbana-Champaign.

Crookes, G. 1991. Second language speech production research: A methodologically-oriented review. *Studies in Second Language Acquisition* 13: 113–132.

Demuth, K. 1997. Variation in acquisition: An optimal approach. In S. Davis (ed.), *Optimal viewpoints,* Bloomington, IN: Indiana University Linguistics Club, 77–88.

Dickerson, W. 1975. The learner's interlanguage as a system of variable rules. *TESOL Quarterly* 9: 401–407.

Eckman, F. 1977. Markedness and the contrastive analysis hypothesis. *Language Learning* 27: 315–330.

Eckman, F. R. 2008. Typological markedness and second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Eckman, F., & Iverson, G. 1993. Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research* 9: 234–252.

Flege, J. 1987. Effects of equivalence classification on the production of foreign language speech sounds. In A. James & J. Leather (eds.), *Sound patterns in second language acquisition*, Dordrecht: Foris, 9–39.

Flege, J. 1990. Perception and production: The relevance of phonetic input to L2 phonological learning. In C. Ferguson & T. Huebner (eds.), *Crosscurrents in second language acquisition and linguistic theories*, Philadelphia: John Benjamins.

Flege, J. 1993. Speech learning in a second language. In C.A. Ferguson, L. Menn, & C. Stoel-Gammon (eds.), *Phonological development: Models, research, implications*, Timonium, MD: York Press, 565–603.

Gibson, E., & Wexler, K. 1994. Triggers. *Linguistic Inquiry* 25: 407–454.

Grosjean, F. 1998. Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1: 131–149.

Hancin-Bhatt, B. 1994. Segment transfer: A consequence of a dynamic system. *Second Language Research* 10: 241–269.

Hancin-Bhatt, B. 1998. Optimality and stages of L2 syllables. Paper presented at the *1998 Second Language Research Forum*, Honolulu, HI.

Hancin-Bhatt, B. 2000. Optimality in second language phonology: Codas in Thai ESL. *Second Language Research* 16(3): 201–232.

Hancin-Bhatt, B., & Bhatt, R. 1997. Optimal L2 syllables: Interaction of transfer and developmental effects. *Studies in Second Language Acquisition* 19: 331–378.

Hancin-Bhatt, B., & Govindjee, A. 1999. A computational model of feature competition in L2 phonology. In P. Broeder & J. Murre (eds.), *Language and thought in development: Cross-linguistic studies*, Tübingen: Gunter Narr, 145–161.

Hardison, D. 1996. Bimodal speech perception by native and nonnative speakers of English: Factors influencing the McGurk Effect. *Language Learning* 46(1): 3–73.

Hayes, R. 1999. Reranking stages in OT analysis of the acquisition of Japanese as a second language. *Carolina Working Papers in Linguistics* 1.

Jacobs, H., & Gussenhoven, C. 2000. Loan phonology: Perception, salience, the lexicon and Optimality Theory. In J. Dekkers, F. van der Leeuw & J. van de Weijer (eds.), *Optimality Theory: Phonology, syntax and acquisition*, New York: Oxford University Press, 193–210.

Kager, R. 1999. *Optimality Theory.* Cambridge: Cambridge University Press.

Lado, R. 1957. *Linguistics across cultures.* Ann Arbor, MI: The University of Michigan Press.

Lombardi, L. 2003. Second language data and constraints on manner: Explaining substitutions for the English interdentals. *Second Language Research* 19(3): 225–250.

Major, R. 1986. The ontogeny model: Evidence from L2 acquisition of Spanish *r*. *Language Learning* 3: 453–504.

Major, R. 1994. Chronological and stylistic aspects of second language acquisition of consonant clusters. *Language Learning* 44: 655–680.

Major, R. C. 2008. Transfer in second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

McCarthy, J. 2002. *A thematic guide to Optimality Theory.* Cambridge: Cambridge University Press.

McCarthy, J., & Prince, A. 1995. Faithfulness and reduplicative identity. In J. Beckman, L. Dickey, & S. Urbanczyk (eds.), *Papers in Optimality Theory: University of Massachusetts Occasional Papers in Linguistics* 18: 249–384.

Prince, A. & Smolensky, P. 2004. *Optimality Theory: Constraint interaction in Generative Grammar.* Blackwell Publishers.

Sato, C. 1987. Phonological processes in second language acquisition: Another look at interlanguage syllable structure. In G. Ioup & S. Weinberger (eds.), *Interlanguage Phonology: The acquisition of second language sound system*, Cambridge, MA: Newbury House, 248–260.

Schwartz, B., & Sprouse, R. 1996. L2 cognitive states and the full transfer/full access model. *Second Language Research* 12: 34–66.

Strange, W., & Shafer, V. L. 2008. Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Tarone, E. 1978. The phonology of interlanguage. In J. Richards (ed.), *Understanding second and foreign language learning*, Rowley, MA: Newbury House, .

Tarone, E. 1982. Systematicity and attention in interlanguage. *Language Learning* 32: 69–84.

Tarone, E. 1983. On the variability of interlanguage systems. *Applied Linguistics* 4: 142–163.

Tesar, B., & Smolensky, P. 1996. The learnability of Optimality Theory: An algorithmn and some basic complexity results. Ms, Department of Computer Science and Institute of Cognitive Science, University of Colorado at Boulder. ROA-2.

Tesar, B., & Smolensky, P. 2000. *Learnability in Optimality Theory.* Cambridge, MA: MIT Press.

Weinreich, U. 1953. *Languages in contact.* The Hague: Mouton.

White, L. 2000. Second language acquisition from initial to final state. In J. Archibald (ed.), *Second language acquisition and linguistic theory*, Oxford: Blackwell Publishers, 130–155.

Young-Scholten, M. 1995. The negative effects of 'positive' evidence on L2 phonology. In L. Eubank & L. Selinker (eds.), *The current state of interlanguage*, Amsterdam: John Benjamins, 107–122.

Young-Scholten, M. 1997. Second language syllable simplification: Deviant development or deviant input? In A. James & J. Leather (eds.), *New Sounds '97: Proceedings of the Third International Symposium on the Acquisition of Second-Language Speech,* Klagenfurt, Austria: University of Klagenfurt, 351–360.

Young-Scholten, M., & Archibald, J. 2000. Second language syllable structure. In J. Archibald (ed.), *Second language acquisition and linguistic theory*, Oxford: Blackwell Publishers, 64–101.

Zampini, M. 1994. The role of native language transfer and task formality in the acquisition of Spanish spirantization. *Hispania* 77: 470–481.

PART II

Second language speech perception and production

Preface

Part I introduced and examined in detail some of the major factors and theoretical isssues related to the acquisition of an L2 sound system. Part II, by contrast, focuses more specifically on the ways in which L2 speech sounds are perceived and produced. The first two chapters of this section address very different issues of L2 speech perception: Chapter 6 examines the perception of L2 speech sounds by L2 learners, whereas Chapter 7 focuses on how L2 learners' speech is perceived by others (especially native listeners of the L2 in question) and how accent affects intelligibility in the L2. The last two chapters of this section likewise examine distinct aspects of L2 speech production. Chapter 8 focuses on research that examines how learners produce L2 speech sounds, i.e., the acoustic and phonetic characteristics of L2 speech. Chapter 9, on the other hand, focuses on how social constraints, such as identity, gender, and interlocutor, affect L2 production, as well as on how and why learners' production of L2 sounds varies across formats, such as task and context.

In Chapter 6 ("Speech perception in second language learners"), Winifred Strange and Valerie L. Shafer explore the state-of-the-art in L2 speech perception research. The authors first outline how selective perceptual processes develop and become automatized in infant/child L1 acquisition and introduce a number of studies that illustrate the range of difficulty with respect to adult listeners' abilities to perceive non-native sound contrasts. Next, the authors provide a detailed discussion of methodological frameworks, including types of experimental tasks, for perception research. Within this context, the experimental paradigms are illustrated and critically evaluated. Strange and Shafer go on to provide an overview of current models of L2 speech perception, including the Native Language Magnet Model, the Perceptual Assimilation Model and Speech Learning Model, and discuss major theoretical issues, such as cross-language phonetic similarity. In addition, they address recent neurobiological approaches to the study of L2 speech perception, including electrophysiological measures of discrimination. They conclude their chapter by presenting a 'tetrahedral model' that provides a framework for considering several experimental variables for the design, interpretation, and evaluation of L2 speech perception studies.

In Chapter 7 ("Foreign accent and intelligibility"), Murray J. Munro addresses issues in how L2 learners' speech is perceived by others, especially in terms of global accent and how this affects communication. Munro first presents an overview of what 'foreign accent' entails, as well as its consequences (both positive and negative), and provides an overview of the segmental and suprasegmental aspects of speech that have been shown to contribute to the detection of a foreign accent by native listeners. The crux of his chapter then centers on the relationship between accent and intelligibility and how different conceptualizations of this relationship affect language teaching, language testing, and human rights litigation. Munro then outlines methodological and theoretical frameworks employed to evaluate L2 speech; in doing so, he also characterizes a number of issues related to the assessment of accent and intelligibility and identifies factors that may affect the ratings of both phenomena. The author concludes the chapter by addressing the implications of research on L2 accent and intelligibility for L2 pedagogy and by describing directions for future research.

In Chapter 8 ("L2 speech production research: Findings, issues, and advances"), Mary L. Zampini addresses research on the nature of the L2 speech sounds produced by L2 learners, as well as bilingual speakers. Providing a comprehensive overview of recent research in both segmental and suprasegmental aspects of L2 speech production, Zampini outlines the theoretical frameworks and major findings in each research domain. Her discussion of segmental research focuses primarily on subsegmental properties of L2 speech, such as voice onset time (VOT) for stop consonants and formant frequencies or duration for vowels; she also discusses the phoneme-level substitutions that occur in L2 speech. The discussion of suprasegmental aspects of L2 speech production focuses on syllable structure, prosodic domains, and stress. Zampini then outlines methodological approaches to the study of L2 speech production, and concludes by addressing the implications of L2 speech production studies for more general theories of language and acquisition and outlining a number of future directions for L2 speech research.

Finally, in Chapter 9 ("Social factors and variation in L2 production"), Jette G. Hansen Edwards provides a synthesis of research that addresses, first, how social factors such as social identity and gender affect L2 phonological acquisition and production, and second, variation in L2 speech production. Central to this discussion is the recognition that not all non-nativelike productions of L2 speech are the result of limitations caused by incomplete knowledge and control of the L2; rather, L2 learners sometimes make a decision to speak the L2 in a particular way for a variety of reasons. In addition, variation in L2 speech may be influenced not only by internal linguistic factors, but by social and other extralinguistic factors as well. In the discussion of social factors, Hansen Edwards summarizes the research on gender, extent of L1/L2 use, social identity, and target language variety. In the second section of the chapter, Hansen Edwards outlines three major areas of research on variation in L2 phonological production: research that focuses on interlocutor/speech accommodation, research on attention/monitoring, and finally, research that focuses on the variable rule analysis of linguistic and social constraints on variation. For each area of research, both theoretical and methodological frameworks are described, and a review of the major studies in the area is presented. Hansen Edwards concludes the chapter with a synthesis of major findings, followed by suggestions for future research.

CHAPTER 6

Speech perception in second language learners

The re-education of selective perception

Winifred Strange and Valerie L. Shafer
City University of New York – Graduate School and University Center

Introduction

One common characteristic of learners of a second/foreign language who acquired the language in late adolescence or adulthood is that their productions of the phonetic segments and sequences of the language are accented. That is, for almost all late second language (L2) learners, the phonetic realization of phonological structures in the L2 is markedly different from native-language patterns. The fact that native listeners can readily identify a) that a speaker is a late learner of their language and b) the native language of the speakers, (i.e., Spanish-accented English speakers; American-accented Japanese speakers) justifies the characterization of the accented phonological patterns as being due, to a considerable extent, to interference from the native language phonology. That is, when producing utterances in an L2, speakers often produce phonetic segments and sequences that appear to be a product of complex interactions between L1 and L2 phonetic realization rules (inter-language phonology).

What is less immediately apparent is that these same late L2 learners also have considerable difficulty with the receptive aspects of phonological processing of the L2. Phonetic segments which are phonologically distinctive in the L2, but not in the learners' native language are often not correctly recognized and categorized, leading to difficulties in comprehension of spoken L2 utterances. Indeed, it is commonly thought that a major determinant of L2 accentedness in production is the underlying problem associated with the perception of L2 phonological structures (Flege 1995). This chapter is intended to provide a brief review of the phenomena that characterize the phonetic perception difficulties of late L2 learners and to describe some current theories of L2 speech perception, as it relates to L2 speech production. In discussing the research on this topic, we will review some theoretical issues concerning the characterization of the underlying representations of L1 and L2 phonological categories and the cognitive processes involved in phonetic perception. We will also provide a framework for considering the many methodological issues in designing and interpreting behavioral and brain research on L2 speech perception. However, before reviewing theoretical and empirical developments in L2 speech perception research, a brief description of the nature of speech perception processes and their development in first language (L1) acquisition will set the stage for the subsequent discussion.

L1 speech perception: Development and automatization of selective perceptual processes

All phonetic features that can serve to distinguish phonological segments (i.e., that may underlie phonological contrasts in a particular language) can be differentiated acoustically by multiple parameters that systematically vary in value along several spectral and temporal dimensions. For instance, voicing contrasts between oral stop consonants in English (e.g. "pet" *vs* "bet" [pʰɛt-b̥ɛt]; "bet" *vs* "bed" [b̥ɛt̚-b̥ɛːd̚]; "bicker" vs "bigger" [b̥ɪkɚ-b̥ɪːgɚ] are differentiated by several temporal parameters (i.e. Voice Onset Time (VOT) for initial stops, duration of consonant closure for medial stops, preceding vowel duration for medial and final stops) and also by spectral characteristics (onset/offset frequency of the first formant transition, fundamental frequency contour after release for initial and medial stops, presence and extent of voicing energy during stop closure for medial stops). There is a complex mapping of values of acoustic parameters to phonetic feature values that is a) highly context-dependent, and b) language-specific.[1] Thus, as an acoustic signal, speech can be considered a *code* in which the phonetic segments and sequences are specified by context-dependent and language-specific complexes of acoustic parameters. Speech perception, then, involves decoding the acoustic signal to recover the phonetic message (Liberman et al. 1967; Liberman & Mattingly 1985). Research using computer-generated speech in which the phonetically-relevant acoustic parameters can be independently varied has shown that native listeners integrate the multiple parameters associated with a phonetic contrast in order to arrive at a phonetic interpretation of the input stimuli (e.g., Polka & Strange 1985). Furthermore, studies of children learning their native lan-

1. This complexity in the acoustic structure of phonetic segments arises from the fact that the speech gestures associated with the realization of phonetic sequences are temporally coordinated movements of laryngeal and superlaryngeal articulators, and that gestures associated with adjacent (and even nonadjacent) phonetic segments overlap temporally, i.e., phonetic segments are coarticulated in real time.

guage have shown that these patterns of perceptual integration develop gradually over the first several years (Nittrouer & Miller 1997a, b) such that the perceptual weighting of the multiple spectral and temporal parameters comes to resemble the adult native-language pattern.

Studies of very young infants have shown that the ability to discriminate differences in phonetically-relevant acoustic parameters is remarkably well developed very shortly after birth. Again, using computer-generated synthetic speech materials, the pioneers of infant speech perception research demonstrated good discrimination by 1–4 month old infants of many phonetically-relevant temporal and spectral acoustic parameters (see Jusczyk 1997, for an overview of this research). Of greatest interest here are the cross–language studies in which infants exposed to different native languages were examined to determine if acoustic parameters associated with non-native phonetic contrasts were discriminated. For example, one study examined the perception of VOT differences underlying the English devoiced vs voiceless aspirated syllable-initial stops [b̥-p^h] and the Spanish fully voiced vs. voiceless unaspirated phones [b–p] (Lasky et al. 1975). When 6-month-old infants from monolingual Spanish environments were tested on both contrasts, they showed good discrimination of the [b̥-p^h] distinction, but not the [b–p] distinction, despite their exposure only to the latter phones in the ambient language. Their pattern of perception did not differ from English learning infants (Eimas et al. 1971). Streeter (1976) also reported that Kikuyu-learning infants discriminated the [b̥-p^h] contrast, even though Kikuyu has only the fully voiced labial stop [b].

More recent studies have employed carefully chosen utterances spoken by native speakers in which the multiple acoustic parameters associated with the phonetic contrasts co-occurred naturally, and phonetically-irrelevant acoustic variations were also present. Janet Werker and colleagues' studies of the perception of Hindi dental vs retroflex initial stop consonants by English-learning infants, children, and adults provide excellent examples of this research (see Werker & Tees 1999, for a review). In both cross-sectional and longitudinal tests, results showed that both Hindi- and English-learning 6-month-old infants could discriminate this difficult place-of-articulation contrast. However, by 11–12 months of age, English-learning infants failed to discriminate the same contrast when tested in the same paradigm, while Hindi-learning infants continued to perform well. Follow-up studies showed that adult English-speaking listeners and 4-year-old, 8-year-old and 12-year-old English-speaking children also failed to discriminate the Hindi place contrast. That is, it appeared that native English speakers had come to ignore the differences between dental and retroflex stops, since this phonetic difference is not phonologically contrastive in English. (It is interesting to note that both dental and retroflex stops occur in English as allophonic variants of /d/ as in "width" vs "drip"[wId̪θ-ɖɹIp].)

These findings corroborated the earlier conclusion that very young infants appear to be *language-general* perceivers. They are able to discriminate voicing and place-of-articulation contrasts in consonants that are not present in the ambient language, or that are present but constitute allophonic variants of a single phonological category. Rapid changes in perceptual abilities occur as a function of exposure to the distributional properties of the native language, such that, by the end of the first year (about the time infants produce their first words), the ability to discriminate non-native consonant contrasts is relatively poor. That is, older infants (and children) display *language-specific* patterns of phonetic perception (Werker & Curtin 2005).

Developmental changes in perception of vowel contrasts are less well documented, but also appear to show a change from language-general patterns of discrimination to language-specific patterns within the first year of life (Kuhl et al. 1992; Polka & Werker 1994). However, Polka and Bohn (1996), in a comparison of German- and English-learning 6 to 8-month old's and 10–12-month olds' perception of two vowel contrasts ([u/y] phonologically distinctive in German but not in English; [ɛ/æ] phonologically distinctive in English but not in German), failed to find either an age effect or an effect of phonological status in the ambient language. Both pairs were discriminated relatively well when the more peripheral vowel [u, æ] was the "change" stimulus against a background of repeated presentations of the more central vowel [y, ɛ], whereas discrimination was poor by both age groups when the more centralized vowel [y, ɛ] was presented as the "change" stimulus against a background of the more peripheral vowels [u, æ].[2] Thus, the course of developmental change from language-general to language-specific perception of vowels is still under study.

Cross-language studies of adult phonetic perception, employing many of the same stimulus materials and age-appropriate perception paradigms have documented language-specific patterns of perception of both consonant and vowel contrasts. Many non-native contrasts are very difficult for adult listeners to perceptually differentiate (see Strange 1995, for a review of this literature). In Section 2, we will show that some of these perceptual difficulties with non-native contrasts are resistant to change even after years of experience with a language for which they are phonologically contrastive. It appears that extensive use of learned patterns of selective perception, in the service of robust and efficient perception of the native language, results in highly automatic patterns of perceptual processing that are not easily modified by subsequent linguistic experience. (Research demonstrating the

2. This is a cross-category version of the within-category "magnet effect" reported by Kuhl et al., 1992; see also Kuhl & Iverson, 1995).

robust nature of native-language phonetic perception is described in more detail in Section 4)

To summarize so far, behavioral research on the perception of phonetically-relevant acoustic properties of speech segments has shown that:

1. Phonetic perception involves the selection and integration of multiple acoustic parameters in order to recognize (categorize) phonetic segments as tokens of phonological categories.

2. The ability to discriminate phonetically-relevant acoustic properties of speech sounds is present at birth or shortly thereafter. Infants' perceptual sensitivities to the complex acoustic patterns of speech provide them with the necessary tools to start to learn the phonological structure of their native language.

3. By the end of the first year of life, infants' perceptual abilities have been *reorganized* so that they begin to reflect the phonological structure of the native language input. That is, they have learned to selectively attend to those phonetic differences that are phonologically relevant in the native language, and to ignore many of the acoustic-phonetic differences not present or not used to distinguish phonological contrasts in the native language.

4. Over the next several years of life, children's selective perceptual processes are further modified such that the weighting of multiple acoustic parameters comes to resemble the adult patterns of the native language. More reliable acoustic parameters are given more weight, while phonologically irrelevant variations are given almost no weight. This allows the child to cope with the inherent variability in the phonetic realization of phonological segments which occurs within and across speakers, phonetic/phonotactic contexts, and speaking rates/styles.

5. In adults, native-language phonetic perception is *robust* and *automatic*. As we will see below, the ability to extract the phonetic message from the acoustic signal, even in non-optimal situations (unfamiliar talkers, competing noise, distracting tasks requiring the listeners' attention) requires few cognitive resources on the part of the native listener.

L2 speech perception: Variable perceptual difficulties with non-native contrasts

Exemplary behavioral studies of cross-language phonetic perception

As briefly stated above, studies of adult listeners' perception of non-native consonant and vowel contrasts have demonstrated markedly poorer performance than for native language listeners for many (but not all) of the phonetic contrasts investigated. A well-documented example of these cross-language difficulties in perception of consonants are the studies of native Japanese listeners' identification and discrimination of the English [ɹ/l] contrast (see also Bradlow's chapter in this volume). In studies using a variety of stimulus materials (both synthetic speech and natural speech utterances), a variety of tasks (oddity discrimination, categorial discrimination, identification), and listeners with a range of experience learning English as a foreign language, performance by most Japanese listeners is usually significantly poorer than for native English speakers (e.g., Miyawaki et al. 1975; Mochizuki 1981; MacKain et al. 1981; Best & Strange 1992; Yamada 1995). These experimental studies corroborate reports by Japanese L2 English learners that [ɹ] and [l] sound alike to them.

As mentioned briefly in the previous section, native English-speaking adults have difficulty perceptually differentiating the non-native dental/retroflex contrast [t̪/ʈ] in Hindi initial stops (Werker et al. 1981; Polka 1991), as well as a velar/uvular place contrast [k'/q'] in ejective stops in Nthlakampx (a North American Indian language) (Werker & Tees 1984a, b). However, Polka (1992) reported that English listeners performed better on the velar/uvular contrast in voiced Farsi stops [g-G] than on [k'/q'], at least when the former were presented first. Best and colleagues (Best et al. 1988) reported very good discrimination by adult English speakers of place and voicing contrasts among Zulu clicks. Werker and Tees (1983) also reported that English listeners did better on a non-native voicing contrast in Hindi stops [d^h/t^h], than on the place contrast. Flege and Wang (1989) reported that native Chinese learners of English could perceive voicing contrasts in final stop consonants when the stimuli included all the acoustic cues associated with clearly articulated syllables (preceding vowel duration, closure voicing, and release cues). However, when the final consonants were unreleased and closure cues were edited out, the Chinese listeners' performance deteriorated relative to the native listeners, who were able to maintain their perceptual differentiation of the distinction. Pikser (2003) reported similar results for native Spanish-speaking learners of English on final stop consonant voicing contrasts, although performance even on the over-articulated (released) stops was significantly below native-listener levels. Thus, perception of non-native place and voicing contrasts in consonants ranges from very poor (no better than chance performance in some studies) to quite good (although rarely as good as native listeners' performance), depending upon a host of variables to be considered below.

Studies of the perception of non-native vowel contrasts have also produced a range in performance by listeners with little or no experience with the L2, and with listeners with varying amounts of L2 experience. Polka (1995) reported that naïve adult English listeners had somewhat more difficulty perceptually differentiating the German lax front /back rounded vowels [ʏ/ʊ], than the tense pair [y/u], but both were distinguished quite well. In contrast, Levy and Strange (in press) reported relatively poor discrimination by both inexperienced and very experienced L2 learners for the French [y/u] contrast, replicating a finding first reported by Gottfried (1984) (see also Levy 2004). Flege (1995) reported that even experienced Spanish L2 English speakers had difficulties perceptually differentiating some non-native vowel contrasts [ɛ/æː, ɑː/ʌ], but could discriminate other contrasts [iː/ɪ, ʌ/ʊ] that either did not occur or were not phonologically distinctive in Spanish.

The above brief review of some of the cross-language phonetic perception studies is not meant as an exhaustive review of the literature, but rather is intended to illustrate the range in findings reported in the empirical research on this topic. Given the inconsistencies in the reported data, no blanket statement can be made about the extent of the perceptual problems facing late L2 learners, either at the outset of L2 learning or later on in their L2 language use. It is clear that significant perceptual problems exist in beginning late L2 learners, and that some of these problems persist over a period of several years. However, if we are to understand the nature of these difficulties and be able to predict which non-native contrasts will cause persistent problems for L2 learners, it is important to try to make sense of the often conflicting data published in this area. In the next section, we will discuss some of the stimulus and task conditions under which non-native perception is relatively easy, even for very difficult contrasts, and the stimulus and task conditions under which significant perceptual problems appear. Variation in performance as a function of these experimental variables sheds light on the nature of the underlying processes and may allow us to more accurately predict performance in real-life L2 learning situations. These considerations will also guide our discussion of current theoretical models of L2 speech perception and the role of perception in production of L2 phonetic segments and sequences.

Experimental tests of phonetic perception

Before discussing how experimental variables may affect the outcome of phonetic perception studies, a brief description of several experimental paradigms that have been used to investigate speech perception in adults is presented for those readers who come from non-experimental scholarly backgrounds. (See also Beddor & Gottfried 1995, for a more detailed discussion of methodological issues in cross-language speech perception studies.) Perception is, by definition, an internal mental (and physiological) process by which the perceiver recognizes incoming stimulus events as instances of mental categories. As stated above, perception of phonetic segments/contrasts involves not only the detection of differences in the acoustic signals that differentiate phonetic categories, but the accessing of internalized phonetic categories in order to make a decision about the identity of the stimuli. Behavioral research paradigms require that the participants indicate the outcome of this internal *categorization process* by making some sort of measurable responses. Experimental tests include the presentation of a set of stimuli arranged in some sequence to the participants, who then make overt responses based on their phenomenological experiences. Perception paradigms can be described in terms of the tasks performed, the test structure, and the stimulus materials presented to the listeners.

Perception tasks and test structures

Perception of phonetic segments/contrasts is experimentally tested using two general kinds of tasks: *identification* and *discrimination*. In an identification task, recorded stimulus materials are presented, one at a time, and listeners indicate their categorization of each presentation as an instance of a phonetic category, either by providing some sort of oral or written response (open-set task) or by selecting one of a set of response alternatives (usually orthographic symbols or key words) provided by the experimenter (closed-set task). In a discrimination task, two or more stimuli are presented, and the listener makes a decision about the relationship between the stimuli, i.e., whether they are the same or different. Many variations of each of these kinds of tasks have been used in L2 speech perception studies with adults.

In an identification task, repeated presentations of instances of each category are typically presented in random order and results are scored in terms of correct classification, relative to native-listener performance patterns. In addition, some experiments may measure *reaction times*, i.e., elapsed time from the presentation of each stimulus to the categorization response. If perception of the non-native segments/contrasts is relatively easy, reaction times (RT) should be faster. This provides a more sensitive measure of relative difficulty of non-native phonetic contrasts for more experienced L2 listeners who may categorize the segments at near perfect levels. While an identification task appears to be the most ecologically valid, in that it mimics what listeners do when attempting to comprehend a speaker's utterance, several problems can arise when testing the perception of non-native segments/contrasts. First is the problem of specifying the response alternatives; orthographic symbols may not be adequate if the listeners are not literate in the non-native language. This problem is exacerbated by the non-transparency of some orthographic systems (e.g., vowels in American English). On the other hand, oral repetition confounds the participants' problems of production of non-native phonetic segments with their perception of them. Thus, for studies of beginning L2 learners, identification tasks may not be feasible.

To avoid the problems of response alternatives, a discrimination task requires that listeners make comparative judgments about two or more stimuli that are presented in sequential order within each trial of the test. In the simplest paradigm (AX), two stimuli are presented; the second stimulus is either the same as the first (AA) or it is different from the first (AB); the correct response is "Same" and

"Different," respectively. No overt categorization response of either stimulus is required. An AX discrimination task in which the A stimulus remains the same from trial to trial within a block of trials (i.e., A is the *constant* stimulus) is thought to have the least *memory load* and *stimulus uncertainty*. Memory load refers to the requirement that the listener retain an auditory trace of previous stimuli while subsequent stimuli are presented for comparison. Stimulus uncertainty refers to the lack of predictability of which stimulus will occur in each sequential location within a trial, and which stimuli will occur from trial to trial. Performance on an AX task with a constant A stimulus is thought to reflect optimally the auditory sensory discrimination capabilities of the listener. However, performance on this task may vary from listener to listener because of differences in the criterion they adopt for deciding what constitutes a relevant difference between the stimuli (typically referred to as *response bias*). While there are ways to determine each listener's response bias and adjust scores accordingly, speech researchers have more often adapted this psychoacoustic task to include more complicated trial structures in which comparisons are required among three stimuli. Three trial structures have been used: ABX, AXB, and Oddity. In the ABX task, A and B are tokens of different phonetic categories and X is the same as A or B; after listening to all three stimuli (retaining auditory traces of them), the listener specifies whether X = A or X = B. In the AXB variation of this task, again A and B are tokens of different phonetic categories, and X is the comparison stimulus. However, this task is thought to constitute a smaller memory load because the comparison stimulus is equidistant in time from both A and B. Finally, in the Oddity paradigm, three stimuli are presented, two from the same category and one from a different category. All six possible combinations are presented randomly over trials within a test (AAB, ABA, BAA, ABB, BAB, BBA). The listener's task is to indicate which of the three sequentially presented stimuli is the different one. This task is thought to have the greatest memory load and stimulus uncertainty of the three trial structures. Thus, as the complexity of the discrimination task increases, performance outcomes begin to reflect not only basic auditory sensory capabilities but increasingly the cognitive processes involved in categorization (including implicit labeling of presented stimuli).

A final discrimination paradigm, referred to as the Category Change (or sometimes Oddball) task, was adapted from an infant speech perception paradigm for use with adults in both behavioral and neurobiological experiments (discussed in Section 4 below). In this task, repeated instances of one phonetic category serve as the Background or Standard. Interspersed from time to time, instances of the contrasting phonetic category are presented (either a single instance or sometimes three instances); this constitutes a Change or Deviant trial. The listeners' task is to indicate when they perceive a switch from the Background to the Change category. The Category Change task is similar to the AX task in terms of memory load and stimulus uncertainty and is thought to tap auditory discrimination capabilities, relatively independently of categorization processes.

Finally, in all of the above discrimination paradigms, the specification of Same and Different can be on the basis of *physical identity* or *phonetic category identity*. For instance, in the AXB task, A and B are physically different tokens (of different categories); however, X can be either physically identical to A or B (e.g., A1 A1 B1) or X can be a physically different token of category A or B (e.g. A1 A2 B1). This latter paradigm is called a *Categorial Discrimination* task (sometimes referred to as *Name Identity*). Categorial discrimination tasks require that the listener ignore acoustic differences that are not phonetically relevant in the target language, while attending to and responding on the basis of acoustic differences that differentiate the phonetic categories in the L2. As such, these tasks require phonetic processing of the stimuli rather than only the detection of acoustic differences. A particularly challenging version of this task (described below in more detail) is when the speaker changes within a trial, that is, when the three stimuli are produced by three different speakers (Gottfried 1984; Beddor & Gottfried 1995). In this task, listeners must categorize the phonetic sequences while compensating for speaker differences.

Stimulus materials

The selection of stimulus materials for speech perception studies varies considerably as a function of the empirical questions being addressed by the experimenter. Cross-language studies may use computer-generated synthetic speech materials that allow for rigorous control over the acoustic parameters that vary and those that remain constant. However, even the best synthetic speech materials often sound somewhat artificial. More importantly, for many phonetic contrasts of interest, the multiple acoustic parameters that are used by native speakers to differentiate the phonetic categories are not well understood. Thus, it may be difficult to relate results of studies using synthetic speech materials to perception of real speech. Alternatively, natural speech materials produced by human speakers can be recorded, and subsets of utterances selected that include both phonetically-relevant and phonetically-irrelevant acoustic variability. The careful selection of multiple tokens of each phonetic category is a time consuming job, requiring both careful listening and detailed acoustic analysis. This is necessary to insure that phonetically-irrelevant acoustic differences (e.g., intonation contour, overall amplitude, speaking rate) are not highly correlated with the phonetic contrast of interest. Ideally, these differences in phonetically-irrelevant acoustic differences should be equally distributed across phonetic categories such that listeners cannot base their perceptual decisions on them.

A second decision about stimulus materials is whether to use nonsense items or real words. L2 perceivers may invoke lexical representations of the real words in performing the perception task, while perceptual differentiation of nonsense items must be performed on the basis of phonetic/phonological knowledge abstracted from lexical knowledge. The confounding of the effects of lexical and phonetic knowledge when using real words may make interpretation of performance differences by naïve listeners (with no experience with the L2), less experienced, and more experienced L2 learners difficult. However, since these processes are always confounded in real world speech perception, experiments using real word materials more accurately reflect the receptive problems of L2 learners.

A third set of considerations is the choice of the phonetic and phonotactic contexts in which the target segments are imbedded. Studies have shown that differences between L1 and L2 phonotactic constraints influence the outcome of perception experiments. Thus, if an experimenter is focusing on the perception of non-native phonetic contrasts, independent of phonotactic aspects of L2 learning, he/she may want to select syllable structures and surrounding phonetic segments that also occur in the native language. The phonetic and phonotactic context in which the target phonetic segments are produced may change their acoustic structure drastically. Thus, results of perception tests of a particular contrast in a particular context cannot be generalized to other contexts. Recently, research on L2 speech perception has begun to examine these contextual effects on performance by systematically varying the phonetic and phonotactic context within the same experiment (e.g., Harnsberger 2000, 2001; Levy & Strange, in press).

Until recently, most studies of non-native phonetic perception have employed materials in which nonsense syllables or real words are produced and presented in isolation (i.e., in a citation-form style of speech), rather than in continuous speech contexts (phrases or sentences). Again, it has been well established that the acoustic parameters differentiating phonetic segments differ substantially as a function of this difference in speech style (e.g., Strange et al. 2007). Thus, studies using citation-form materials may not yield results that are easily generalizable to real-world situations in which perceivers are usually listening to and trying to comprehend continuous speech. Indeed, even "read speech" (i.e., speakers producing phrases or utterances as they read a protocol) differs considerably in acoustic structure from spontaneous conversational speech. However, the experimental assessment of perception of phonetic segments/contrasts in read speech contexts may be generalized to some real-world situations, such as the language classroom. (See also, the influence of "plain" vs "clear" speech styles on non-native listeners' perception of phonetic information [Bradlow & Bent 2002].)

The preceding brief description of some of the major design variables in experimental tests of speech perception studies provides a framework for a discussion of how these experimental variables interact with participants' knowledge of the L2 to determine the outcome of empirical studies of the perception of non-native phonetic segments and contrasts (see Werker & Curtin 2005, for a similar discussion of the role of task demands on L1 speech perception). In the next section, these issues will be addressed further by describing in some detail some key experiments that demonstrate how perception of non-native contrasts varies with the cognitive demands of the task and the nature of the stimuli.

Effects of experimental variables on perception of non-native contrasts

Native-language phonetic perception was characterized above as a process involving learned patterns of selection and integration of those acoustic properties of speech stimuli that are phonologically relevant in the native language. Under this analysis, it is assumed that the basic auditory sensory capabilities of children and adult speakers of different languages do not differ as a function of specific experience with a particular phonological system. Rather, the language-specific patterns of perception reflect differences in (higher-order) categorization processes (cf., Werker & Tees 1999; Strange 2002). Thus, if experimental paradigms test perception of non-native contrasts using stimuli and tasks that assess basic auditory sensory capabilities, rather than categorization processes, language-specific differences across listeners should be minimized (cf., Kewley-Port, Bohn, & Nishi 2005).

Support for this view of the locus of language-specific patterns of speech perception comes from studies using methods more closely resembling those of basic psychoacoustic research. These include tasks in which stimulus uncertainty and memory load are minimized, and in which participants receive enough practice with the materials and tasks to show optimal performance. A study by Strange and Dittmann (1984) illustrates how perceptual differentiation of a difficult non-native contrast can be shown to be native-like when performance is assessed with these techniques. In that study, Japanese L2 speakers of English were tested on 16 natural speech minimal-pair words contrasting [ɹ/l] in word-initial, initial cluster, intervocalic and word-final position, and on two synthetic [ɹ/l] 10-step stimulus series ("male" rock-lock; "female" rake-lake) in which temporal and spectral cues to the contrast co-varied appropriately. After pre-testing on all materials, participants took part in an All-Step (AX) discrimination task using the rock-lock synthetic series. On each day of testing, the constant stimulus (A) was either a good token of [ɹ] or a good token of [l]. For Different trials, the comparison stimuli (X) were the 9 other stimuli presented in random order; for Same trials, (A) was repeated, i.e., the two stimuli in Same trials were physically identical. During these tests, the participants received immediate feedback about the correctness of their responses. Subjects completed 14–18 half-hour sessions over the course of about 3 weeks. After these training sessions, performance on all natural and synthetic stimulus materials was again assessed (post-test).

Over the course of training with the rock-lock series, performance improved markedly and the post-test showed excellent discrimination of stimuli that American listeners labeled as different ([ɹ/l] pairs). Indeed, discrimination performance equaled or exceeded that of native English listeners with no training. When the Japanese listeners were asked to label the rock-lock stimuli in an identification task (with stimuli presented one at a time), their functions also looked native-like. Thus, we can conclude that Japanese listeners' ability to detect the acoustic parameters associated with this contrast was intact, and that, after some practice learning to attend to those acoustic parameters, their performance on the more demanding identification task reached native-like levels of accuracy. However, when post-test performance on the rake-lake series and on the minimal-pair real words was evaluated, there was little improvement over pretest levels by the Japanese listeners. That is, when memory load/stimulus uncertainty was greater and the materials were unfamiliar, the Japanese listeners still had difficulty categorizing the phonetic segments as [ɹ] or [l]. Thus, the training experience, in which listeners were attending to physical differences in particular (synthetic) stimuli, did not lead to a reorganization of phonetic perceptual patterns of categorization. (See Bradlow in this volume for a more extensive discussion of training procedures that lead to successful change in phonetic categorization.)

Another set of studies by Werker and her colleagues (Werker & Tees 1984b; Werker & Logan 1985) also illustrates how experimental variables tap into different *levels* or *modes* of processing of non-native phonetic contrasts. In the second study, stimuli were multiple natural tokens of Hindi dental and retroflex stops in consonant-vowel (CV) syllables and an AX discrimination task was used in which the time interval between the two stimuli (ISI) of each comparison pair varied (ISI = 250 ms, 500 ms, 1500 ms). Three types of pairs were presented: Physical Identity (PI) pairs (X=A), Name Identity (NI) pairs (two different tokens of a single phonetic category for Hindi listeners, e.g., two dental stops, or two retroflex stops) and Different (Diff) pairs (one dental token and one retroflex token). If English listeners were able to tap into auditory sensory capabilities, they would be able to discriminate physical differences in both NI and Diff pairs. If they were only able to tap into native-language categorization processes, they would be unable to discriminate either NI or Diff pairs. Hindi listeners, on the other hand should be able to discriminate Diff pairs easily even in the longest ISI condition (this is a phonological contrast for them). Results indicated that when the stimuli were temporally very close together (250 ms ISI), American listeners could discriminate NI and Diff pairs, despite the fact that these stimuli were all heard as the same phoneme. At the longest ISI, however, performance was poor on Diff pairs (the non-native contrast) initially, but improved with practice. Performance on the NI pairs (which were acoustically more similar than Diff pairs) did not improve. The Hindi listeners discriminated the Diff pairs at the longest interval, but not the NI pairs; that is, under the increased memory load, they utilized native-language categorization processes to respond. Werker and Logan interpreted these findings as supporting a three-factor theory of speech perception. Under conditions of high stimulus uncertainty and memory load, listeners will (at least initially) reflect *phonemic* processing, responding on the basis of native-language categorization processes. With practice and/or in less demanding tasks, performance may come to reflect language-general *phonetic* processing, in which non-native phonetic distinctions can be made. Finally, under minimally demanding conditions, auditory sensory abilities are reflected in that subjects can discriminate physical differences between tokens of the same phonetic category, i.e., acoustic differences that do not serve a phonological function in any language.

The above examples show that, in experiments in which the stimulus and task conditions tap auditory sensory capabilities (such as those shown by very young infants), adult listeners are able to demonstrate good perceptual differentiation of speech stimuli differing in phonetically-relevant acoustic parameters, whether or not those acoustic cues differentiate phonological categories in their native language. That is, despite years of employing learned patterns of selective perception and integration, adults can access those language-general processing abilities that they were born with. However, under more cognitively demanding conditions, listeners revert to their (automatic) language-specific patterns of perception. (See Strange 2002, and below for a further discussion of these modes of perception.)

In the above studies, good performance on difficult non-native contrasts was demonstrated by adult listeners only when the cognitive demands of the task were minimized. This included familiarity and practice (with feedback) with the stimuli, and simple discrimination tasks with short ISIs that allowed the listeners to make comparisons of auditory traces of sequentially presented stimuli. These sorts of discrimination tasks using carefully controlled stimuli, usually produced and presented in isolation, assess the listeners' perceptual abilities in a situation very different from real-world contexts. In the L2 language environment, L2 learners hear phonetic sequences (usually in continuous speech contexts) and must perceptually differentiate non-native phonetic segments "on the fly" in order to recognize the words of the utterance. That is, spoken word recognition requires rapid *identification* or *categorization* of phonetic segments, by reference to internalized representations of those categories. In the laboratory, then, tasks that examine categorization processes are probably more ecologically valid tests of L2 perception problems. As discussed above, identification tasks may be most appropriate for L2 learners who have some knowledge of the non-native phonetic categories and for whom the orthographic labels are unambiguous. This type of identification task cannot be used with naïve listeners who have no response labels for L2 phonetic categories; rather categorial discrimination tasks may be used.

Gottfried (1984) developed a *cross-speaker* categorial discrimination task to test both naïve English listeners and experienced L2 learners of French on French vowel contrasts. An ABX trial structure in which three stimuli were presented

(ISI = 1000 ms) was employed: the first two stimuli were tokens of different phonetic categories (e.g. [y] and [u]) spoken by two different speakers); the third stimulus was a token of either category A or category B, produced by a third speaker. For instance, a trial could consist of the following: speaker 1 [ty], speaker 3 [tu], speaker 2 [ty]. In this case, the correct response is "1" to indicate that the third syllable contained the same vowel as the first syllable. This task requires that the listeners categorize each vowel as an exemplar of the phonetic category [y] or [u] and ignore inter-speaker variation in the phonetic realization of those vowels. The relatively long ISI further increased the memory load of the task.

Using this task, Gottfried (1984) reported that both naïve English listeners and late L2 French speakers with many years of French experience had considerable difficulty with the front/back rounded contrast (as well as other contrasts). Levy and Strange (in press) extended this study in two ways: First, the French vowels were imbedded in nonsense disyllables [raCVC] produced in the sentence, "J'ai dit neuf raCVC a des ami." The sentences were then edited to include only "neuf raCVC a des ami" for presentation to listeners. Thus, the vowels were produced and presented in contexts more closely resembling continuous speech, and the cognitive demands of the task were markedly greater than for studies in which isolated vowels or monosyllables produced by a single speaker are presented. Second, the consonantal context was varied in the disyllables (rabVp, radVt) to examine the effects of consonant-vowel coarticulation on vowel perception. Results replicated those reported by Gottfried in that both naïve and very experienced L2 speakers of French had considerable difficulty with the [y/u] contrast. Moreover, the consonantal context in which the vowels were produced and presented had a significant effect on performance. Naïve listeners made more errors on front/back rounded [y/u] in alveolar than in labial context, whereas they made more errors on the front unrounded/rounded pair [i/y] in labial than in alveolar contexts. In contrast, experienced listeners made errors on the [y/u] contrast in both contexts, while their discrimination of [i/y] was very good in both contexts. Native French speakers could do this very difficult task with almost no errors. Thus, we can conclude that this cross-speaker categorial discrimination task is a sensitive measure of differences in higher-order phonetic perception processes in L1 and L2 listeners, and may be a better measure of their perceptual capabilities in real-world situations.

Even when the perception task imposes considerable cognitive load, as in the above paradigm, proficient late L2 learners may perform well under favorable listening conditions, i.e., when they are listening in a quiet environment with no distractions and their task is well defined. Again, outside the laboratory or language classroom, it is often the case that we must perceive speech in non-optimal conditions, and it is in these conditions that late L2 learners report that they have greatest communication difficulties. Research has corroborated their reports experimentally. For instance, Mayo et al. (1997) tested native English speakers and proficient native Spanish late L2 learners of English on the perception of key words produced at the ends of sentences (SPIN test). The sentences were either high predictability (semantic cues for the key word available) or low predictability (no semantic contextual cues). Perception was measured in two conditions: quiet and with competing speech babble (cafeteria noise) added. Both English and Spanish listeners performed almost errorlessly in quiet. For the noise condition, the level of the noise relative to the sentences (signal to noise ratio = S/N) in which each listener could identify 50% of the key words was used as the measure of performance.

Results showed that the native speakers could tolerate much more competing speech babble than the late L2 learners. Late learners, on average, made 50% errors identifying words when the S/N ratios were +3dB (high predictability sentences) and + 6dB (low predictability sentences). These levels of background noise are similar to everyday conversational situations; speakers typically increase the intensity of their speech to about 3 to 6 dB above the background noise. Thus, L2 listeners were making many errors even in these typical listening conditions. This contrasts markedly with the noise levels at which native English speakers performed at 50% accuracy (–3 dB for high, +2 dB for low predictability sentences). These differences in S/N ratios represent a doubling of the intensity (6 dB) of speech input needed for proficient non-native listeners to perform as well as native speakers even when semantic cues were available. That is, the L2 learners were less able to utilize semantic context to disambiguate words in the noisy environment than were native speakers.

In a recent study, Bradlow and Alexander (2007) replicated and extended this finding by varying the speech style used in the production of the stimuli from *plain* speech to *clear* speech. The latter style of speech is typically used when speakers are told that the listeners are hearing impaired or foreign speakers of the language. In this study, non-native listeners could take advantage of contextual information only when the sentences were produced in clear speech style, thus enhancing the phonetic information available.

The pattern of results in experiments on non-native phonetic perception by naïve listeners and late L2 learners described in this section can be summarized as follows:

1. Experimental paradigms which tap basic auditory sensory capabilities show that native listeners, naïve L2 listeners, and experienced L2 listeners are able to discriminate phonetically-relevant acoustic parameters which distinguish vowels and consonants, independently of the phonological function of those acoustic cues in their native language in optimal listening conditions. That is, years of experience

with a particular phonological structure does *not* result in changes in low-level sensory capabilities.

2. As the cognitive demands imposed by the stimulus materials and perceptual tasks increase, native-language perceptual patterns are more likely to be demonstrated. If we are interested in cross-language differences in phonetic *categorization* processes, stimulus materials should include within-phonetic-category variability as well as cross-category systematic differences. Embedding the to-be-differentiated segments in variable phonetic and prosodic contexts also renders the task more ecologically valid and will shed light on the relative difficulties of particular phonetic realizations of non-native contrasts. Such studies of L2 categorization indicate that even highly experienced late L2 learners continue to have perceptual difficulties with some L2 contrasts. However, the patterns of perceptual confusion differ as a function of L2 experience and reflect a reorganization of selective perceptual processes.

3. Perception in more realistic listening conditions reveals, further, that even L2 learners who have supposedly mastered L2 perceptual distinctions may nevertheless need to employ more cognitive resources for the task of phonetic categorization and word recognition. Even in conditions where stimuli are produced in enhanced speech style and semantic cues are provided, their performance suffers relative to the robust performance of native listeners in the same difficult listening conditions.

This pattern of results across different testing paradigms and different listening conditions supports the model of language-specific phonetic perception outlined in the first section. Language-specific patterns of performance are *not* due to differences in basic auditory capabilities of adult speakers of different languages. Rather, they reflect highly over-learned and efficient patterns of selection and integration of acoustic-phonetic information by which phonetic sequences are recognized. In adult listeners, these language-specific patterns of categorization have become automatic (requiring few cognitive resources) and highly robust even in difficult listening conditions. In her *Automatic Selective Perception* (ASP) model of speech perception, Strange (2006) refers to these automatic, language-specific patterns of perception as Selective Perceptual Routines (SPRs).

Beginning L2 learners initially come to the L2 listening task using their automatic L1 SPRs, which, in some cases, are not attuned to the most appropriate acoustic information for L2 phonetic segments (i.e., L1 interference). This results in perceptual difficulties on some non-native contrasts; when tested with stimulus materials and perception tasks that tap these selective perception processes, they show significant perceptual deficits, relative to native listeners. However, because basic auditory sensory capabilities remain intact, perception of non-native contrasts can and usually does improve with experience with the L2 phonological structures. Selective perceptual processes are *re-educated* with L2 experience such that many late L2 learners come to be able to perceptually differentiate even difficult contrasts under optimal listening conditions. That is, L2 SPRs can be acquired in adulthood. However, due to the influence of the L1, L2 SPRs may be based on different (non-optimal) weightings of acoustic parameters than those used by native listeners, even after years of immersion experience. Under difficult listening conditions which challenge the perceptual capacities of the listeners, L2 speakers' performance deteriorates more rapidly than native speakers' performance. It appears, then, that L2 SPRs differ from those of native listeners and may never be as fully automated as L1 SPRs.

Theoretical issues in L2 phonetic perception research

Current models of L2 speech perception

Most theories of L2 speech perception (Best 1995; Flege 1995; Kuhl & Iverson 1995) have focused primarily on characterizing the nature of the underlying perceptual representations of L1 and L2 phonological categories. Kuhl's *Native Language Magnet* (NLM) theory (cf., Kuhl & Iverson 1995) characterizes the initial changes in the underlying perceptual representations of phonological categories in a multidimensional acoustic-phonetic parameter space brought about by experience with L1 input. According to this model, perceptual reorganization from language-general to language-specific patterns of phonetic perception is due to the *warping* of phonetic space on the basis of distributional properties of L1 input (see Kuhl's Neural Commitment Model on her website for a more recent version of this theory).

Best's *Perceptual Assimilation Model* (PAM) (Best 1995) also employs the metaphor of a phonological space in which native categories (described as gestural constellations) are arrayed according to similarities/differences in their articulatory-phonetic structure. PAM was developed primarily to account for patterns of non-native segmental perception by naïve listeners with no experience with the L2. For these listeners, L2 phonetic segments are *perceptually assimilated* to L1 phonological categories on the basis of their gestural similarity to L1 phonetic segments, unless they are so phonetically disparate that they are heard as *uncategorizable* speech sounds, or if, indeed, they are not perceived as speech at all (*unassimilable*) (cf., Best et al. [1988] in which the perception of voicing and place contrasts among unassimilated Zulu click consonants was examined).

According to the PAM, when contrasting L2 phonetic segments are both categorizable as exemplars of L1 phonological categories, three patterns of perceptual assimilation predict relative discrimination difficulties by naïve listeners. If both members of the contrast are perceived as equally good exemplars of a single L1 category, discrimination will be most difficult (Single Category pattern). If both phonetic segments are assimilated as exemplars of a single L1 category, but differ in their perceived category goodness, then discrimination will be better (Category Goodness pattern). Finally, if the contrasting L2 phones are perceptually assimilated to separate L1 categories, then discrimination is expected to be excellent (Two Category pattern). (See Best, McRoberts & Goodell 2001, for an example of the application of this model to predictions of relative difficulties in discrimination of non-native consonant contrasts.)

More recently, Best and Tyler (2007) extended the PAM to predict patterns of speech perception by L2 learners (PAM-L2). They describe several patterns of cross-language assimilation at both the phonetic (allophonic, dialectal) level, and at the phonological level (lexical minimal pairs). Thus, L2 phonetic segments can be assimilated as more or less "good" exemplars of L1 phonological categories, based on differences in the details of their articulatory-phonetic realization in the two languages, or on the basis of similar phonological functions (e.g., phonotactic distribution, as in French /r/ vs American /r/). Patterns of phonetic/phonological assimilation, as well as the functional load in the L2, jointly determine the probability that an L2 contrast will come to be perceptually differentiated with L2 experience.

Flege's *Speech Learning Model* (SLM) (1995; Flege et al. 2003) focuses on how underlying phonological representations change with L2 experience. He claims that L1 and L2 phonetic subsystems exist within a single phonological space in experienced L2 learners. L1 and L2 phonetic segments can be related along a continuum from *identical* through *similar* to *new*, defined empirically in terms of acoustic similarity or perceived cross-language similarity. The degree of phonetic (dis)similarity determines whether L2 phonetic segments will be assimilated into existing L1 phonetic categories through a process of *equivalence classification* (for identical and more similar L2 phones) or whether, with L2 experience, separate L2 phonetic categories will be formed (for less similar and new L2 phones).

As applied to perception of non-native contrasts, the SLM (like PAM) predicts that if contrasting L2 phones are both assimilated to the same L1 category, discrimination will be difficult, as will be the differentiation of the L1 phones from the L2 phone. For example, for Spanish listeners, both American English [ɑ] and [ʌ] might be perceptually assimilated to Spanish [a], resulting in continuing difficulty discriminating this contrast, to accented production of both L2 vowels, and even to changes in the production of Spanish [a] (due to the dissimilation of L1 and L2 categories). If the L2 phonetic segment is very different from any L1 category (new in SLM, uncategorizable in PAM-L2), then it will not be assimilated to any L1 category. For instance, Flege and Hillenbrand (1984) considered French

[y] a *new* vowel for American listeners, and predicted that late L2 French learners would produce it accurately and perceptually differentiate it from both back, rounded and front, unrounded French vowels, as well as from English vowels (but see Levy & Strange, in press; Strange et al. 2004, 2005; Levy 2004, for conflicting data on perception of front rounded vowels by American listeners). According to this model then, L2 phonetic segments and contrasts that are very different from any distinctive L1 phonetic category will come to be perceived and produced relatively accurately, whereas L2 phonetic segments that are more similar to L1 segments will continue to be misperceived and mispronounced.

Empirical measures of cross-language phonetic similarity

In both SLM and PAM-L2, the concept of *cross-language* (L1/L2) *phonetic similarity* plays an important role in predicting initial and continuing difficulties in the perceptual differentiation of non-native contrasts. However, definitions of phonetic similarity differ across these models, and are not well-specified generally. In order that this concept not be completely circular, independent measures of cross-language phonetic similarity must be established which do not include *discrimination* of L2 and L1 segments or L2 contrasts. Four techniques have been employed in recent L2 speech perception research: (1) qualitative descriptions of articulatory-phonetic similarities (e.g. Best and Strange 1992; Best et al. 2003), (2) qualitative perceptual comparisons, involving the (narrow) transcription of L2 segments (e.g., Best et al. 2001), (3) acoustic comparisons of L1 and L2 phones (e.g., Strange et al. 2007; Flege 1987), including the use of a correlational technique called discriminant analysis; (4) direct measures of perceived similarity that involve presenting the L2 segments for listeners to categorize in terms of L1 phonetic categories. In a recent chapter, Strange (2007) describes and critiques acoustic and perceptual methods of establishing cross-language phonetic similarity. Recent research suggests that direct measures of perceived similarity are not always predictable from acoustic comparisons (cf. Strange et al. 2004, 2005). Thus, direct perceptual measures are probably a more valid way to determine L1/L2 perceived similarities for naïve listeners and L2 learners.

As an example of the direct method, the task used by Strange and her colleagues to examine the perceived similarity of German and American vowels by naïve American listeners (Strange et al. 2004, 2005) is described here. The non-native vowels were presented to listeners in different prosodic and phonetic contexts. Multiple tokens of each category were presented multiple times so that both within- and across-listener consistency in perceived similarity could be determined. On each trial, listeners first selected the L1 vowel category (indicated with key words) to which the L2 vowel was most similar, then they rated the *goodness* of the L2 vowel as an exemplar of that L1 category on a 7-point Likert scale

(1 = very foreign sounding; 7 = very English sounding). Scoring of such data included the consistency (over trials and listeners) with which each L2 vowel was assimilated to a particular L1 category and the goodness ratings (median value) assigned. If no particular L1 category was chosen on a majority of trials within and across listeners, we concluded that the L2 vowel was uncategorizable. Both the relative consistency in categorization and the judged category goodness were used to determine whether contrasting L2 vowels constituted Single Category, Category Goodness, or Two Category patterns (PAM). For instance, German [u] and [y] were both categorized as most similar to AE [u] in all contexts. However, in citation-form (hVp) syllables, [y] was judged a poorer exemplar of AE [u] than was German [u] (Category Goodness pattern). In bVp, dVt, and gVk syllables embedded in a sentence context, however, German [u] and [y] were categorized as equally good exemplars of AE [u] (Single Category pattern).[3] We predicted, then, that in continuous speech contexts, English L2 learners of German would have considerable difficulty differentiating this vowel contrast.

Cognitive mechanisms in L2 phonetic perception

The theoretical models of L2 phonetic perception summarized above, and the research generated by those models, are primarily concerned with predicting relative difficulties in the perception (and production) of non-native contrasts by naïve listeners and late L2 learners. They also make somewhat different claims about the nature of the representations of phonetic categories in adult language users, and the basis for characterizing cross-language similarities. However, neither model can be considered a theory of the mechanisms of speech processing. That is, they do not consider in detail the online processes involved in recovering the phonetic message from acoustic signals, and how those processes may differ for perception of L1 vs. L2 phonetic sequences, or for inexperienced vs experienced L2 learners. While both SLM and PAM/PAM-L2 use processing metaphors (perceptual equivalence classification, perceptual assimilation) to characterize L2 listeners' perception of non-native contrasts, neither model directly addresses issues about the nature and role of attention processes or the employment of cognitive re-

3. A surprising finding was that the front rounded vowels were more similar to front unrounded American vowels in terms of spectral structure (formant frequencies) when produced in citation-form syllables, while perceptually they were considered much more similar to American back rounded vowels in this context. However, when acoustic comparisons of AE and German vowels produced in labial, alveolar and velar contexts were performed, German front rounded vowels were more similar to back rounded AE vowels which are "fronted" in alveolar, and to a lesser extent in velar contexts. That is, perceptual similarity judgments appeared to be based on context-independent similarities between distributions of native language categories.

sources in the course of phonetic categorization. In PAM-L2, Best and Tyler (2007) introduce the notion of *attentional focus* at either a phonetic or a phonological level of analysis, but do not address under what conditions these differences in attentional focus are invoked.

Strange's ASP model is being developed to begin to address these issues. In this model, two *modes* of online processing of speech materials are proposed: a context-specific *phonetic mode* of processing that requires attentional resources, and a *phonological mode* of processing that is fully automatic for L1 speech processing (requiring minimal cognitive resources). The extent to which these modes of processing are tapped in perceptual tests of non-native speech is a complex function of the listeners' L1 and L2 experience, and the stimulus and task structure. When stimulus materials are relatively simple (e.g. citation-form utterances) and task structure (and instructions) direct listeners to attend to the context-specific, phonetic structure of the stimuli, naïve listeners and L2 learners can respond on the basis of detailed phonetic (dis)similarities between L2 and L1 segments and between non-native L2 contrasting segments. As the stimulus materials become more complex and task demands greater, online perceptual processing of L1 input is accomplished via automatic selective perceptual routines (SPRs); this phonological mode of processing is rapid and robust in non-optimal listening conditions.

It is hypothesized that L2 learners, when faced with complex L2 stimulus input and greater task demands, may resort to their automatic L1 SPRs. Thus, when their attention is otherwise occupied, as when they are trying to comprehend the semantic intent of the message, they may fail to differentiate non-native phonetic contrasts that, under simpler conditions, they can discriminate. Alternatively, if they have established L2 SPRs after considerable experience with the L2, they may use them to perform the required task. However, these L2 SPRs may not be based on optimal weightings of the acoustic-phonetic parameters specifying the L2 phonological categories and may not be as fully automated as L1 SPRs. Thus, performance may suffer, especially in non-optimal listening conditions. These issues of the role of attention and automaticity in L1 and L2 speech processing are discussed further below.

Theoretical implications of L2 phonetic perception research

While the models of L2 phonetic perception described above can be considered "works in progress" that are constantly being modified as more studies of the phenomena are completed, there is some convergence among theorists and researchers about the nature of L2 phonetic perception:

1. Phonetic perception by late L2 learners can be said to reflect interference from L1 phonological structures. Thus, *contrastive analysis* of L1 and L2 phonological

structures should provide specific predictions about beginning L2 learners' problems in perceiving (and producing) L2 phonetic segments. However, such contrastive analyses must be performed using descriptions of phonological structures that capture the details of phonetic realization and allophonic variation. Flege (1995) proposes that the appropriate units of analysis are *context-sensitive systematic allophones* of phonological categories (see also Strange et al. 2004). Detailed descriptions of the acoustic and articulatory characteristics of L1 and L2 phonetic segments as they are produced in a variety of phonotactic and phonetic contexts will help in fully characterizing cross-language phonetic similarities and differences that play a role in L2 perception and production problems (Strange et al. 2007). However, acoustic or articulatory descriptions alone may not be sufficient.

2. The concept of *perceived* cross-language phonetic similarity is central to predictions of relative difficulty in perception of non-native contrasts. Direct measures of L1/L2 perceived similarity have been developed, and appear to be more predictive of discrimination and categorization problems than either abstract analyses of phonological structures or context-specific acoustic comparisons of phonetic categories.

The theories described above focus on L1/L2 relationships as they are predictive of L2 phonetic perception problems. Implicit in these theories is the characterization of these perceptual problems as being due to learned patterns of selective perception (rather than to sensory deficits), and therefore, subject to modification by experience with a new phonological system. As described above, perceptual responses can reflect (non-linguistic) auditory, language-general or language-specific phonetic, or language-specific phonological modes of processing. Because perceptual responses can (under some conditions) reflect basic auditory sensory abilities, independent of the phonetic relevance of the acoustic parameters, it should be mentioned that phonetic contrasts differ in their *psychoacoustic salience,* i.e., in the distinctiveness of their acoustic structures. For instance, it has been suggested that temporally-cued phonetic contrasts are perhaps more salient than spectrally-cued contrasts (e.g., Bohn 1995). Thus, in general, place-of-articulation contrasts in consonants, cued primarily by spectral differences of short duration, may be considered less salient than voicing contrasts, cued primarily by temporal parameters. For vowels, contrasts in vowel quantity (length) may be more acoustically salient than contrasts in vowel quality (height and position).

Contrasts in manner of articulation (e.g., fricative vs. stop) may be considered very salient, in that they are differentiated by differences in source sound characteristics (e.g., presence of sustained noise vs. silence). For instance, while neither [z] nor [t] occur word-finally in Spanish, and final [s] and [d] are usually not realized phonetically in New World (Caribbean) dialects, Pikser (2003) showed that native Spanish L2 learners of English had no difficulties discriminating English syllable-final fricative/stop contrasts [z/d, s/t] in an AX categorial discrimination task with VC monosyllables. They did, however, have difficulty on voicing contrasts in both stops and fricatives. Much more work is needed to establish the psychoacoustic salience of acoustic-phonetic parameters, and to examine how language-specific (learned) patterns of perceptual weighting of those parameters interact with their (non-linguistic) psychoacoustic salience to determine discrimination performance.

Neurobiological studies of cross-language phonetic perception

Electrophysiological measures of discrimination

Behavioral methods have provided detailed information concerning the endpoint of speech processing. However, they are limited in their ability to present a clear picture of the sequence of internal processes leading up to the behavioral response. Neurobiological methods provide us with means to examine more directly the physiological processes preceding the behavioral response. In particular, the electrophysiological method of event-related potentials (ERPs) has already been useful in examining speech perception processes in L1 and L2 listeners. ERPs provide fine-grained temporal information concerning the auditory processing of input stimuli well before any behavioral response is planned or executed.

The ERP is the average of portions (i.e., epochs) of the electroencephalogram (EEG) that are time-locked to some stimulus event (e.g., acoustic onset of a syllable). The EEG is the product of electrical activity resulting from firing of neurons, which propagates to the scalp surface where it is recorded by electrodes. The electrodes are labeled with standard names indicating there location on the scalp (F = frontal; C = central; P = parietal; T = temporal; O = occipital; odd numbers are on the left, even numbers on the right, and "z" on the midline, e.g., F3 = frontal left, Fz = frontal midline). ERPs have a characteristic time-course to particular types of auditory stimuli at particular scalp-electrode locations (e.g., Central midline, Cz). For example, Figure 1, top left, shows an average of the ERP across ten participants recorded to a Hindi retroflex stop-vowel syllable [ɖɑ] and a bilabial stop-vowel syllable [bɑ] that were presented to participants in a sequence of repeating syllables (ISI = 535 ms). In this particular experiment, stimuli were presented in the so-called Oddball (Category Change) paradigm in which one stimulus is presented frequently (Standard) and a second stimulus infrequently (Deviant). For each participant, the ERP to the standard stimuli consisted of an average of approximately 1200 trials (80% of total), while the ERP to the deviant stimulus consisted of an average of approximately 240 trials (20% of total). At the frontal central midline sites (e.g., Cz) a slight positive (P1) deflection is observed peaking around 90 ms, followed by a negative deflection (N1) peaking around 140 ms and a second positive deflection (P2) peaking around 240 ms. ERP peaks are often named according to their polarity (P = positive, N = negative), sequence of occurrence in time (e.g., P1 = first positivity) or latency (e.g., P300 = positivity peaking at 300 ms). This sequence of deflections inverts in polarity at the mastoid sites (behind the ears; left mastoid [LM]), as shown in the top right graph. This pattern of deflections in time (morphology) and across the scalp (topography) is characteristic of ERPs evoked to auditory stimuli. The latency, amplitude and topography of the sequence of peaks, P1, N1, P2 are dependent on the physical properties of the stimulus, and, thus, are called *obligatory components*.

The ERP can be modulated by factors that are not directly related to stimulus properties. In particular, a negative-going deflection is observed at frontal and central superior scalp sites (including Fz and Cz) if a rare auditory event is presented in a sequence of frequent auditory events. For example, less frequent [ba] syllables (deviants) presented among frequent [ɖa] syllables (standards) lead to a negativity peaking between 210 and 240 ms in adults (Shafer, Schwartz & Kurtzberg 2004). The top left graph in Figure 1 illustrates the negative deflection of the ERP to the deviant stimulus compared to the standard. This negative deflection has been named the *Mismatch Negativity (MMN)*. The nature of MMN makes it an ideal

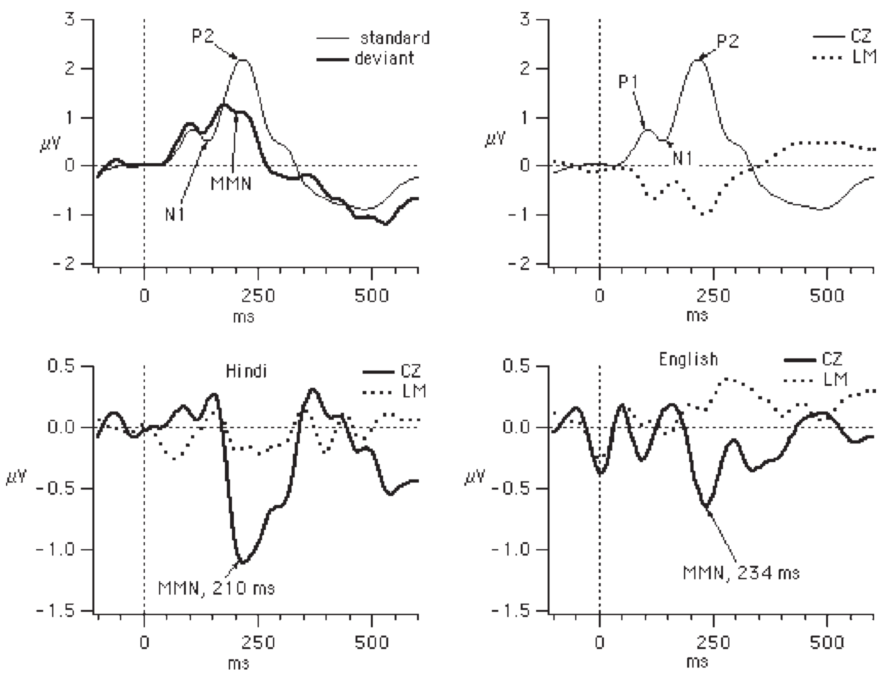


Figure 1.

tool for examining how language experience affects speech processing. The MMN can be elicited even when the participant is asked to ignore the incoming auditory stimuli and attend to some other input (e.g., a video). Thus, there is considerable evidence that the MMN indexes a *pre-attentive* comparison process because it can be obtained without attention to the auditory stimulus input (see Naatanen 1990) in L1 listeners. Attentional focus, however, can affect whether an MMN is elicited to more complex stimuli or patterns (e.g., Sussman, Ritter & Vaughan 1999). The latency and amplitude of the MMN are correlated with the difficulty of the discrimination between standard and deviant stimuli, and appear to be influnced by attentional focus when listeners are presented non-native contrasts (Hisagi 2007).

An MMN can be elicited, not only to an increase, but also to a *decrease* in amplitude or duration of the deviant stimulus, relative to the standard. A decrease in these stimulus parameters leads to less activation of cortex involved in the detection of auditory input. Thus, if the MMN were an index of detecting physical stimulus properties alone, it would not be elicited to a decrease in stimulus intensity or duration. The presence of an MMN to both increases and decreases in these stimulus properties supports the claim that it is an index of *change detection* by the auditory system, rather than lower-level sensory processing of acoustic differences (i.e., new afferent activation).

Several additional ERP components can also be elicited to a deviant stimulus in an oddball paradigm. These components provide further information regarding the processes leading up to a behavioral response. The P3a component indexes orienting to a deviant stimulus in a passive (i.e., ignore) task. It is largest at frontocentral sites and follows the MMN in time. Thus, this component indexes whether a participant is consciously aware of a change to a deviant stimulus. The N2b and P3b are components elicited to the deviant when it serves as a target in a behavioral task (e.g., press a button to the deviant, or count the deviants). The N2b is a negativity following (and sometime overlapping) the MMN in time and is largest at central sites. The P3b is a positive component following N2b and is largest at central–parietal sites. The amplitude and latency of these two components are more directly related to the behavioral response than MMN.

Results of cross-language MMN studies of phonetic perception

Beginning in the 1990's, researchers began exploring the question of whether the pre-attentive discrimination indexed by MMN reflected only the auditory processing of the acoustic properties of speech sounds, or whether it might also reflect higher-order perceptual properties relevant to categorization processes. Results, to date, have been somewhat equivocal. Some studies have found evidence indicating that MMN is sensitive to acoustic differences between speech sounds, in that the MMN increases in amplitude and decreases in latency with increasing acoustic difference between the standard and deviant (e.g., Aaltonen et al. 1987; Sams et al. 1990; Maiste et al. 1995; Sharma & Dorman 1998). This implies that the MMN indexes a basic auditory sensory level of processing. However, other studies suggest that the MMN reflects phonetically-relevant differences (e.g., Aaltonen et al. 1997, Sharma & Dorman 1999; Szymanski et al. 1999). Evidence that MMN is sensitive to phonetic category differences is seen as larger amplitude MMNs to a pair of stimuli crossing a native-language phonetic boundary (e.g., [t] to [d]) than to a pair within the same phonetic category (e.g., allophones of /d/) even though the cross-category and within-category pairs differ acoustically by the same amount in Voice Onset Time.

Of greatest interest here are the more recent investigations examining MMN patterns when listeners are presented native versus non-native phonetic contrasts. Results generally support the suggestion that MMN can reflect both language-general phonetic and language-specific phonemic levels of processing (Naatanen et al. 1997; Winkler et al. 1999; Winkler et al. 1999; Shafer et al. 2004). For example, Naatanen and colleagues (1997) found a smaller amplitude MMN to a non-native Estonian vowel contrast for Finnish speakers than for a native contrast that was actually acoustically less differentiated than the non-native contrast. In a series of investigations with Finnish and Hungarian speakers, Winkler and colleagues found that experience with a pair of contrasting vowels, either through exposure from birth (Winkler et al. 1999a) or through L2 learning (Winkler et al. 1999b; Jacobsen et al. 2004), led to a larger MMN than shown for a group with no prior experience with the contrasting vowels. Finally, Shafer and colleagues (2004) employed the MMN to investigate the perception of Hindi stop consonants by Hindi and American listeners. They observed a shorter latency in MMN to a bilabial/retroflex contrast [b/ɖ] for native Hindi listeners compared to English-speaking listeners (see Figure 1 bottom left and right).

Despite these general findings in support of MMN reflecting a language-specific level of processing, there are some less easily explained results from these investigations that will need to be examined further in future studies. One finding, in particular, has not been explained in previous investigations. First, it has generally been believed that any contrast (speech or non-speech) that can be behaviorally discriminated will elicit an MMN. However, this claim was not supported by Shafer and colleagues (2004). In the first experiment, an MMN was not elicited to the dental/retroflex contrast [d̪/ɖ] in Hindi stops, even though both the Hindi and English-speaking groups demonstrated behavioral discrimination of this contrast in a Category Change task in which they were asked to press a button anytime they heard a stimulus differing from the frequent standard (see Dehaene-Lambertz 1997 for a similar result with French speakers). Both groups showed an MMN to the bilabial/dental contrast [b/d̪], which differed on the relevant acoustic parameters of formant transition onset frequencies by the same amount as did the dental/retroflex stimuli (F2 difference 400 Hz, F3 difference 200 Hz). The Hindi-speakers had been expected to show both good behavioral discrimination and an MMN to the native dental/retroflex contrast; thus the absence of an MMN in this case is somewhat puzzling. One possible explanation for this finding is that attention is necessary to elicit MMN to this very difficult phonetic distinction, even for speakers for whom it is phonologically contrastive. As discussed above, the participants were instructed to ignore the speech stimuli and read a book or watch a silent video in this passive attention condition.

An investigation comparing native and non-native listeners' processing of Japanese vowel and consonant duration contrasts under differing conditions of attention revealed that attending to the duration contrast minimized the difference in MMN between the native (Japanese [JP]) and non-native (American English [AE]) listeners (Hisagi 2007). Specifically, the AE group showed smaller MMNs to the vowel duration contrast than the JP group in a task directing their attention away from the auditory stimuli, but less difference when attention was directed to the auditory stimuli. However, no meaningful group differences in the MMN to the consonant duration contrast were observed, and the MMN to this contrast was much smaller than to the vowel duration contrast for both language groups. This finding is similar to that for the Hindi dental/retroflex contrast in that the MMN was much smaller for a consonant than a vowel contrast and did not reveal group differences. The stimuli used by Hisagi included variable tokens of the standard and the deviant, which increased the difficulty of identifying the duration difference and increased the likelihood that listeners' would rely on categorical knowledge. The JP listeners showed better categorization of the consonant duration contrasts (in a behavioral task) than the AE group and also showed larger MMNs to this contrast when attending to it compared to attending to a visual task. These findings support the suggestion that some phonetic distinctions are psychoacoustically more difficult than others and may require attention to perceptually differentiate even by native listeners.

We recently examined the role of attention on MMN indices of phonetic perception, using a vowel contrast that is phonemic in English, but not Spanish [ɪ/ɛ]. In general, we found that early L2 learners of English (age of onset [AO] under 5 years) showed robust MMNs, regardless of the task, while late L2 learners (AO over 18 years) showed no MMN (Garrido, Hisagi, & Shafer 2005). Early L2 learners, however, showed some differences in a later ERP component (a late negativity [LN]} related to attention (see Shafer et al. 2005; Datta et al. 2006). The participants' attention was manipulated by asking them to respond to an infrequent [ba] and ignore an infrequent [da] that occurred among the [ɪ/ɛ] vowels in one task. In a second task they responded to an infrequent high-pitched tone and ignored a low-pitched tone among the vowels. In the third task, they ignored all the auditory stimuli and watched a video with the sound turned off. Participants were instructed to press a button when they heard the interspersed target stimulus. Monolingual, but not bilingual listeners' showed enhancement of the LN in the task in which they responded to [ba] compared to the tone task. That is, attending to the spectral cues in the speech target task led to enhancement of the LN response to the vowel difference for the monolinguals, but not bilinguals. These findings suggest that at an early level of preattentive processing there is little difference between monolinguals and bilinguals, but that at a later level requiring attention, they differ in how they treat contrasts present in one language but absent in the other. Our interpretation of these findings are that the bilinguals are less ready to commit resources to processing stimuli as members of one of their two languages without further (perhaps semantic) cues to the language.

Theoretical and methodological issues in electrophysiological studies of phonetic perception

Previous investigations using ERPs to examine speech perception have generally focused on whether a certain measure (e.g. MMN, or P3b) is sensitive to acoustic, phonetic or phonemic modes of processing. This narrow focus has limited the usefulness of these studies in answering fundamental questions regarding the time course of phonetic perception. Specifically, one interesting question is what modes of speech processing are affected by L1 versus L2 learning? This question can be examined more fully if researchers use ERP components as indices of the time course of processing rather than using speech perception as a test of a specific component (e.g., MMN). To illustrate this point, the ERP components N1, MMN, N2b and P3b can all be examined together to determine the time course of processing a stimulus. N1 reflects an early stage of cortical processing. The amplitude of the N1 is sensitive to the ISI in that it is smaller in amplitude to shorter ISIs (called refractoriness). A great deal of research has shown that N1 reflects new afferent input to primary auditory cortex. Thus, the amplitude of N1 to a deviant in an oddball paradigm can indicate the acoustic similarity (and therefore, sensory resolution) between the standard and the deviant. As discussed above, MMN reflects detection of a pattern, and thus indexes a later stage of processing than the N1; finally, the N2b and P3b reflect even later stages of perceptual processing leading to a behavioral response. The latency of these late components, in part, reflects the difficulty of the decision. Examination of all of these components during the same task will allow for a more global view of the time course of phonetic processing.

It will be important in future studies to examine the neurophysiology of phonetic perception using more ecologically valid stimuli. To date, most studies have used synthetic V or CV stimuli, although a few studies have begun to use words (e.g., Winkler et al. 2004; Dehaene-Lambertz, Dupoux & Gout 2000; Hisagi 2007). Combining the results of ERP studies with brain imaging studies will also be an important direction for future studies, since methods such as functional Magnetic Resonance Imaging (fMRI) can provide high-resolution localization of speech processing (e.g., Binder & Price 2001).

Tetrahedral framework for speech perception experiments

The review of the empirical literature presented above points to a clear need to consider a whole host of variables when designing experiments and interpreting the data obtained in behavioral and neurophysiological studies of phonetic perception. Since perception is, by its very nature, an unobservable mental event, it is especially important to consider how the methods used to elicit overt behaviors and brain responses might affect the outcome of the experiment and the answers to our theoretical questions about the nature of L1 and L2 phonological representations and the cognitive processes by which L2 learners decode the phonetic message from L2 speech utterances. To provide a framework within which to consider these methodological variables as they relate to theoretical questions, the Tetrahedral Framework, first introduced by James J. Jenkins for the study of memory, is described here (Jenkins 1979; see also Strange 1992).

In this framework, four sets of variables define the points of a tetrahedron, with lines drawn between all combinations of points. The four sets of variables include: *subject variables*, *stimulus variables*, *orienting task variables,* and *criterial task variables*. The lines connecting the points are meant to represent the complex interactions among all of these variables that determine the outcome of any experiment. In studies of L2 phonetic perception, important *subject variables* include: native language and continued L1 use, L2 language experience (type of instruction, years of immersion, daily use, etc.). Age of acquisition is also a very important determinant of L2 speech perception abilities (see Ioup in this volume for a discussion of critical or sensitive periods for language learning). Finally, it should be mentioned that most studies of adult L2 learners' perception of non-native phonetic contrasts report a considerable range in the performance of individual participants that cannot be easily correlated with other subject variables listed here; these individual differences have often been labeled *talent*.

Stimulus variables include the particular contrasts studied (as they relate to L1 categories) and the psychoacoustic salience of the acoustic cues for the contrasts. It has long been known that perception of vowels and consonants differs markedly, due to the inherent acoustic differences in these classes of speech sounds (see Strange 1995). In addition to the type of contrasts examined, experiments differ in their use of computer-generated vs. natural speech materials, and for the latter, in the selection of tokens and speakers producing the materials. In addition, decisions must be made about the choice of contexts in which the contrasting phones are produced and presented, and whether to use citation-form syllables/words (produced in lists) or test syllables/words embedded in short phrases or sentences. The sequencing of stimuli may have significant effects on performance. For instance, tests in which different speakers' utterances are blocked, rather than sequenced randomly, or systematically varied (as in the cross-speaker categorial discrimination task) will have an effect on the outcome. If multiple phonetic and phonotactic contexts are used, then the use of a blocked or mixed list design affects the cognitive load. In the former, listeners can anticipate the context in which the target phones will occur (i.e., stimulus uncertainty is lower); in mixed lists, they cannot. Finally, as we have reported above, a seemingly trivial decision about the timing of trials (ISI) within a test can often determine the mode of processing tapped by the experiment.

Orienting task variables refer to decisions and procedures that affect a) the participants' understanding of what is being examined and the nature of the task, and b) the activities they participate in before and during testing. For instance, in the MMN studies, it appears that performance may vary significantly as a function of whether or not the participant attends to the incoming stimuli (with or without a required response). This may interact with subject and stimulus variables in that L2 learners may have to attend more than native listeners, and even the latter may have to attend to non-salient contrasts to show detectable MMN patterns. In behavioral studies, instructions and familiarization procedures can bias the participants toward using different modes of processing in making their response decisions: i.e., whether to attend to physical differences in the stimuli (auditory level processing), or to process the incoming utterances in a language-general or language-specific mode. The number of trials (i.e., within-experiment familiarity with the stimuli) and overall task difficulty may also influence participants' motivation and attention to the task.

Criterial task variables refer to the type of responses required of the listeners and partially overlaps with orienting task variables described above. As discussed earlier, physical-identity discrimination tasks (in which "different" responses refer to physical differences of the stimuli) tap different processes than do categorial discrimination tasks. In the latter, only phonetically-relevant physical differences define what is "different," while physical differences that constitute within-category variations must be ignored. Finally, in identification tasks (and some perceptual assimilation tasks) the stimuli must be compared against internal representations of phonetic/phonological categories.

The ASP model of speech perception outlined above (see also Best & Tyler 2007; Werker & Curtin 2005), which characterizes perception of speech sounds as reflecting both phonetic and phonological modes of processing, provides a framework within which choices about what stimuli to employ, what instructions and familiarization tasks to use, what criterial task to use, and how to manipulate attention, memory and cognitive load can be motivated. In designing and interpreting studies of phonetic perception, the balance between experimental rigor (control of confounding variables that threaten internal validity) and ecological validity (external validity) must be considered. No one experimental paradigm is *best* in examining the phenomena of L2 phonetic perception. A good research strategy is to vary experimental designs along several of the dimensions known to have important influences on the outcomes. When experimental data from several paradigms, within which subject, stimulus, orienting-task and criterial-task variables interact in various ways, all converge on the same answers to experimental questions, we can have more confidence that the findings reflect the true nature of the perceptual processes in which we are interested.

Conclusions

In this chapter, we described current theoretical and empirical issues in research on the perception of phonetic contrasts by adult L2 learners. Auditory perception of phonetic contrasts was characterized as an *active process* which involves the *selective* detection and integration of multiple acoustic parameters in order to recover the phonetic segments/sequences that differentiate lexical items. Using this conception of phonetic perception, we can summarize the nature of the problems encountered by adult learners of a foreign language and interpret the sometimes conflicting empirical research on the nature and extent of the perception problems of L2 learners as follows:

1. In the course of acquisition of the native language, patterns of *selective perception* become language-specific; phonetically-relevant acoustic information that serves reliably to distinguish phonological contrasts in L1 is weighted more heavily than acoustic information that is not as relevant to L1 phonology. In adult native-language users, these L1 selective perceptual routines are highly robust and automatic, requiring few cognitive resources (little attention) even under non-ideal listening conditions. Thus, cross-language differences in perception of consonant and vowel contrasts do *not* reflect differences in basic psychoacoustic discrimination abilities across language groups, but rather reveal automatic categorization processes by which listeners selectively detect phonetically-relevant differences in the stimuli while ignoring other acoustic variations which constitute within-L1-category variability.

2. For adult learners of a foreign language, these L1 automatic selective perception routines may interfere with their ability to perceptually differentiate some phonetic contrasts in the new language. Initially, non-native phonetic segments may be perceptually assimilated to native phonological categories, resulting in percep-

tual confusions in tasks in which categorization is assessed. However, the ability to discriminate non-native phonetically-relevant acoustic parameters remains intact in adults and can be accessed under stimulus and task conditions that reduce cognitive demands and that allow the listeners to (learn to) attend to the appropriate acoustic structures. Thus, adult learners of an L2 can and do improve in their ability to differentiate non-native contrasts, i.e., they can develop L2 selective perception routines. However, phonetic perception of non-native contrasts may never become as automatic and robust as perception of native contrasts.

3. Cross-language perception studies of L2 learners have used a variety of stimulus materials and tasks to explore perception by a variety of L2 learners who differ in their L1/L2 experience and usage. Different experimental paradigms reflect different modes of online processing of speech input. If we are interested in *phonetic and phonological categorization processes*, we must design studies that begin to reflect real-world stimulus and task constraints in language processing (including those of the language classroom and the L2 work environment) while maintaining experimental control and rigor. In that way, we will be better able to predict the initial and continuing perceptual difficulties of L2 learners and to design better instructional materials and tasks to improve perception of difficult L2 phonetic structures.

4. More research is needed on the cognitive mechanisms involved in phonetic perception. Cross-language research using event related potentials and brain imaging to study physiological processes involved in speech perception may shed light on the structures and functions associated with psychoacoustic and linguistic modes of processing phonetic sequences. However, to date, the stimulus materials and tasks used in most of these studies are rather restricted, and may not tap selective phonetic perceptual processes appropriately. Future studies, using both brain and behavioral indices of perception, as well as laboratory training studies (see Bradlow, this volume), should be designed to investigate these underlying mechanisms, how they differ for L1 and L2 speech perception, and how they change as a result of L2 language experience in the laboratory (training), the language classroom (instruction), and the L2 environment (immersion experience).

Suggested readings

Basic Texts on Acoustic Phonetics and Speech Perception:

Johnson, K. 2003. *Acoustic and auditory phonetics,* 2nd edition. Malden, MA: Blackwell Publishing.

Pickett, J. M. 1999. *The acoustics of speech communication: Fundamentals, speech perception, theory, and technology.* Needham Height, MA: Allyn & Bacon.

Theoretical and Empirical Reviews of Cross-language and L2 Speech Perception and Production:

Bohn, O.-S., & Munro, M. J. (eds.) 2007. *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam/Philidelphia: John Benjamins.

Flege, J. E. 2003. Assessing constraints on second-language segmental production and perception. In N. O. Schiller and A. S., Meyer (eds.), *Phonetics and phonology in language comprehension and production*. Berlin-New York: Mouten de Gruyter.

Piske, T., MacKay, I., & Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215.

Strange, W. (ed.). 1995. *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press.

Neurobiological Measures of Auditory Perception:

Friedman, D., Cycowicz, Y. M., & Gaeta, H. 2001. The Novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Behavioral Reviews* 25: 355–373.

Donchin, E., & Coles, M. 1988. Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences* 11: 357–374.

Naatanen, R. 1999. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 38: 1–21.

References

Aaltonen, O., Niemi, P., Nyrke, T., & Tuhkanen, J. M. 1987. Event related brain potentials and the perception of a phonetic continuum. *Biological Psychology* 24: 197–207.

Aaltonen, O., Eerola, O., Hellstrom, A., Uusipaikka, E., and Lang, A. H. 1997. Perceptual magnet effect in the light of behavioral and psychophysiological data. *Journal of the Acoustical Society of America* 101: 1090–1105.

Beddor P. S., & Gottfried, T. 1995. Methodological issues in cross-language speech perception research with adults. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, Timonium, MD: York Press, 207–232.

Best, C. T. 1995. A direct realist view of cross-language speech perception. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, Timonium, MD: York Press, 171–204.

Best, C. T., & Strange, W. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20: 305–330.

Best, C. T., Hallé, P. A., Bohn, O.-S., & Faber, A. 2003. Cross-language perception of nonnative vowels: Phonological and phonetic effects of listeners native language. In M. J. Solé, D. Recasens, & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences,* Barcelona, Causal Productions, 2889–2892.

Best, C. T., McRoberts, G. W., & Goodell, E. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listeners' native phonological system. *Journal of the Acoustical Society of America* 109: 775–793.

Best, C. T., McRoberts, G. W., & Sithole, N. M. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English–speaking adults and infants. *Journal of Experiment Psychology: Human Perception and Performance* 14: 345–360.

Best, C. T., & Tyler, M. D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*, Amsterdam/ Philidelphia: John Benjamins, 13–34.

Binder, J., & Price, C. J. 2001. Functional neuroimaging of language. In R. Cabeza & A. Kingstone (eds.), *Handbook of functional neuroimaging of cognition*, Cambridge, MA: MIT Press, 187–250.

Bohn, O.-S. 1995. Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, Timonium, MD: York Press, 273–304.

Bradlow, A. R. 2008. Training non-native language sound patterns: Lessons from training Japanese adults on the English /ɹ/-/l/ contrast. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Bradlow, A. R., & Alexander, J. S. 2007. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America* 121: 2339–2349.

Bradlow, A. R., & Bent, T. 2002. The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America* 112: 272–284.

Datta, H., Hisagi, M., Wroblewski, M., Garrido, K., Vidal, N., & Shafer, V. L. 2006. Vowel perception in bilingual adults & children: Electrophysiological measures. Poster presentation to Language Acquisition and Bilingualism: Consequences for a Multilingual Society, Toronto, Canada.

Dehaene-Lambertz, G. 1997. Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport* 8(4): 919–924.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. 2000. Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience* 12: 635–647.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. 1971. Speech perception in early infancy. *Science* 171: 304–306.

Flege, J. E. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–65.

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, Timonium, MD: York Press, 233–277.

Flege, J. E., & Hillenbrand, J. 1984. Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America* 76: 708–720.

Flege, J. E., & Wang, V. 1989. Native phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. *Journal of Phonetics* 17: 299–315.

Flege, J. E., Munro, M. J., & Fox, R. A. 1994. Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America* 95: 3623–3641.

Flege, J. E., Schirru, C., & MacKay, I. R. A. 2003. Interaction between the native and second language phonetic systems. *Speech Communication* 40: 467–491.

Garrido, K., Hisagi M., & Shafer, V. L. 2005. ERP Indices of Vowel Processing in Spanish–English bilinguals. Poster presentation to 12th Annual Meeting of the Cognitive Neuroscience Society, New York.

Gottfried, T. L. 1984. Effects of consonantal context on the perception of French vowels. *Journal of Phonetics* 12: 91–114.

Harnsberger, J. D. 2000. A cross-language study of the identification of non-native nasal consonants varying in place of articulation. *Journal of the Acoustical Society of America* 108: 764–783.

Harnsberger, J. D. 2001. On the relationship between identification and discrimination of non-native nasal consonants. *Journal of the Acoustical Society of America* 110: 489–503.

Hisagi, M. 2007. Perception of Japanese temporally-cued phonetic contrasts by Japanese and American English listeners: Behavioral and Electrophysiological Measures. Unpublished doctoral Dissertation, City University of New York Graduate School and University Center.

Ioup, G. 2008. Exploring the role of age in the acquisition of a second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Jacobsen, T., Horvath, J., Schroger, E., Lattner, S., Widmann, A., & Winkler, I. 2004. Pre-attentive auditory processing of lexicality. *Brain and Language* 88: 54–67.

Jenkins, J. J. 1979. Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik (eds.), *Levels of processing in human memory,* Hillsdale, N.J. Erlbaum Associates, 429–446.

Jusczyk, P. W. 1997. *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kewley-Port, D., Bohn, O.-S., & Nishi, K. 2005. The influence of different native language systems on vowel discrimination and identification. *Journal of the Acoustical Society of America* 117: 2339.

Kuhl, P. K., & Iverson, P. 1995. Linguistic experience and the "perceptual magnet effect." In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, Timonium, MD: York Press, 121–154.

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

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. 1975. VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology* 20: 215–225.

Levy, E. 2004. Effects of language experience and consonantal context on perception of French front rounded vowels by adult American English learners of French. Unpublished doctoral dissertation. City University of New York-Graduate School and University Center.

Levy, E., & Strange, W. in press. Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychological Review* 74: 431–461.

Liberman, A. M., & Mattingly, I. G. 1985. The motor theory of speech perception revised. *Cognition* 21: 1–36.

Maiste, A. C., Wiens, A. S., Hunt, M. J., Scherg, M., & Picton, T. W. 1995. Event related potentials and the development of speech sounds. *Ear and Hearing* 16: 68–90.

Mayo, L. H., Florentine, M., & Buus, S. 1997. Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research* 40: 686–693.

MacKain, K. S., Best, C. T., & Strange, W. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics* 2: 369–390.

Miyawaki, K., Strange, W., Verbrugge, R. R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. 1975. An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception & Psychophysics* 18: 331–340.

Morr, M. L. 2002. Mismatch negativity elicited to vowels in children and adults and its association to behavioral performance. Unpublished doctoral dissertation. City University of New York, Graduate School and University Center.

Mochizuki, M. 1981. The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics* 9: 283–303.

Naatanen, R. 1990. The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive functions. *The Behavioral and Brain Sciences* 13: 201–288.

Naatanen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Livonen, A., Vainio, M., Alku, P., Limoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385: 432–434.

Nittrouer, S., & Miller, M. E. 1997a. Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America* 101: 2253–2266.

Nittrouer, S., & Miller, M. E. 1997b. Developmental weighting shifts for noise components of fricative-vowel syllables. *Journal of the Acoustical Society of America* 102: 572–580.

Pikser, R. 2003. Perception of word-final /t/, /d/, /s/, and /z/ by adult native Spanish-speaking learners of English. Unpublished Master's thesis. City University of New York, Graduate School and University Center.

Polka, L. 1991. Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America* 89: 2961–2977.

Polka, L. 1992. Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics* 52: 37–52.

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., & Bohn, O.-S. 1996. A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America* 100: 577–592.

Polka, L., & Strange, W. 1985. Perceptual equivalence of acoustic cues that differentiate /r/ and /l/. *Journal of the Acoustical Society of America* 78: 1187–1197.

Polka, L., & Werker, J. F. 1994. Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 20: 421–435.

Sams, M., Aulanko, R., Aoltonen, O., & Naatanen, R. 1990. Event related potentials to infrequent changes in synthesized phonetic stimuli. *Journal of Cognitive Neuroscience* 2: 344–357.

Shafer, V. L., & Garrido, K. 2007. The neurodevelopmental basis of speech and language. In M. Shatz & E. Hoff (eds.), *The handbook of language development*, Oxford: Blackwell, 21–45.

Shafer, V. L., Morr, M. L., Datta, H., Kurtzberg, D., & Schwartz, R. G. 2005. Neurophysiological indices of speech processing deficits in children with specific language impairment. *Journal of Cognitive Neurosciences* 1168–1180.

Shafer, V. L., Schwartz, R. G., & Kurtzberg, D. 2004. Language-specific memory traces of consonants in the brain. *Cognitive Brain Research* 18: 242–254.

Sharma, A., & Dorman, M. F. 1998. Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential. *Journal of the Acoustical Society of America* 104: 511–517.

Sharma, A., & Dorman, M. F. 1999. Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America* 106: 1078–1083.

Strange, W. 1992. Learning non-native phoneme contrasts: Interactions among subject, stimulus, and task variables. In Y. Tohkura, Y. Sagisaka, & E. Vatikiotis-Bateson (eds.), *Speech perception, production and linguistic structure*, Tokyo, JAPAN: OHM, 197–219.

Strange, W. (ed.) 1995. *Speech perception and linguistic experience: Issues in cross-language speech research.* Timonium, MD: York Press.

Strange, W. 2002. Speech perception and language learning: Wode's developmental model of speech perception revisited. In P. Burmeister, T. Piske, & A. Rohde (eds.), *An integrated view of language development: Papers in honor of Henning Wode*, Trier: WVT: Wissenschaftlicher Verlag Trier.

Strange, W. 2006. Second-language speech perception: The modification of automatic selective perceptual routines. *Journal of the Acoustical Society of America* 120: 3137.

Strange, W. 2007. Cross-language phonetic similarity of vowels. In O.-S. Bohn & M.J. Munro (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*, Amsterdam/Philidelphia: John Benjamins, 35–55.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., & Jenkins, J. J. 1998. Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics* 26: 311–344.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., & Nishi, K. 2001. Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *Journal of the Acoustical Society of America* 109: 1691–1704.

Strange, W., Bohn, O.-S., Trent, S. A., & Nishi, K. 2004. Acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America* 115: 1791–1807.

Strange, W., Bohn, O.-S., Nishi, K., & Trent, S. A. 2005. Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America* 118: 1751–1762.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics* 36: 131–145.

Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., & Nishi, K. 2007. Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *Journal of the Acoustical Society of America* 122: 1111–1129.

Streeter, L. A. 1976. Language perception of two-month old infants shows effects of both innate mechanisms and experience. *Nature* 259: 39–41.

Sussman, E., Ritter, W., & Vaughan, H. 1999. An investigation of the auditory streaming effect using event-related potentials. *Psychophysiology* 36; 222–34.

Szymanski, M. D., Yund, E. W., & Woods, D. L. 1999. Phonemes, intensity and attention: Differential effects on the mismatch negativity (MMN). *Journal of the Acoustical Society of America* 106: 3492–3505.

Werker, J. F., & Curtin, S. 2005. PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development* 1: 197–234.

Werker, J. F., & Logan, J. 1985. Cross-language evidence for three factors in speech perception. *Perception & Psychophysics* 37: 35–44.

Werker, J. F., & Tees, R. C. 1983. Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology* 37: 278–286.

Werker, J. F., & Tees, R. C. 1984a. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7: 49–63.

Werker, J. F., & Tees, R. C. 1984b. Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America* 75: 1866–1878.

Werker, J. F., & Tees, R. C. 1999. Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology* 50: 509–535.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. 1981. Developmental aspects of cross-language speech perception. *Child Development* 52: 349–353.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csepe, V., Aaltonen, O., Raim, I., Alho, K., Lang, H., Iivonen, A., & Naatanen, R. 1999a. Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research* 7: 357–369.

Winkler, I., Kujala, T., Titinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csepe, V., Ilmoniemi, R. J., & Naatanen, R. 1999b. Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36: 638–642.

Winkler I., Kujala, T., Alku, P., & Naatanen, R. 2003. Language context and phonetic change detection. *Cognitive Brain Research* 17: 833–844.

Yamada, R. A. 1995. Age and acquisition of second language speech sounds. Perception of American English /ɹ/ and /l/ by native speakers of Japanese. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, Timonium, MD: York Press, 305–320.

CHAPTER 7

Foreign accent and speech intelligibility

Murray J. Munro
Simon Fraser University

Introduction

The occurrence of foreign accents provides some of the clearest evidence that knowledge of a first language (L1) influences the acquisition of a second (L2). Nonnative speakers of English are often readily recognized because of their pronunciation, and in many cases their specific L1 backgrounds can be identified, even by casual interlocutors. Ellis (1994) commented that the phenomenon of accented speech "is so well attested that it hardly requires documenting" (p. 316). Nevertheless, as Pennington (1996) observed, there is no widely-accepted definition of *foreign accent*, and the growth in importance of nativized varieties of English will likely continue to blur the distinction between what is called a "native" accent and what is considered "foreign." Despite this difficulty, foreign-accented speech has attracted the attention of teachers, researchers, and clinicians for a very long time, though conceptions of its significance have varied considerably, especially over the past century. Greene and Wells (1927),[1] for instance, wrote that

> Foreign accent, being of the nature of imperfect or defective speech, is the result of incorrect articulation and enunciation and is therefore classified, from our therapeutic viewpoint, as stammering speech. (p. 24)

Several decades later, in an article directed at teachers, Griffen (1980) presented a less disparaging, but still disapproving, view of foreign accents as inherently undesirable and in need of repair:

> The goal of instruction in pronunciation is that the student (or patient) should learn to speak the language as naturally as possible, free of any indication that the speaker is not a clinically normal native. (p. 85)

1. I thank Anna Marie Schmidt for drawing this material to my attention.

However, current views of foreign-accented speech hold that native pronunciation in the L2 is not only uncommon but unnecessary. This understanding has been reached in part because of research on L2 speech learning, a field of study that has been motivated by a wide range of concerns, including those having to do with central issues in language acquisition and those relating to the effects of accent on communication. The first of these areas encompasses such diverse topics as the putative critical period for language acquisition (Oyama 1982; Scovel 1988; see also Chapter 2 by Ioup in this volume), the underlying perceptual basis of accented speech (Flege 1995), and universal factors and learnability in phonological acquisition (see Chapters 3 and 4 of this volume by Major and Eckman, respectively, as well as Eckman 1977; Major 1987, 2001).

One of the most commonly cited observations in applied linguistics research is the close relationship between age of L2 learning and foreign accentedness. Flege, Munro and MacKay (1995), for instance, identified a strong positive correlation between these variables in a group of 240 Italian immigrant adults living in Canada. In general, later learning was associated with stronger accents. Of particular interest was the finding that the overwhelming majority of participants who had begun to learn English after early childhood – "late" L2 learners – could be identified as nonnative as a result of their pronunciation. Although other work has indicated that at least some late learners may speak with native or native-like accents (Bongaerts 1999), these cases are the exception rather than the rule. Some of them appear to result from special talent in second language acquisition (Ioup, Boustagui, El Tigi, & Moselle 1994), while others have been associated with high motivation for learning L2 pronunciation (Moyer 2004).[2] The possibility that talent or motivation may help overcome, even partly, age-related limitations on phonetic learning is intriguing (and conversely, as Hansen Edwards notes in Chapter 9, that social factors such as identity can affect L1 accent retention). However, whether or not one accepts the existence of a critical period for speech learning, the available evidence leads to the inescapable conclusion that having a foreign accent is a common, normal aspect of late second language acquisition. As such, it is not a disorder, and the fact that millions of second language users around the world communicate successfully using foreign-accented speech indicates that accent-free pronunciation is *not* a necessary goal for either learners or teachers of second languages (cf. Hansen Edwards, Chapter 9, in this volume).

This chapter addresses the second area of concern identified above – the ways in which accented speech is received by those who hear and interact with L2 speakers, primarily learners of English as a second language. Research within this area

2. Chapter 3 by Ioup in this volume provides a more detailed discussion of age and foreign language accent.

of inquiry has already led to some valuable insights into the role of accent in communication. However, many critical issues have yet to be explored in detail. The sections that follow will outline some of the established facts about the perception of L2 speech and identify a number of problems that deserve further attention.

Foreign accent and its consequences

Flege (1988) noted that speaking with a foreign accent entails a variety of possible consequences for L2 users, including accent detection, diminished acceptability, diminished intelligibility, and negative evaluation. To this list we might add the potential benefit that an accent may serve as a marker of non-native competence, such that interlocutors adjust the speech input they provide to L2 users (Gass & Varonis 1984). An accent may therefore trigger foreigner talk from native speakers (Varonis & Gass 1982), thus enhancing communication. This effect and other responses to L2 speech are possible because accents are highly salient to both native and second language speakers. Even when they have no phonetic training, listeners can often recognize when someone comes from outside their own speech community on the basis of very little speech material. This sensitivity has been explored in a number of accent detection studies designed to establish the basis on which listeners identify a speaker as nonnative.

The bases of accent detection

Flege (1984) presented utterances of varying durations to listeners who indicated whether or not the speaker's L1 was English. In general, the listeners performed the task successfully, regardless of whether they heard phrases, words, single phones, or even parts of a phone. The accuracy of the listeners was evidently the result of noticing characteristics of L2 production that differed from the patterns that the listeners themselves might use. Thus, it is possible that they did not actually recognize the speech as non-native *per se*, but that they were able to say which speakers spoke a different variety of English from their own by attending to speech characteristics at a variety of levels.

At the segmental level, accented speech is signaled by the omission or insertion of phones, the substitution of one phone for another, or the production of phones that differ at the subphonemic level from native-like segments (see Zampini, Chapter 8 of this volume, as well as Chapter 2 by Ioup, for discussion of several studies that illustrate these phenomena). Any of these might be used by listeners in order to determine the nativeness of a speech sample. In fact, several studies have shown that when listeners rate the degree of foreign-accentedness of utterances, their scores correlate with the numbers of insertions, deletions, and substitutions that have been identified by phonetically-trained assessors (Anderson-Hsieh, Johnson, & Koehler 1992; Brennan & Brennan 1981; Munro & Derwing 1995a). Thus, when assigning judgments listeners seem to take into account how frequently such phenomena occur.

However, other work reveals that non-segmental phenomena also contribute to foreign accent detection. Van Els and DeBot (1987) used digital signal processing techniques to remove pitch variation in native Dutch and Dutch L2 utterances so that they were heard as monotone. This manipulation reduced listeners' success in identifying the speech as native or foreign-accented and therefore demonstrated that the intonation component of Dutch speech can convey accentedness. In a related study, Munro (1995) presented low-pass filtered English utterances that sounded like a low-pitched murmur to native English listeners and found that they could judge which ones exhibited a Mandarin accent. This unintelligible speech preserved some of the prosodic properties of the original utterances but contained little or no useful segmental information. The listeners may have succeeded at the task by noting nonnative rhythmic or intonation patterns. In fact, other research indicates that L2 intonation and rhythm are indeed influenced by properties of the L1 sound system (Grover, Jamieson, & Dobrovolsky 1987; Shah 2003; Tajima, Port, & Dalby 1997).

More recently Munro, Derwing, and Burgess (2003, forthcoming) made digital recordings of native and non-native speech and presented them backwards in an accent detection task. Once again the listeners showed high levels of accuracy that held for different kinds of accents, including Mandarin, Cantonese and Czech; for utterances of various durations, ranging from a whole sentence to a single disyllabic word; and for stimuli that had undergone severe temporal disruption. The stimuli used in these studies contained no useable segmental information and sometimes were not prosodically intact. It is possible that the listeners made correct judgments by paying attention to long-term characteristics of speech, such as articulatory settings that had been transferred from the L1 to the L2. Although the potential role of speech settings such as breathiness, creakiness, dentalization, and retroflexion in foreign accented speech has been discussed in the pedagogical literature (Esling 1994; Esling & Wong 1983), this issue has, as yet, received only minimal attention in L2 speech research. However, recent ultrasound and optotrack analyses by Wilson (2006) point to cross-linguistic differences in interspeech postures as a promising new line of investigation in this area. (See also Gick, Bernhardt, Bacsfalvi, & Wilson, Chapter 11 of this volume.)

The accent-intelligibility distinction

While the salience of foreign accents is well established, less is known about how an accent might affect communication during social interactions. Of course, it is widely recognized that L2 users at times have difficulty making themselves understood, sometimes because of pronunciation errors that make their speech unintelligible. Applied linguists also recognize, however, that perfect formal correctness in linguistic output is not a necessary condition for communicative competence. One reason is that interlocutors are often able to understand L2 utterances containing grammatical or pronunciation errors by invoking top-down or other processes. Nonetheless, a detailed understanding of the situations in which pronunciation errors lead to communication breakdowns has yet to be developed.

Researchers have not often distinguished between those aspects of L2 speech that cause it to sound foreign and those that reduce its intelligibility (see Giles & Powesland 1975; Van Els & de Bot 1987). In fact, as recently as 1988, Anderson-Hsieh and Koehler wrote that "Only a few studies have been reported in the literature on the comprehensibility of nonnative speech..." (p. 563). Even in studies of pronunciation error gravity and of the effectiveness of pronunciation instruction, the *nativeness* of L2 speakers' pronunciation has often been selected as the dependent variable rather than speech intelligibility (Anderson-Hsieh, Johnson, & Koehler 1992; Macdonald, Yule, & Powers 1994). Yet, the accent-intelligibility distinction is a fundamental one that holds considerable importance in several domains, including language teaching, language testing, and even human rights litigation. There are a number of reasons for studying this relationship.

Language teaching

With the rise and fall in popularity of various language teaching methods over the past half century, interest in the phonological aspects of ESL has waxed and waned. Current trends suggest a renewed interest in the teaching of pronunciation, at least among researchers and teacher educators (Chela-Flores 2001; Derwing & Munro 2005; Derwing & Rossiter 2002), and in the applications of research findings to the classroom (see Chapter 12 by Chun et al. and Chapter 13 by Derwing, this volume, for extensive discussions of L2 pronunciation pedagogy). Several kinds of pedagogical concerns might be addressed through research on speech intelligibility. For example, the continued emphasis on communicative competence in language instruction leads teachers to focus on aspects of pronunciation that have a demonstrable effect on communicative success. They would like to have specific advice on what aspects of English pronunciation are most important for communication. Also, because teachers and learners do not have unlimited time for instruction in pronunciation (any more than in any other skill), it is important to establish a set of priorities for teaching. If one aspect of pronunciation instruction is more likely to promote intelligibility than some other aspect, it deserves more immediate attention. Finally, it is important to know what aspects of intelligibility are teachable and what ways of teaching are most likely to be successful. Research examining speech intelligibility can be valuable to pedagogy, then, when it helps identify critical pronunciation problems that are actually experienced by L2 learners and leads to successful ways of addressing those problems.

Language testing

Several standardized tests, such as the Test of Spoken English (TSE®) and the International English Language Testing System (IELTS®) focus on or include assessment of L2 speaking skills. Because they aim at evaluating effectiveness in oral communication, such tests must use appropriate evaluation instruments to obtain assessments from adequately trained evaluators who judge speech samples reliably. While it seems obvious that evaluators must recognize the difference between speech that is unintelligible and speech that is accented but still intelligible, this distinction has not always been clearly observed. Anderson-Hsieh, Johnson, and Koehler (1992), for instance, describe the pronunciation criteria for the Speaking Proficiency English Assessment Kit (SPEAK® test), which combines intelligibility and acceptability into a single 4-point scale ranging from "heavily accented *and* unintelligible" at one end to "near-native" at the other. This is a serious confusion of two partially independent dimensions of L2 speech.

Another problem in language testing is that evaluators of L2 speech, being very familiar with it, may not perceive it in the same way as interlocutors outside the classroom. On the one hand, they may approach it more analytically than linguistically less-sophisticated listeners, and may be able to identify problem areas using appropriate metalinguistic terminology. This may explain why several studies have shown that phonetic training and experience with L2 speech seem to correlate with high levels of inter-judge reliability in L2 speech assessment (Brennan & Brennan 1981; Thompson 1991). On the other hand, from a more holistic standpoint, teachers and testers may understand utterances that listeners with less familiarity do not. Yet they may respond to L2 speech more critically in general (Schairer 1992), perhaps because of a heightened awareness of the pronunciation difficulties that learners experience. Thus, in some respects, sophisticated evaluators may not be an ideal audience to render an opinion on the intelligibility of a particular L2 speaker. It is not clear to what extent they may be able to estimate the difficulties a third party might have in understanding L2 speech (see Schairer 1992). While additional research on the role of experience and expertise on L2 speech perception is clearly needed, this concern could be mitigated considerably if evaluators could make use of research findings about speech intelligibility. After analyzing L2 output, their identifications of problem areas and pronunciation errors needing correction could be based on research findings about which errors have the largest impact on communication, rather than on their personal impressions about the intelligibility of a particular speaker.

Human rights litigation

High levels of immigration in Inner Circle countries such as Canada and Australia have resulted in rich linguistic diversity, along with a growing awareness of the problem of language-based discrimination (see e.g., Lippi-Green 1997). Munro (2003) reviews several cases of accent discrimination that have been brought before human rights tribunals, including incidents in which L2 speakers were harassed, denied employment or terminated. In many of these cases, the central issue has been precisely the accent-intelligibility distinction being considered here. Under human rights legislation in many countries, language proficiency is seen as a bonafide occupational requirement in some circumstances. For instance, it is accepted that employees who work with the public, such as telephone receptionists, teachers, and nurses should be able to communicate effectively in the language of clients, students, and patients, and it is reasonable for employers to use language proficiency as a criterion for hiring in such cases. In other words, it is justifiable to expect intelligibility. However, it is not reasonable to require that employees speak without a foreign accent. In evaluating language-related human rights complaints and civil lawsuits, tribunals must often use the criterion of intelligibility in judging whether a particular action (e.g., not hiring an applicant) is justified. For instance in the case of Mirek Gajecki v. Board of Trustees, School District No. 36 (Surrey) (1990), it was determined that a Canadian school teacher was denied employment simply because of his accent and not because he had any difficulty communicating with his students. The complainant was awarded compensation by the tribunal. In a contrasting complaint, however (Jacques Clau v. Uniglobe Pacific Travel 1995), a job applicant at a travel agency was found by a tribunal to have serious difficulty communicating over the telephone because of his accent, and his case was dismissed. In fact, a large proportion of language-related human rights cases center around the accent-intelligibility distinction, and must often be resolved by eliciting testimony from linguistically unsophisticated witnesses. Whether or not the decisions rendered in these two cases were correct, research on the accent-intelligibility distinction might have made possible expert testimony that could have clarified key issues for complainants, respondents, and adjudicators.

The evaluation of L2 speech: Approaches, problems, and findings

Dimensions for assessment

The empirical study of L2 speech entails assessments on various dimensions. Despite some inconsistencies in usage, terms such as *acceptability*, *comprehensibility*, *intelligibility*, and *fluency* are frequently discussed in this area of research. Three general approaches to speech assessment are of particular importance here: (1) responses from unsophisticated listeners, (2) impressionistic analyses from expert evaluators, and (3) acoustic phonetic analyses. The first of these often entails holistic ratings of L2 speech samples and does not require much specific metalinguistic knowledge on the part of the raters. Listeners who are not phonetically trained may be asked to rate speech according to how accented, comprehensible, or fluent it is. They may also be asked to identify what words, phrases, or sentences have been produced. The second means of assessment often refers to phoneticians' counts or ratings of specific phonetic phenomena, such as segmental errors, prosodic accuracy, or voice quality. The third approach entails computer measurements of quantifiable aspects of speech, including voice onset time, formant frequencies, pitch, and duration.

To establish the effects of accent on communication, unsophisticated listeners' judgments are especially important because they may provide insight into how understandable L2 speakers are when they interact with other members of their community. Focusing on L2 pedagogical concerns, Munro and Derwing (1995a, 1995b) and Derwing and Munro (1997) examined two kinds of perceptual judgments from untrained listeners, with *accentedness* referring to their perceptions of strength of accent in an utterance and *comprehensibility* referring to their estimation of difficulty in understanding the utterance. In both cases, they collected ratings from native English listeners on 9-point scales. This understanding of accentedness and comprehensibility – as the experiences of a listener – rules out using expert judgments or acoustic phonetic analyses on their own as a sufficient means of evaluating L2 speech. In the first place, as already observed, phonetically-trained evaluators do not necessarily respond to L2 speech in the same way as unsophisticated listeners. In the second, instrumental measurements might, in principle, reveal differences between native and non-native speech that are not noticed by listeners and that therefore do not result in an accent. From the standpoint of communication, there is no useful way to assess accentedness and comprehensibility, except through listener responses of some sort, and therefore, there is no reason to use expressions like "*perceived* accentedness" or "*perceived* comprehensibility" because, in fact, there is no other kind of accentedness or comprehensibility.

Numerous L2 studies have used listeners' judgments of accentedness, and a few have examined comprehensibility. In terms of communicative competence, however, *intelligibility* is usually seen as *the* critical concern in L2 speech production. Subtelny (1977) identified it as the single most important index of oral communicative competence and Pennington (1996) saw it as the most pressing goal in pronunciation instruction. With respect to L2 speech, Munro and Derwing (1995a) proposed that it be defined as the degree to which a speaker's utterance is actually understood by a listener. Although some researchers have had listeners rate intelligibility in scalar fashion (e.g., Fayer & Krasinski 1987), rating data are of limited use in evaluating how much comprehension has actually taken place, because listeners sometimes mistakenly believe that they have understood an utterance, and may therefore rate it highly, when they have not understood it well at all. Munro (1998), for instance, found that listeners incorrectly thought that they had understood L2 utterances about 13% of the time. This finding suggests that some measure is required that compares the speaker's intended message with what the listener has understood. This raises an important issue in the elicitation of L2 utterances because it presupposes that the researcher knows or can determine what that message is.

Evaluating intelligibility

A wide array of techniques for assessing the intelligibility of normal and disordered native speech have been in use for many years. Kent, Miolo, and Bloedel (1994), for instance, describe 19 different procedures for pediatric speech assessment. The study of intelligibility in L2 speech, however, is still in its early stages, with only a small number of studies actually assessing intelligibility using a variety of different approaches. Lane (1963) assessed the intelligibility of individual foreign-accented words in quiet and noise by presenting them to listeners who indicated what they heard. Smith and Rafiqzad (1979) had listeners complete a cloze test based on a passage read by speakers from various L1 backgrounds. Smith and Bisazza (1982) used a standardized test in which L2 speakers read sentences or paragraphs aloud. Listeners were then required to select a picture corresponding to the read material and to provide multiple-choice responses to questions based on the reading. Perlmutter (1989) had listeners summarize the main idea of short presentations by L2 users. Brodkey (1972) proposed a now-common technique, the dictée task, in which listeners heard sentence-length samples and wrote them out in standard orthography. The data were then scored in terms of words correctly transcribed. To prevent ceiling effects in intelligibility scores, dictée tasks are sometimes carried out with speech embedded in noise as in Bent and Bradlow (2003), who scored transcriptions on the basis of keywords correctly transcribed. Anderson-Hsieh and Koehler (1988) had L2 speakers read passages aloud and then presented them to native listeners, who responded to comprehension questions. Finally, Munro and Derwing (1995b) used a verification task in which listeners heard short true and false sentences read aloud by native and nonnative speakers, and indicated their comprehension through true or false responses.

Though not intended to be exhaustive, the preceding summary indicates a fairly wide range of choices for L2 intelligibility assessment. Each of these approaches has its advantages and limitations, but none gives a complete picture of all aspects of speech intelligibility. The choice of a particular approach depends on the type of speech material that is available or that can be elicited, the kinds of demands that can be placed on listeners and speakers, and the specific research questions to be addressed.

One problem worthy of close examination here is that of obtaining a suitable speech sample for assessing intelligibility. In general, controlled production tasks in which speakers read a text or repeat a recorded model have the advantage of allowing the researcher to obtain particular speech material from the speaker. As noted earlier, if intelligibility is defined as the amount of a message that is actually understood, a comparison of the intended message with the received message is essential. However, the researcher can be certain of the content of the intended message only if that content is pre-determined, as in tasks in which the L2 speaker reads or repeats words, sentences, or longer texts. The drawback of such controlled tasks is that they may yield language that lacks naturalness because it was not actually formulated by the speaker. Read material, for instance, may include mispronunciations due to lack of word familiarity or because of orthographic interference, dysfluencies, and unnatural prosody that is noticeably different from that found in spontaneous speech. The resulting intelligibility assessment may therefore underestimate the speaker's actual capabilities. Asking a speaker to simply repeat utterances after a model raises problems as well, because the speaker may produce better-than-normal output by closely imitating the model voice.

As an alternative to controlled production tasks, it is sometimes preferable to elicit extemporaneous speech through picture story tasks, personal narratives, or interactive tasks. While the evaluation of extemporaneous speech eliminates some of the problems identified above, it introduces new difficulties. For example, the output may well contain grammatical errors that could inappropriately influence pronunciation ratings, and the speakers may not produce particular words or sounds that the investigator wishes to examine. Furthermore, if two groups of speakers are to be compared directly, it may be necessary to present evaluators with identical speech material from each, a situation that cannot be achieved with extemporaneous utterances.

One further approach that has been used in some studies is a "delayed repetition task," in which speakers provide a repetition of modeled speech, but only after they hear some intervening speech material (Flege, Munro, & MacKay 1995). This approach is believed to reduce short-term recall of the model voice that might otherwise allow close imitation.

Because of the various drawbacks of all approaches to speech elicitation, it seems inadvisable to rely on any single one as a basis for drawing firm conclusions about particular speakers or accents, or about L2 speech in general. The best that can be hoped for is that multiple approaches in research will lead researchers to converging conclusions. Even so, as noted by Fayer and Krasinski (1987), intelligibility is "one aspect of the total communicative effect of a nonnative message" (p. 313). Other non-linguistic aspects include various responses of the listener such as irritation, distraction or the listener's perception of a negative or positive relationship with the speaker. As Derwing notes in Chapter 13, this volume, loudness of speech, voice quality, and clarity can also affect judgments. Such aspects may be very difficult or even impossible to assess, but they are clearly relevant to communication. The study of accent in actual human interactions is further complicated by the fact that prejudice and discrimination can play a role. Lippi-Green (1997) reviews a number of studies showing that listeners sometimes devalue those who speak with an accent that is different from their own. Thus, one cannot assume that interlocutors always perform to their full potential when it comes to understanding what others say. Not only may they sometimes choose not to comprehend or consciously decide to feign a lack of comprehension, but they may sometimes be influenced unconsciously by their expectations, even when there is no willful resistance to understanding. Rubin (1992), for example, found that American undergraduate students actually understood less of a lecture when they were led to believe that the speaker was Chinese as opposed to Caucasian, even though the voice they heard was identical under both conditions. Given these findings, it would obviously be impossible in any single study to assess the contributions of all the factors that might affect a particular listener's comprehension of some utterance. However, this does not preclude the possibility of obtaining meaningful findings in intelligibility-oriented research, especially when the outcomes of diverse studies are taken into account.

Interrelationships among assessment dimensions

In a series of perceptual investigations, Munro and Derwing (1995a, b, 2001) and Derwing and Munro (1997) have examined the interrelationships among accentedness, comprehensibility and intelligibility using various types of speech material, tasks, and groups of listeners. The findings of these studies strongly suggest that while accented speech is a normal phenomenon, unintelligible L2 speech is much less common. In particular, listeners often judge an utterance as heavily accented, even when they transcribe it perfectly and when they do *not* rate it as difficult to understand. This finding is illustrated in Figure 1, which compares distributions of accentedness and comprehensibility ratings observed by Derwing and Munro (1997). In that study, a single group of listeners rated a single set of utterances on these two dimensions. The two resulting distributions are somewhat skewed in opposite directions. Although many utterances received strong accent ratings, considerably fewer were actually rated as hard to understand.

This difference in distributions, which has been repeatedly observed in L2 speech research, indicates that listeners have an underlying awareness of the distinction between speech that is merely "different" from native speaker output and speech that is difficult to understand. It is important to note that this dif-

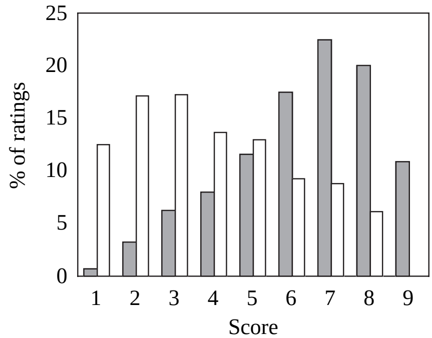


Figure 1. The distributions of accentedness (shaded) and comprehensibility (unshaded) scores assigned by a single group of listeners to one set of utterances. (1 = no accent or easy to understand; 9 = heavily accented or very difficult to understand)

ference is not simply due to a tendency for raters to use the accentedness and comprehensibility scales differently. In fact, a growing body of psycholinguistic data appears to support the view that accentedness and comprehensibility are partially independent aspects of L2 oral output. In the sentence verification task conducted by Munro and Derwing (1995b), listeners heard simple native and non-native statements, which they verified by pressing buttons marked *true* and *false*. Their accuracy scores and response latencies were then measured. The foreign-accented utterances took significantly longer to evaluate than the native ones, a finding that suggests that listeners might allocate additional processing resources to understand accented sentences. A similar conclusion was reached by Schmid and Yeni-Komshian (1999), and Bürki-Cohen, Miller, and Eimas (2001) found evidence of differences in the way individual native and nonnative words were processed. Also, work by Weill (2003) has revealed a relationship between response times recorded in a repetition task and corresponding comprehensibility ratings.

However, in Munro and Derwing (1995b) the amount of processing time was *not* related to the strength of the foreign accent. Heavily-accented sentences took no longer to verify than moderately- or weakly-accented utterances. This finding underscores the observation that aspects of an accent can be strong without necessarily affecting the listener's comprehension. On the other hand, processing time *was* somewhat related to comprehensibility ratings. In particular, sentences that were rated as harder to understand tended to take longer for listeners to process. Thus, in that study, the listeners' impressions of difficulty in understanding L2 speech appeared to reflect actual processing difficulty (see also Weill 2003). A further relevant finding is that intelligibility scores tend to correlate somewhat more highly with comprehensibility scores than with accentedness scores (Derwing & Munro 1997; Munro & Derwing 1995a).

Other evidence about the relationship between accent and intelligibility comes from pedagogical research. For instance, Derwing, Munro, and Wiebe (1998) examined improvements in narrative productions of ESL speakers due to global pronunciation instruction. A key finding of that study was that an instructed group improved significantly in comprehensibility and fluency, but not in accentedness. This finding not only adds support to the conception of accent and comprehensibility as partially independent dimensions of L2 speech, but suggests that pronunciation teaching need not focus on so-called "accent reduction" in order to help learners make themselves understood.

Stimulus properties vs. listener factors

A central issue in the study of L2 speech is the extent to which listeners share a response to particular speakers and utterances, a question that relates to the construct validity of such notions as accentedness, comprehensibility, and intelligibility. In particular, it is important to establish whether listeners generally agree on whether a particular utterance is intelligible or not. Gass and Varonis (1984) pointed out that the degree to which an utterance is understood depends partly on properties of the stimulus utterance itself, including both grammatical and phonological properties, as well as on various listener characteristics, such as the amount of experience the listener has with accented speech. Although their use of terminology differs from that employed here, Varonis and Gass (1982) and Gass and Varonis (1984) proposed a model of intelligibility in which a number of stimulus variables and listener factors would be assigned different weightings in order to account for the overall intelligibility of a particular utterance regardless of the speaker's background. Their model can be revised and extended as follows to provide a useful way of conceptualizing accentedness, comprehensibility, and intelligibility:

$$\text{SCORE} = \text{SP} + \text{LF} + \text{CF} + \ldots + \text{error}$$

Where

SCORE refers to one of accentedness (A), comprehensibility (C), or intelligibility (I)

and

SP (Stimulus properties) = $\alpha\text{Seg} + \beta\text{Pros} + \gamma\text{Gram} + \delta\text{Flue} + \ldots$

LF (Listener factors) = $\varepsilon\text{FTop} + \zeta\text{FSpkr} + \nu\text{FAcc} + \ldots$

CF (Contextual factors) = θCtxt

In the reconceptualized model shown above, the SCORE for an utterance on any of the three dimensions depends partly on stimulus properties (the SP component) and on listener factors (the LF component). Here it is assumed that accentedness (A) scores range from low values meaning "not foreign accented" to high values meaning "heavily foreign accented"; that comprehensibility (C) ranges from "easy to understand" to "very difficult to understand"; and that intelligibility (I) ranges from good scores to poor scores. Therefore, low numerical values for A, C, and I would generally indicate more native-like speech, though it should be noted that native speakers are not necessarily fully comprehensible or intelligible. The SP component can be broken down into segmental, prosodic, and grammatical problems, as well as fluency difficulties, with the Greek-letter coefficients indicating how much a particular kind of problem affects the SCORE. For instance, a large number of segmental errors leads to a larger value for the Seg contributor. However, the actual contribution of Seg errors to the total score is weighted by the value of the α coefficient on Seg. Moreover, the values of the coefficients vary, depending on whether A, C, or I is being evaluated. If, for instance, segmental errors have a large effect on A, but a smaller effect on I, the α coefficient on Seg would be larger for accentedness than for intelligibility. The LF component includes a number of listener-specific factors having to do with novelty. The more novel (i.e., unfamiliar) a topic (FTop), speaker (FSpkr), or type of accent (FAcc) is, the larger are the values of the relevant variables. Other factors such as context and listener error also influence the A, C, and I scores.

This model is not to be viewed as a computational model, but merely as a way of understanding the various possible influences on the perception of an L2 utterance. Suppose that for comprehensibility, the total contribution of LF were very small, while that of SP were large. In such a case, we would expect strong agreement among listeners about which utterances were comprehensible and which were not, and about the degree of comprehensibility of any particular utterance. In other words, the comprehensibility of an utterance could be largely predicted by analyzing the utterance itself. The reverse situation – a large contribution of LF and near-zero contribution of SP – would mean that comprehensibility was strongly influenced by characteristics of the perceiver and that it could be expected to vary radically from one rater to another. If the latter were the true state of affairs, there would presumably be little point in teaching pronunciation to improve ESL speakers' comprehensibility because there would be no consensus about any particular speech sample. Thus, an improvement in prosody might lead to better comprehension on the part of some listeners but not others.

To date, the research examining the contribution of the LF component in the perception of L2 speech has been quite limited. In one of the few pertinent studies, Gass and Varonis (1984) found that listeners' familiarity with topic, accent, speaker, and L2 speech in general had a positive impact on intelligibility, with topic familiarity having the largest effect. However, their study led to no conclusions about the relative contributions of SP and LF factors in the comprehension of L2 speech; nor did it deal with the question of how accentedness, comprehensibility, and intelligibility might entail different weightings of the same list of contributors. One more recent study (Munro, Derwing, & Morton 2006) provided evidence that, when listeners from a diverse range of language backgrounds are considered, listener factors play a much smaller role overall than stimulus properties. Still other work (Munro & Derwing 2006) indicates that accent and comprehensibility ratings can be affected in different ways by the same segmental substitutions in L2 speech. Nevertheless, more work needs to be carried out on these issues, and empirical data from a number of studies will have to be taken into account. Here we can consider four kinds of results: those addressing inter-listener reliability, those indicating the amount of variance in listener data that is explained by SP phenomena, pedagogical studies examining the effect of teaching on the three dimensions; and particular listener effects that may be explained by the structure of the rating task.

Reliability

As in any research involving rating-scale data, a finding of good interrater reliability supports the construct validity of the dimensions being assessed. Moreover, the extent to which listeners agree on the accentedness or comprehensibility of an utterance can provide insight into the relative contributions of the SP and LF components. In general, the more agreement among listeners, the less "subjectivity" there must be in their judgments, and the more evident it is that the listeners share a response to particular stimulus properties. Most studies in which accentedness ratings have been collected reveal moderate to high inter-rater agreement (e.g., Anderson-Hsieh, Johnson, & Koehler 1992; Derwing & Munro 1997; Munro & Derwing 1995b, to name a few). However, because not all researchers report the same kinds of reliability statistics, it is difficult to make comparisons across studies. Brennan, Ryan, and Dawson (1975) reported high levels of agreement among linguistically untrained undergraduate raters who assessed the accentedness of Spanish-English bilinguals, while Thompson (1991) reported high Spearman-Brown values (> .95) for experienced raters who judged Russian ESL speakers using a 5-point scale. In the latter study, reliability varied somewhat for different kinds of speech samples, however, and raters who had little or no experience with foreign-accented speech tended to be considerably less reliable than experienced listeners. One of the few studies to report low reliability for accentedness ratings was Southwood and Flege (1999), though it is unclear why their results differed from those of other researchers.

Few studies have reported reliability scores for comprehensibility judgments and intelligibility assessments. However, the limited available data suggest that the reliability for comprehensibility ratings tends to be comparable to that observed for accentedness. Although their work did not involve L2 speech, Barefoot, Bochner, Johnson, and vom Eigen (1993) reported high reliability among professional ratings of the comprehensibility of deaf speakers' productions. Derwing and

Munro (1997) reported intraclass correlation coefficients of .95 and .94 respectively for accentedness and comprehensibility based on ratings of four different accents from 26 undergraduate listeners. Although Derwing, Munro, and Wiebe (1998) and Munro and Derwing (2001) did not report the same type of scores for their untrained listeners, inter-rater correlations were above .70, indicating moderately good agreement.

From the studies cited above, which entailed native listener judgments of foreign-accented speech, it seems clear that good inter-rater agreement, even from untrained listeners is possible, though not inevitable. At the very least, the evidence indicates reflects a sizeable "shared experience" of L2 speech among those who have been asked to rate it. However, considerably more work must be carried out to establish the extent to which listeners agree with one another on intelligibility. One of the many remaining problems in this area is that very little work has explored whether responses to L2 speech from native listeners tend to agree with those of non-native listeners. The latter issue will be discussed further below.

Variance explained

Several studies have used regression analysis to examine the effects of various stimulus properties in the perception of L2 speech. In general, such research uses listener ratings as a dependent variable, and expert analysis or acoustic measurements to determine one or more independent variables. Obviously, the more variance in listeners' ratings that is explained by the measured stimulus properties, the smaller the contribution of idiosyncratic listener factors must be. In one of the earliest such studies, Brennan and Brennan (1981) had linguists evaluate utterances from nine Mexican Americans by identifying the frequencies of various pronunciation errors, mainly segmental substitutions and deletions, in a reading passage. After ensuring that the linguists had provided reliable error counts, they computed an accentedness index for each speaker based on the assessments, and compared the index values to accentedness judgments made by a group of 80 high school students. The experts' accentedness index accounted for 69% of the variance in the students' ratings. Because only selected utterances were used in the analysis, this result must be interpreted with some caution. Nonetheless, it provides an important indication not only that expert evaluations of stimulus properties are related to untrained listeners' judgments, but that the former can account for a sizeable amount of the variability in those judgments. Anderson-Hsieh, Johnson, and Koehler's (1992) results from an examination of 60 ESL speakers support this finding. In that study, trained judges' assessments of segmental, prosodic, and syllable structure errors accounted for 89% of the variance in global pronunciation ratings from ESL teachers. Munro and Derwing (2001) considered ratings from 27 undergraduate listeners who judged 10 Mandarin speakers of English on both accent and comprehensibility scales. Their analysis accounted for 39% of the variance in accent judgments and 41% of the variance in comprehensibility ratings using only phonological error counts (from experts) and speaking rate (measured using digital techniques) as predictor variables. Given the previous findings that prosodic factors also correlate with listener judgments (Anderson-Hsieh, Johnson, & Koehler 1992; Derwing & Munro 1997; Munro & Derwing 1995a), it seems very likely that even more variance would have been explained had prosody been taken into account. Taken together, these studies provide quite convincing evidence that listeners' ratings of accentedness and comprehensibility can be predicted to a sizeable extent on the basis of SP phenomena. Although it seems clear that these ratings are also influenced by contributions of the LF component of the model, the research evidence provides no reason to believe that LF factors generally outweigh SP factors in importance. In fact, the reverse may well be true.

Evidence from pedagogical research

Pedagogical studies also have the potential to shed light on the SP-LF distinction. In particular, suppose that we collect recordings of L2 learners' utterances before and after we provide them with pronunciation instruction, as did Perlmutter (1989). We may then mix the "before" and "after" recordings with those of an untrained control group and present them in random order to listeners for evaluations of comprehensibility and intelligibility. Assuming that the rating task is "blind" and that the control group shows no effect, higher ratings for the after-training utterances must be due to improvement in the speech itself and hence to SP factors that influenced the responses of listeners. As yet, relatively few studies of this type have actually assessed intelligibility and comprehensibility, though preliminary results have been positive. Perlmutter (1989), for instance, reported greater intelligibility among international teaching assistants after instruction, though she did not use a control group. In addition, Derwing, Munro, and Wiebe (1997, 1998) obtained similar outcomes for intelligibility and comprehensibility.

Listener effects in rating tasks

One final concern in examining listeners' responses to L2 speech is the effect that the structure of the rating task itself has on the results. Flege and Fletcher (1992), for example, found some degree of instability in listeners' accentedness judgments, noting that a judgment of a particular utterance depended to some degree on comparisons with other utterances that were presented during the same listening session. They also found that familiarity with a particular utterance led to harsher accentedness ratings, a finding supported by Munro and Derwing (1994). At first the latter outcome may seem to conflict with previous evidence of positive effects of familiarity on intelligibility (Gass & Varonis 1984). However, it is important to note that in these more recent studies, it was accentedness, and not intelligibility that was assessed. Munro and Derwing (1994) proposed that familiarity with a speaker or a specific speech sample might actually heighten the listener's awareness of particular errors and might therefore lead to a harsher judgment on the accent scale. In general, a requirement for accentedness judgments is a focus on form, while for intelligibility assessment, the listener must focus on meaning. As a result, different effects of familiarity might well be expected for the two dimensions.

Some specific implications for L2 pedagogy

In a review of the implications of the work discussed in this chapter, it should come as no surprise that, at this relatively early stage in L2 speech research, few definitive statements can be made about intelligibility in pronunciation pedagogy. However, one finding that has emerged in several of these studies is that speech does not necessarily become harder to understand simply as a result of being "different." In fact, L2 speech can be very different from native speech, yet still intelligible and comprehensible. From this, it follows that L2 learners who have difficulty making themselves understood cannot necessarily make their speech more intelligible simply by making it "less different." Thus the concept of *accent reduction* as a way of improving communicative competence seems poorly motivated. Rather, pedagogically speaking, it is more important to focus on those L2 speech phenomena that interlocutors commonly find hard to understand. As discussed earlier, further evidence suggests that comprehensibility and intelligibility are not entirely "in the ear of the beholder" and that there is considerable shared ground among those who respond to L2 speech. Most importantly, research indicates that pronunciation instruction that focuses on certain types of problems can lead to improved intelligibility in L2 speech.

One belief about English pronunciation that has been widely accepted for some time is that prosodic errors can be especially problematic in L2 speech output and that a focus on such concerns in the classroom is likely to have benefits for learners (see also Derwing, Chapter 13 of this volume). Of course, such a claim can be true only if the learners actually have difficulty with L2 prosody to begin with. In fact, several studies have supported the importance of prosody in L2 output. Among others, Anderson-Hsieh, Johnson, and Koehler (1992) found that prosodic errors contributed significantly to foreign accentedness ratings, while Munro and Derwing (1995a) and Derwing and Munro (1997) found that prosodic errors detracted significantly from both comprehensibility and intelligibility. Derwing, Munro, and Wiebe (1998) also concluded that instruction focusing on global aspects of pronunciation, including general speech habits and prosody, had greater overall benefits for learners' intelligibility than did instruction focusing exclusively on segmentals. Because prosody encompasses a wide range of speech phenomena, further research is needed to pinpoint those aspects of prosody that are most critical. One intriguing finding of Tajima, Port, and Dalby (1997) was that correction of the rhythmic properties of foreign-accented English using computer techniques led to improved intelligibility (see Bradlow, Chapter 10, and Chun, Hardison & Pennington, Chapter 12 of this volume for detailed discussions of issues related to technology and pronunciation training).

Most discussion of issues in pronunciation instruction focuses on what learners can do to improve their oral production. Consequently, there is a tendency to assume that there is nothing that can be done about the LF component in the intelligibility model described above – that is, the component that captures the experience or background of the listener. However, there is no reason to suppose that the L2 speaker must always carry full responsibility for the listener's comprehension in an interaction. Rather, it is worthwhile to consider ways of teaching people how to listen to and understand L2 speech, while maintaining a positive and receptive attitude towards it. Preliminary work by Derwing, Rossiter, and Munro (2002) suggests that these are realistic goals.

Future directions for research

Many of the issues raised in this discussion continue to be explored in an expanding body of research on the perception of L2 speech. In the future, this work can be expected to grow to include studies of perception by a more diverse audience of listeners and interlocutors, to cover explorations of the effects of a broader range of speaking and listening conditions on the comprehension of L2 speech, and to entail development of more wide-ranging and sophisticated methods of assessing such dimensions as intelligibility and comprehensibility.

One of the most significant gaps in the current literature concerns the mutual intelligibility of different accents. Over 20 years ago, Smith and Bisazza (1982) noted the lack of research addressing how nonnative listeners perceive the speech of other nonnative speakers. They pointed out that this bias was unrealistic, because learners of English often interact in their L2 with other learners, an observation that is even more true today than it was in 1982, not only because of heterogeneity in ESL classrooms, but because of the growing number of culturally diverse contexts where English serves as a lingua franca. Yet it is only recently that more attention has been paid to the mutual intelligibility of different L2 accents and the ways in which different native accents are perceived by L2 users.

In one early study, Smith and Rafiqzad (1979) found that the intelligibility rankings of a diverse group of English speakers were quite consistent even for listeners from very different L1 backgrounds. However, their findings are hard to interpret, because the difficulty of the content of the speakers' output was confounded with accent. More recently, Major, Fitzmaurice, Bunta, and Balasubramanian (2002) examined L2 listeners' comprehension of lectures presented by L2 speakers who shared or did not share the same L1. Some of their listeners understood lectures in their own accent better than those in other nonnative accents, though the effect was small and inconsistent. Recent work by Munro, Derwing, and Morton (2006) appears to confirm that the actual advantage of hearing speech in one's own accent is small. In that study, listeners from different L1 backgrounds tended to agree with each other in terms of their evaluations of accent and comprehensibility. Moreover, speakers who were unintelligible to listeners from one particular L1 background were perceived in the same way by listeners from other backgrounds.

Some work has indicated that nonnative listeners may report irritation and distraction when listening to foreign-accented speech and that they may actually be less tolerant of it than native speakers (Fayer & Krasinski 1987). Nevertheless, Bent and Bradlow (2003) found that, in some cases, L2-speaking listeners might find foreign-accented speech just as intelligible as native-produced speech, even when they are exposed to an L2 accent that differs from their own. Research of this type needs to be extended to gain a better understanding of mutual intelligibility and the underlying causes of what Bent and Bradlow (2003) refer to as an "interlanguage speech intelligibility benefit."

A second way in which L2 speech research may expand is through work on age effects in the comprehension of accents. Given the trend toward aging populations in many western countries at a time when cultural and linguistic diversity is increasing, such work has considerable sociolinguistic importance. Recent work with elderly listeners reveals a noticeable decrement in the comprehension of L2 speech by geriatric listeners, even when their hearing loss is within typical ranges (Burda, Scherz, Hageman, & Edwards 2003). Because geriatric caregivers may speak with a different accent from their clients, this decrement needs to be explored in more detail to reveal its causes and to identify possible ways of counteracting it.

More work is also required on the effects of experience with L2 speech on comprehension. While some research indicates that familiarity leads to improved comprehension, little is understood about the underlying psycholinguistic processes that lead to such improvement. It is not known, for example, whether listeners who are familiar with particular accents develop effective strategies for processing L2 speech in general, or whether their awareness of the specific details of accents assists them.

Because communication with and among nonnative speakers occurs in a wide range of physical settings, it is also important to examine the intelligibility of L2 speech under non-ideal listening conditions, including noisy environments and over the telephone. Preliminary work on this topic by Munro (1998) indicates considerable inter-speaker variability in the way noise affects accented speech. However, very little published work exists on the issue.

The empirical study of L2 speech raises many other issues having to do with the ways in which data are collected, the kinds of speech samples that are evaluated, and the nature of the tasks and rating scales that are used. While many techniques used in this area parallel those used in clinical and more general speech research settings, it is still important to address questions concerning the validity and reliability of data obtained from listeners. In this chapter it has been possible to consider only a few of the substantive problems.

Conclusions

A consideration of the research evidence reviewed in this chapter leads to a view of L2 speech that contrasts sharply with the pedagogical outcome sought by Griffen (1980). Rather than requiring native-sounding oral output, L2 users need intelligible speech, and the latter does not necessitate perfect formal "correctness." It may therefore be concluded that, in language teaching contexts where communication is the highest priority, the goal of pronunciation instruction should be to help learners realize their full communicative potential in second language acquisition. In most cases, this requires that speakers learn to produce comfortably intelligible speech from the perspective of the diverse community of interlocutors, whether native or nonnative, in which they interact. For both researchers and pedagogical specialists this perspective entails acceptance of several underlying principles:

1. Rather than see foreign accentedness as inherently problematic in L2 oral output, we should accept it is as part of normal variation in human speech.
2. Rather than view "accent reduction" as automatically desirable for L2 speakers, we should focus on intelligibility as a more important concern. There is no reason to believe that "reducing" a speaker's accent will automatically lead to improved communication.
3. In interactive situations, we should not assign L2 speakers all the responsibility for intelligibility. Rather, we should acknowledge the role of listeners as active participants in the interaction and recognize that they may be capable of enhancing their ability to understand L2 speech.
4. We should not define the intelligibility of L2 speech solely in terms of native listeners' perceptions, but should understand it as the response of a linguistically diverse audience to the L2 speaker.

Acknowledgments

The author is grateful for Tracey Derwing's invaluable contributions to the ideas expressed in this chapter. Much of the work described here was supported by the Social Sciences and Humanities Research Council of Canada.

Suggested readings

Brennan, E. M., & Brennan, J. S. 1981. Measurements of accent and attitude toward Mexican-American speech. *Journal of Psycholinguistic Research* 10: 487–501.

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, 233–277.

Gass, S., & Varonis, E. 1984. The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning* 34: 65–89.

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

Munro, M. J. 2003. A primer on accent discrimination in the Canadian context. *TESL Canada Journal* 20 (2): 38–51.

References

Anderson-Hsieh, J., Johnson, R., & Koehler, K. 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning* 42: 529–555.

Anderson-Hsieh, J., & Koehler, K. 1988. The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning* 38: 561–613.

Barefoot, S. M., Bochner, J. H., Johnson, B. A., & vom Eigen, B. A. 1993. Rating deaf speakers' comprehensibility: An exploratory investigation. *American Journal of Speech-Language Pathology* 2: 31–35.

Bent, T., & Bradlow, A. R. 2003. The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America* 114: 1600–1610.

Bongaerts, T. 1999. Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (ed.), *Second language acquisition and the Critical Period Hypothesis*, Mahwah, NJ: Erlbaum, 133–159.

Bradlow, A. R. 2008. Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/ – /l/ contrast. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Brennan, E. M., & Brennan, J. S. 1981. Measurements of accent and attitude toward Mexican-American speech. *Journal of Psycholinguistic Research* 10: 487–501.

Brennan, E. M., Ryan, E. B., & Dawson, W. E. 1975. Scaling of apparent accentedness by magnitude estimation and sensory modality matching. *Journal of Psycholinguistic Research* 4: 27–36.

Brodkey, D. 1972. Dictation as a measure of mutual intelligibility: A pilot study. *Language Learning* 22: 203–220.

Burda, A. N., Scherz, J. A., Hageman, C. F., & Edwards, H. T. 2003. Age and understanding of speakers with Spanish or Taiwanese accents. *Perceptual and Motor Skills* 97: 11–20.

Bürki-Cohen, J., Miller, J. L., & Eimas, P. D. 2001. Perceiving nonnative speech. *Language and Speech* 44: 149–169.

Chela-Flores, B. 2001. Pronunciation and language learning: An integrative approach. *International Review of Applied Linguistics* 39: 85–101.

Chun, D. M., Hardison, D. M., & Pennington, M. C. 2008. Technologies for prosody in context: Past and future of L2 research and practice. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Derwing, T. M. 2008. Curriculum issues in teaching pronunciation to second language learners. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

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

Derwing, T. M., & Munro, M. J. 2005. Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly* 39: 379–397.

Derwing, T. M., Munro, J. J., & Wiebe, G. 1997. Pronunciation instruction for "fossilized" learners: Can it help? *Applied Language Learning* 8: 217–235.

Derwing, T. M., Munro, M. J., & Wiebe, G. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48: 393–410.

Derwing, T. M., & Rossiter, M.J. 2002. ESL learners' perceptions of their pronunciation needs and strategies. *System* 30: 155–166.

Derwing, T. M., Rossiter, M. J., & Munro, M. J. 2002. Teaching native speakers to listen to foreign-accented speech. *Journal of Multilingual and Multicultural Development* 23: 245–259.

Eckman, F. 1977. Markedness and the contrastive analysis hypothesis. *Language Learning* 27: 315–330.

Eckman, F. 2008. Typological markedness and second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Ellis, R. 1994. *The study of second language acquisition*. Oxford: Oxford University Press.

Esling, J. H. 1994. Some perspectives on accent: Range of voice quality variation, the periphery, and focusing. In J. Morley (ed.), *Pronunciation pedagogy and theory: New views, new directions*, Alexandria, VA: Teachers of English to Speakers of Other Languages, Inc., 49–63.

Esling, J. H., & Wong, R. F. 1983. Voice quality settings and the teaching of pronunciation. *TESOL Quarterly* 17: 89–94.

Fayer, J. M., & Krasinski, E. 1987. Native and nonnative judgments of intelligibility and irritation. *Language Learning* 37: 313–326.

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

Flege, J. 1988. The production and perception of foreign language speech sounds. In H. Winitz (ed.), *Human communication and its disorders, a review – 1988*, Norwood, NJ: Ablex, 224–401.

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, 233–277.

Flege, J. E., & Fletcher, K. L. 1992. Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America* 91: 370–389.

Flege, J., Munro, M., & MacKay, I. 1995. Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97: 3125–3134.

Gass, S., & Varonis, E. 1984. The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning* 34: 65–89.

Gick, B., Bernhardt, B., Bacsfalvi, P., & Wilson, I. 2008. Ultrasound imaging applications in second language acquisition. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Giles, H. & Powesland, P. F. 1975. *Speech style and social evaluation*. New York: Academic Press.

Greene, J., & Wells, E. 1927. *The cause and cure of speech disorders: A textbook for students and teachers on stuttering, stammering and voice conditions*. New York: The Macmillan Company.

Griffen, T. 1980. A nonsegmental approach to the teaching of pronunciation. *Revue de Phonétique Appliquée* 54: 81–94.

Grover, C., Jamieson, D. G., & Dobrovolsky, M. B. 1987. Intonation in English, French, and German: Perception and production. *Language and Speech* 30: 277–295.

Hansen Edwards, J. G. 2008. Social factors and variation in production. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Ioup, G. 2008. Exploring the role of age in the acquisition of a second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. 1994. Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition* 16: 73–98.

Jacques Clau v. Uniglobe Pacific Travel. 1995. *Canadian Human Rights Reporter* 23: D/515 – D/518.

Kent, R. D., Miolo, G., & Bloedel, S. 1994. The intelligibility of children's speech: A review of evaluation procedures. *American Journal of Speech Language Pathology* 3: 81–95.

Lane, H. 1963. Foreign accent and speech distortion. *Journal of the Acoustical Society of America* 35: 451–453.

Lippi-Green, R. 1997. *English with an accent: Language, ideology, and discrimination in the United States*. New York: Routledge.

Macdonald, D., Yule, G., & Powers, M. 1994. Attempts to improve English L2 pronunciation: The variable effects of different types of instruction. *Language Learning* 44: 75–100.

Major, R. C. 1987. Phonological similarity, markedness, and rate of L2 acquisition. *Studies in Second Language Acquisition* 9: 63–82.

Major, R. C. 2001. *Foreign accent: The ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Major, R. C. 2008. Transfer in second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Major, R. C., Fitzmaurice, S. M., Bunta, F., & Balasubramanian, C. 2002. The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly* 36: 173–190.

Mirek Gajecki v. Board of Trustees, School District No. 36 (Surrey). 1990. *Canadian Human Rights Reporter*11: D/326–332.

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

Munro, M. J. 1993. Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech* 36: 39–66.

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

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

Munro, M. J. 2003. A primer on accent discrimination in the Canadian context. *TESL Canada Journal* 20 (2): 38–51.

Munro, M., & Derwing, T. 1994. Evaluations of foreign accent in extemporaneous and read material. *Language Testing* 11: 253–266.

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

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

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

Munro, M. J., & Derwing, T. M. 2006. The functional load principle in ESL pronunciation instruction: An exploratory study. *System* 34: 520–531.

Munro, M. J., Derwing, T. M., & Burgess, C. S. 2003. The detection of foreign accent in backwards speech. In M. J. Solé, D. Recasens, & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Futurgraphic, 535–538.

Munro, M. J., Derwing, T. M., & Burgess, C. S. (forthcoming). The bases of foreign accent detection.

Munro, M. J., Derwing, T. M., & Morton, S. L. 2006. The mutual intelligibility of foreign accents. *Studies in Second Language Acquisition* 28: 111–131.

Oyama, S. 1982. A sensitive period for the acquisition of a nonnative phonological system. In Krashen, S., Scarcella, R., & Long, M. (eds.), *Child-adult differences in second language acquisition*. Rowley, MA: Newbury House.

Pennington, M. C. 1996. *Phonology in English language teaching*. London: Longman.

Perlmutter, M. 1989. Intelligibility rating of L2 speech pre- and postintervention. *Perceptual and Motor Skills* 68: 515–521.

Rubin, D. L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education* 33: 511–531.

Schairer, K. E. 1992. Native speaker reaction to non-native speech. *Modern Language Journal* 76: 309–319.

Schmid, P. M., & Yeni-Komshian, G. H. 1999. The effects of speaker accent and target predictability on perception of mispronunciations. *Journal of Speech, Language, and Hearing Research* 42: 56–64.

Scovel, T. 1988. *A time to speak: A psycholinguistic investigation into the critical period for human speech*. New York: Harper and Row.

Shah, A. P. 2003. Acoustic indicators of Spanish-accented English. *Canadian Acoustics* 31: 42–43.

Smith, L. E., & Bisazza, J. A. 1982. The comprehensibility of three varieties of English for college students in seven countries. *Language Learning* 32: 259–269.

Smith, L. E., & Rafiqzad, K. 1979. English for cross-cultural communication: The question of intelligibility. *TESOL Quarterly* 13: 371–380.

Southwood, M. H., & Flege, J. E. 1999. Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics and Phonetics* 13: 335–349.

Subtelny, J. D. 1977. Assessment of speech with implications for training. In F. H. Bess (ed.), *Childhood deafness: Causation, assessment, and management*, New York: Grune & Stratton, 183–194.

Tajima, K., Port, R., & Dalby, J. 1997. Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics* 25: 1–24.

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

Van els, T., & De Bot, K. 1987. The role of intonation in foreign accent. *Modern Language Journal* 71: 147–155.

Varonis, E. M., & Gass, S. 1982. The comprehensibility of non-native speech. *Studies in Second Language Acquisition* 4: 114–136.

Weill, S. A. 2003. The impact of perceptual dissimilarity on the perception of foreign accented speech. Unpublished doctoral dissertation, The Ohio State University.

Wilson, I. L. 2006. Articulatory settings of French and English monolingual and bilingual speakers. Unpublished doctoral dissertation, Department of Linguistics, University of British Columbia.

Zampini, M. L. 2008. L2 speech production research: Findings, issues and advances. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

CHAPTER 8

L2 speech production research

Findings, issues, and advances

Mary L. Zampini
Le Moyne College

Introduction

Despite being published a half century ago, the Contrastive Analysis Hypothesis (CAH) presented by Lado (1957) continues to be invoked in much second language (L2) speech acquisition research today. It is well known (and repeated throughout this volume) that many adult learners speak their L2 with a foreign accent and that, furthermore, the learner's first language (L1) can play an inhibitive role in L2 speech perception, processing, and production. The CAH highlights this role in its simplest form by predicting that those aspects of the L2 sound system that are similar to the L1 will be easy to acquire, while those aspects that are different from the L1 will be difficult (the CAH is also discussed in detail in Chapter 2 by Major, this volume). Much subsequent research, however, has found that the role of the first language (L1) in L2 phonological acquisition is not so straightforward. Some L2 sounds that are very different from the L1 may be relatively easy to acquire, while other sounds that are similar to the L1 may be difficult. In addition, there are a myriad of other factors that affect L2 phonological acquisition that may mitigate or heighten the role of the L1, such as age, markedness, and social factors (see Chapters 2, 4, and 9 in this volume, respectively, for a further discussion of these factors). The goal of this chapter is to examine research on L2 speech production, to highlight the relevant findings, theoretical contributions, and methodological approaches, and to offer some insights regarding the future directions and promising avenues of research.

The field of L2 speech production is vast, and it would be impossible to cover all the areas of focus within the confines of this chapter. This chapter will therefore focus on studies that examine the nature of the L2 speech sounds produced by learners. The relationship between these studies and other areas of focus in this book (e.g., L2 speech training, speech perception, L1 transfer, etc.) will be made where appropriate, but will not be discussed in detail. The chapter is outlined as follows. The first section will survey major research findings on the articulation of L2 speech sounds from recent years. It will examine a range of empirical studies that focus on particular aspects of L2 speech, including those that focus on sounds at the (sub)segmental level (stop consonants, liquid consonants, and vowels), as well as the suprasegmental level (syllable structure, prosodic domains, and stress). On the basis of the literature review, the next section will examine methodological options for carrying out research on L2 speech production. The implications that the research findings have for more general models of L2 phonology and acquisition will then be discussed. Finally, the chapter will conclude by examining a number of outstanding issues and considering future directions in the field given both recent trends in the literature and technological advances.

Review of the literature

Virtually all research on L2 speech production assumes that the learner's L1 sound system impacts L2 pronunciation, at least some of the time or in certain stages of L2 acquisition.[1] This transfer of knowledge of the L1 to the L2 can have a facilitative effect on L2 pronunciation (e.g., for those areas where both the L1 and L2 sound systems are the same), or may hinder acquisition. As such, it is impossible to address the findings and contributions of the body of literature on L2 speech without reference to the role of L1 transfer. The reader is urged, therefore, to consult Chapter 3 by Major in this volume, which provides a more complete discussion of the role of transfer in L2 speech.

For ease of presentation, the literature review is divided into subsections that examine a particular group of sounds or some suprasegmental aspect of speech. Segmental studies are addressed first, with subsections on the L2 production of stop consonants, vowels, liquids, and phoneme-level or allophonic substitutions. Affricates, fricatives, and nasals are not addressed, since very few studies on their L2 production appear in the literature. Suprasegmental aspects of L2 speech production are examined next, including syllable structure, prosodic domains and stress.

---

1. As Hansen Edwards notes in Chapter 9, it may not only be the L1 sound system, but possibly also the L1 identity, that affects L2 pronunciation.

(Sub)segmental aspects of L2 speech production

Stop consonants

Stop consonants, such as voiceless /p t k/ and voiced /b d g/, are one of the most widely examined classes of sounds in L2 speech production studies. One of their primary articulatory characteristics is voice onset time, or VOT, which refers to the time that elapses between the release of the obstructed airflow (release burst) and the beginning of vocal cord vibration (voicing). Stops are classified in one of three categories according to the average VOT duration: *long-lag* stops are produced with long VOT durations (generally longer than 35 milliseconds [ms.]), *short-lag* stops have short VOT durations (0–35 ms.), and *prevoiced* stops exhibit voicing throughout the closure and are therefore often expressed as having negative VOTs (Lisker & Abramson 1964; Keating 1984). Figure 1 illustrates this category distinction with waveforms showing stop consonants of three different VOT durations. Some languages (e.g., Thai) have a three-way phonemic distinction among the categories shown in Figure 1. Other languages have just a two-way phonemic distinction. Crucially, however, languages may have the same *phonemic* distinction (e.g., voiceless /p t k/ vs. voiced / b d g/), but differ with respect to the *phonetic* realization of those phonemes. For example, the voiceless stops /p t k/ are long-lag stops in English, while voiced /b d g/ are short-lag stops. In Spanish and French, on the other hand, voiceless /p t k/ are short-lag stops, while voiced /b d g/ exhibit prevoicing. Thus, short-lag VOTs characterize the voiceless phonemes /p t k/ in Spanish and French, but the voiced phonemes /b d g/ in English.

Because of this kind of cross-linguistic variation, many researchers have examined L2 stop consonants to determine the extent to which learners produce them with native-like VOT durations[2] (see Ioup, Chapter 2 of this volume, for discussion of similar VOT studies related to age of acquisition and foreign accent detection). In an early study, Flege (1987b) examined the production of L2 French /t/ by three groups of first language (L1) speakers of English with varying degrees of experience in French. Among his results, Flege found that the least experienced learners (American students of French) produced L2 French /t/ with a mean VOT duration that was like that of monolingual English speakers' VOT for English /t/; the more experienced learners (American instructors of university-level French and Americans living in Paris), on the other hand, produced French /t/ with significantly shorter VOT durations than those for monolingual English /t/. Flege also found

2. It should be noted, however, that most VOT studies focus on word-initial voiceless stops in L2 English (/p t k/). Less work as been done on the production of L2 voiced stops (e.g., /b d g/), word-medial stops, and L2 stops in languages other than English. VOT is not typically examined in word-final position, because such sounds may be unreleased; word-final L2 stops may also trigger processes related to syllable structure (e.g., deletion or epenthesis).

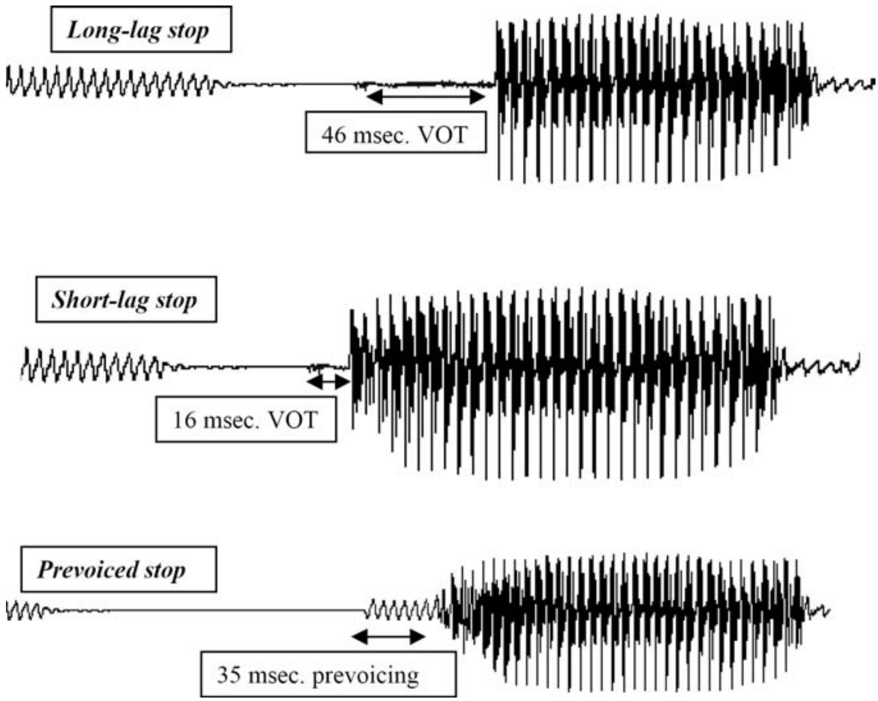


Figure 1. Phonetic categories of stop consonants

that only the most experienced speakers of L2 French (Americans living in Paris) produced French /t/ with a mean VOT duration that was like that of monolingual French speakers. The American professors of French produced L2 French /t/ with a mean VOT that was unlike either L1 or L2–significantly shorter than monolingual English /t/, but still significantly longer than monolingual French /t/.

Two key findings emerge from this study. First, inexperienced English-speaking learners of French do not differentiate L1 and L2 /t/. Flege (1981, 1987b; see also Flege & Hillenbrand 1984) proposes that this is due to *equivalence classification*, "...a basic cognitive mechanism which permits humans to perceive constant categories in the face of inherent sensory variability found in the many physical exemplars [of a given] category" (Flege 1987b: 49). In other words, although French /t/ is different from English /t/, it is similar enough that English-speaking learners consider French /t/ to be an instance of English /t/. This equating of the L1 and L2 sounds inhibits the establishment of a separate category for the L2 sound. Because of this, Flege proposes that "similar" L2 sounds, such as English and French /t/, may be harder to acquire than L2 sounds that are unlike anything found in the L1. The notion of equivalence classification became a central notion in Flege's Speech Learning Model (SLM) of second language acquisition (Flege 1995) and has had enormous impact on studies of L2 speech production. The SLM will be discussed in more detail below (see also Chapter 6 by Strange and Shafer, this volume).

The second important finding from Flege (1987b) is that even experienced L2 learners (e.g., the American instructors of French of this study) may produce L2 French /t/ with intermediate VOT durations that fall in between those for monolingual speakers of either L1 or L2. In fact, many researchers have reported that bilinguals show evidence of intermediate, or "compromise," VOT values for L2 (e.g., Williams 1977; Flege & Hillenbrand 1984; Flege & Eefting 1987; Flege 1991; Hazan & Boulakia 1993; Thornburgh & Ryalls 1998). Moreover, researchers have found evidence of compromise VOT values in bilinguals' L1 productions as well. For example, Flege (1987b) also found that L1 speakers of French who had been living in the United States for many years produced French /t/ with a mean VOT duration that was significantly longer than that for monolingual speakers. In another study, Major (1992) found compromise VOT values for L1 English stops produced by bilinguals from the United States who had been living in Brazil for several years. Moreover, he found that the bilinguals' L1 English productions were even less English-like and more Portuguese-like in casual, rather than formal, speech. These results suggest that L1 phonetic representations may be restructured in response to the acquisition of L2.

Beyond a strict examination of learners' or bilinguals' productions of L2 stop consonants, many researchers have considered factors that may influence the degree to which learners are able to produce native-like VOT durations. One of the most commonly studied variables is age of acquisition of L2 (cf. Ioup, Chapter 2 of this volume). Many researchers have found that learners and bilinguals who acquire L2 at an early age (before the age of six) are more likely to produce L2 stops with monolingual-like VOT durations than those who acquire L2 after the age of six or seven. In an early study, for example, Williams (1977) found that children who had learned English and Spanish by the age of six were able to produce stops in both languages with VOT durations that were similar to those for monolingual speakers. Mack (1989) showed similar results for children who had learned English and French. In a comparative study of both early and late learners, Flege (1991) examined the production of the L2 English voiceless stops /p t k/ by L1 speakers of Spanish. He found that early learners produced L2 /p t k/ with mean VOT durations that were similar to those for monolingual English speakers. Late learners, however, produced L2 /p t k/ with compromise VOT values. Flege concluded that the early learners had been able to establish separate phonetic categories for L1 Spanish and L2 English stops, while the late learners were inhibited by equivalence classification.[3]

3. Flege and colleagues (e.g., Flege, Munro & MacKay 1995a, 1995b) have also examined age of arrival and length of residence as factors affecting L2 speech production. In such cases, the

Another area of research has examined the effect of speaking rate on the production of L2 stop consonants (e.g., Schmidt & Flege 1995, 1996; Magloire & Green 1999). The duration of VOT may change as a function of speaking rate; for example, the VOT of the English voiceless stops /p t k/ shortens as speaking rate increases (Miller, Green & Reeves 1986; Volaitis & Miller 1992; Kessinger & Blumstein 1997). The L2 learner, therefore, must learn to make rate-related VOT adjustments in order to approximate native speaker norms in the L2. As Schmidt and Flege (1996) note, the fact that L2 learners or bilinguals may produce L2 stops accurately at a normal speaking rate does not necessarily mean that they have established separate phonetic categories for the L2 sounds. It may simply mean that they can produce such sounds under "conscious (i.e., nonautomatic) control" (Schmidt & Flege 1996: 166). If the learner, on the other hand, can show native-like norms with respect to VOT at a variety of speaking rates, one can make a stronger argument in favor of the learner having established accurate L2 phonetic categories.

In their study, Schmidt and Flege (1996) examined the production of /p/ and /t/ at three different speaking rates (normal, fast, and slow) by monolingual speakers, early Spanish-English bilinguals, and late Spanish-English bilinguals. Among their results, they found that the early bilinguals produced L2 English VOTs at each of the three speaking rates that were similar to those of the monolingual English speakers, whereas the late bilinguals exhibited compromise VOTs in L2 English. As for changes in speaking rate, the English monolinguals produced /p, t/ with significantly longer VOTs at the slow rate compared to the normal rate and significantly shorter VOTs at the fast rate compared to the normal rate, as expected. The early bilinguals also produced the L2 English stops with significantly shorter VOTs at the fast rate compared to the normal rate, but did not produce a significant VOT difference in the normal vs. slow rates. The late bilinguals, on the other hand, did not show significant VOT differences in either the fast-vs. normal rates, or slow vs. normal rates. Thus, Schmidt and Flege found additional age-related differences in the production of L2 English stops, since the early learners came closer to approximating native speaker norms across speaking rates than the late learners.

The large number of studies showing age-related differences in L2 speech production led Flege to develop the SLM. This model, as presented by Flege (1995), consists of four postulates and seven hypotheses that attempt to predict how L2

---

subjects are immigrants to a country where the L2 is spoken. The hypotheses and findings generally conform to those for age of acquisition: the subjects who immigrate at a young age are more likely to reflect native speaker norms in L2 speech production than those who immigrate as adults. Likewise, the longer an immigrant resides in the country of the L2, the more likely s/he is to reflect native-like norms in production (although age of arrival may be a stronger predictor of native-like attainment in L2 production).

learners will behave with respect to the L2 sound system. As such, it has proved to be an influential and useful tool for researchers, because it "...generates testable predictions and can be used to organize and interpret a wide range of empirical data" (Flege 1995: 238). According to the SLM, there is no critical period after which the learner will be unable to acquire an L2 sound system – the mechanisms used to acquire the L1 sound system remain in place throughout adulthood, and the L1 phonetic categories evolve over time as new sounds and languages are learned, such that they "...reflect the properties of all L1 or L2 phones identified as a realization of each category" (p. 239). Equivalence classification, however, can inhibit L2 acquisition and prevent the establishment of a separate L2 phonetic category; as discussed above, this occurs when the L2 sound is sufficiently similar to the L1 sound that it is interpreted as another instance of the L1 category. For a complete description of the hypotheses of the SLM, see Flege (1995: 239); see also Strange and Shafer, this volume.

In spite of the age-related trends in L2 acquisition presented thus far, not all studies show an age effect on L2 stop consonant production. For example, Magloire and Green (1999) found that both early and late Spanish-English bilinguals produced L2 English stops with English-like VOT durations, even under extreme changes in speaking rate. They argue that their results were due, at least in part, to careful control of "language mode" throughout the experimental procedure (see also Zampini & Green 2001). That is, their subjects were recruited for English and Spanish production studies independently, and in neither case were they told of the researchers' interest in their bilingual capabilities. In addition, all interaction with the subjects (including recruitment, explanation of the experimental tasks, and interaction with the laboratory assistant) was conducted in the language of the experiment. In this way, the researchers hoped to place their subjects in a "monolingual mode" as much as possible, whereby the language under study was the primary one activated. If issues of language mode and methodological design are born out by future studies, the SLM will have to be modified to take into account language activation. That is, it may be that when bilinguals are in a monolingual mode and have one language activated, their speech is more likely to reflect native-like norms than if they are in a bilingual situation where both languages are activated (see also Grosjean 1998). Moreover, work by Green, Zampini and Magloire (1997) showed that while bilingual speakers may be able to approximate native-like norms with respect to some characteristics, like VOT, both early and late learners may still differ from monolingual speakers with respect to other characteristics of stop consonant articulation, such as the mean duration of the voiceless closure that precedes the release burst (see Zampini & Green 2001 for further discussion).

To summarize, research on the production of L2 stop consonants has focused most often on L2 learners' production of one acoustic cue, VOT. Many of these studies have found that learners who acquire their L2 before six years of age are more likely to produce L2 stops with native-like VOTs, whereas late learners are more likely to exhibit compromise VOT durations that are in between those found for the L1 and L2. Similar differences are found across speaking rates. Some recent research, however, has found that even late learners can produce L2 stops with native-like norms when language activation is controlled. In addition, learners may be more likely to produce some characteristics of L2 stops (e.g., VOT) more accurately than others. Thus, future studies will need to examine a variety of acoustic cues in order to gain a more complete picture of the nature of L2 stops. It is also worth noting that most studies of L2 stop production have focused on bilinguals who are fluent in the L2. In fact, Flege (1995) even asserts that the SLM is primarily concerned with the "ultimate attainment of L2 pronunciation" (Flege: 1995: 238). Few studies have examined the production of L2 stops by beginning learners, nor have they focused on the acquisition of stops over time (but see Zampini 1998b).

Vowels

Languages differ greatly with respect to the number and types of vowels in their phonemic inventory, and as a result, they provide a wealth of opportunities for researchers in L2 acquisition. In phonological terms, vowels are classified and distinguished in part by the relative position of the tongue in the mouth during articulation; that is, vowels may be classified in terms of tongue height (e.g., high, mid, low) and frontness / backness (e.g., front, central, back). These properties are reflected acoustically to some degree in the *formant frequencies* associated with each vowel. The formant frequencies refer to the characteristic "pitch overtones" of a given vowel as a function of the size and shape of the articulatory tract (Ladefoged 2001). There are two primary formants that distinguish vowels: the first formant (F1) and second formant (F2). F1 increases as vowel height decreases (that is, high vowels have lower F1s than low vowels), and F2 generally decreases as the vowel's backness increases. To illustrate, Figure 2 shows the average F1 and F2 values (in Hertz) of American English /i ɪ ɛ æ ɑ ɔ u/.

In addition to the acoustic, or spectral, quality of vowels, quantity may also play a distinctive or phonetically prominent role in a given language. Thus, some languages, like Finnish, have a phonological contrast between long and short vowels that have otherwise similar spectral properties. Other languages, like English, have long and short vowels, but the long-short pairs also exhibit spectral differences (e.g., English /i/ has a lower F1 value and higher F2 value than /ɪ/). Still other languages, like Spanish, do not show any significant durational differences for vowels at all. Both vowel quality and quantity may be measured fairly readily through acoustic analysis. With regard to L2 speech research, therefore, a common methodological approach is to examine the L2 vowels produced by learners

	F1	F2
/i/ (*beat*)	280	2250
/ɪ/ (*bit*)	400	1920
/ɛ/ (*bet*)	550	1770
/æ/ (*bat*)	690	1660
/ɑ/ (*pot*)	710	1100
/ɔ/ (*bought*)	450	1030
/u/ (*boot*)	310	870

Figure 2. Average frequencies (in Hz) for F1 and F2 of American English /i ɪ ɛ æ ɑ ɥ u/ (Ladefoged 1982: 176)

and compare characteristics such as the average formant frequencies or duration of articulation to those for monolingual speakers.

Many of the questions that drive experimental studies in L2 vowel production are similar to those seen for stop consonants. For example, a number of studies have examined the role that age of acquisition plays in L2 vowel production and have found that early L2 learners are more likely to produce L2 vowels accurately than late L2 learners (e.g., Jun & Cowie 1994; Munro, Flege & MacKay 1996). In addition, Bohn and Flege (1992) found length of residence to be an important factor in L1 German speakers' ability to accurately produce L2 American English /æ/. The experienced learners (who had resided in an English-speaking environment for an average of 7.5 years) produced English /æ/ like monolingual English speakers, whereas the inexperienced learners (who had resided in an English-speaking environment for 0.6 years) did not.

Flege, Bohn and Jang (1997) followed this study with another that examined the production of L2 English vowels by experienced and inexperienced L1 speakers of German, Spanish, Mandarin, and Korean. Many of their results also suggested that experience with English contributed to more accurate production of the L2 English vowels. For example, native English speakers produce a significant height difference between the vowel pairs /i–ɪ/ and /ɛ–æ/. Flege et al. found that the experienced Mandarin subjects produced a significant height distinction between /i–ɪ/, but the inexperienced Mandarin speakers did not. In addition, the experienced German and Mandarin subjects showed a significant height distinction between /ɛ–æ/, but the inexperienced German and Mandarin subjects did not. As for the other language groups, both the inexperienced and experienced subjects showed similar results. Both the inexperienced and experienced German subjects produced a significant height distinction for /i–ɪ/, whereas neither the experienced nor inexperienced Korean and Spanish speakers did. As for /ɛ–æ/, neither the experienced nor inexperienced Korean speakers produced a significant height distinction, but both the inexperienced and experienced Spanish speakers did.

These results illustrate some of the complexities of L2 speech production research, since length of residence does not seem to be a uniform predictor of performance, and one must look to other factors that may help explain the data. In this case, Flege et al. (1997) considered the vowel inventory of the different subject groups' L1 and the "perceived phonetic similarity" between the L1 and L2 vowels. To illustrate, German has a phonemic distinction between /i/ and /ɪ/, similar to English. It is not surprising, therefore, that both the inexperienced and experienced German subjects produced a significant height distinction between these two L2 phones. Korean, Mandarin, and Spanish all have /i/, but none has /ɪ/ in its phonemic inventory. L1 speakers of these languages may therefore perceive L2 English /ɪ/ as equivalent, or similar, to /i/ (equivalence classification); this would explain the finding that neither the inexperienced nor the experienced Korean and Spanish speakers produced a significant height distinction between L2 English /i–ɪ/ (see Flege et al. for discussion regarding the remaining results).

In a recent study, McAllister, Flege and Piske (2002) also examined L1 influence on L2 vowel production, but in this case, they considered the role that vowel quantity (or duration) plays in the L1 and L2. The authors examined the L2 production of Swedish vowels by L1 speakers of Estonian, American English, and Latin American Spanish. Four long-short vowel pairs were tested: high /uː, u/, mid /oː, o/ and /ɛː, ɛ/, and low /aː, a/. These vowel pairs may also show spectral differences in Swedish, but McAllister et al. (p. 233) report that duration is the primary cue used to distinguish the mid vowels, whereas spectral characteristics are the primary cue used to distinguish the high and low vowels. The Swedish situation differs from the L1s of the subjects in the McAllister et al. study in the following ways (p. 232): Estonian, like Swedish, has a phonemic contrast between long and short vowels, but Estonian speakers use duration, rather than spectral quality, as the primary cue to distinguish long-short vowel pairs.[4] English also has long and short vowels, but spectral qualities are the primary cue used to distinguish these pairs. Finally, Spanish vowels are all short in duration; quantity, therefore, does not play a significant role in distinguishing Spanish vowels.

Based on the role of duration in the subjects' L1, McAllister et al. predicted (pp. 233–234) that L1 speakers of Estonian would be most successful at producing the L2 Swedish long-short contrast, since duration serves a contrastive function in their L1. They also predicted that, of the four L2 vowel pairs, the mid vowels would be most difficult for L1 speakers of English and Spanish, since duration is the primary cue used to distinguish these pairs. The results of their production

4. There are also co-occurrence restrictions for the long and short vowels in Swedish, but not Estonian: if the vowel is long in Swedish, the following consonant is short (V:C), whereas if the vowel is short, the following consonant is long (VCC, where CC may be a geminate or consonant cluster).

experiment showed that all speakers were relatively successful at producing the L2 Swedish quantity distinction. The L1 speakers of Spanish, however, produced a significantly smaller quantity distinction for the mid vowels than the L1 speakers of English and Estonian, as well as the monolingual speakers of Swedish. This, together with a corresponding perception study, led the authors to conclude that the results supported their hypothesis that the quantity distinction would be harder to learn for the mid vowels than the non-mid vowels. In addition, the results indicated that L1 speakers of English are more successful than L1 speakers of Spanish in acquiring the quantity distinction. McAllister et al. suggest that this is due to the fact the English vowels do exhibit durational differences, even though spectral qualities are more prominent.

To summarize, studies of L2 vowel production tend to examine spectral characteristics (especially F1 and F2) and/or durational characteristics of the vowels produced by L2 learners or bilinguals. In addition, age of acquisition and length of residence have been examined as factors that may affect the extent to which learners produce L2 vowel with native-like norms. The nature of the L1 vowel inventory, however, has also been shown to influence L2 production (see Major, Chapter 3 of this volume) and may even prove more influential than age of acquisition or length of residence in some cases.

Liquids

Like vowels, liquid consonants (e.g., /r/ and /l/) show much cross-linguistic variation with respect to their phonemic status and phonetic realization; as such, they can prove difficult for speakers of one language learning another. The majority of research on L2 liquids has focused on L1 Japanese learners of L2 English and, furthermore, the extent to which training can improve learners' perception and production of L2 English /ɹ/ and /l/. Studies on training will not be addressed here, because they are discussed extensively by Bradlow in Chapter 10 in this volume (see also Chapter 6 by Strange and Shafer for related issues dealing with perception). Instead, this section will focus on two other kinds of studies that have dealt with L2 liquids.

First, and following in the same vein as some of the research discussed above, a recent study by Aoyama, Flege, Guion, Akahane-Yamada and Yamada (2004) examined the role of the L1 and perceived phonetic (dis)similarity between L1 and L2 sounds in the production of L2 English /ɹ/ and /l/ by L1 Japanese speakers. For Japanese speakers, English /l/ is perceptually more similar to Japanese /r/ than English /ɹ/. Thus, the authors hypothesized that L1 Japanese learners would have more difficulty in acquiring L2 English /l/ than /ɹ/. They examined the L2 perception and production of these phones by L1 Japanese children and adults at two different intervals (T1 and T2, separated by one year) and found some support for their hypothesis. For example, the Japanese children's perception of the /l–ɹ/ and

/ɹ–w/ contrasts was significantly better at T2 than at T1; the Japanese adults, on the other hand, performed better than the children at T1 but did not show improvements over time. In addition, the children showed greater improvement from T1 to T2 in the production of L2 English /ɹ/ than /l/.[5] The adults showed better production of L2 English /ɹ/ than the children at T1, but there were no significant differences between the children and the adults at T2. Thus, the children, but not the adults, improved their production of L2 English /ɹ/. As for the production of L2 English /l/, neither the children nor the adults showed significant improvement over time. In summary, this study showed better acquisition of the more dissimilar L2 sound (/ɹ/) than the similar one (/l/), as predicted. Age–related differences in the rate of acquisition were also apparent, however, since the observed gains over time occurred only for the L1 Japanese children.

In another type of study dealing exclusively with rhotics, Major (1986) examined the acquisition of the Spanish apical flap /ɾ/ (as in the word *pero*, '*but*') and trill /r/ (as in *perro*, '*dog*') by L1 speakers of English enrolled in an intensive beginning-level Spanish course. The subjects were tested seven times at approximately weekly intervals; each time, they were asked to answer questions and read a word list and sentence list. Their productions of the Spanish rhotics were subsequently transcribed and analyzed by the author. Major used this case as a test for his Ontogeny Model of phonological development, which makes predictions regarding the types of production errors that L2 learners will make over the course of acquisition. Specifically, the model predicts that errors due to L1 transfer will be numerous in the early stages of acquisition but will decrease over time. Errors that are due to developmental factors, on the other hand, will be infrequent at first but will increase over time and eventually decrease again as acquisition is attained. Developmental factors are due to universal language acquisition processes (Major 1986: 461) and may include the deletion or insertion of segments, approximation (the pronunciation of a sound not found in either the L1 or the L2), assimilation, and overgeneralization. The results of Major's study supported his model: in general, the number of errors due to L1 transfer (e.g., the production of English /ɹ/ for Spanish /ɾ/) were numerous at first but began to decrease over time. In addition, developmental errors, such as the deletion of L2 Spanish /r/, the substitution of /l/ for /r/, and the pronunciation of a voiced uvular trill [ɾ] or voiceless uvular fricative [X] for L2 Spanish /r/, began to increase over the course of the experiment. This study is important, because it was one of the first in which Major set forth the Ontogeny Model as a means for predicting L2 speech errors. Like the SLM,

5. The production data were evaluated for intelligibility by native English speakers. This method of assessment is common in many L2 speech production studies (e.g., Riney & Flege 1998; Guion, Flege & Loftin 2000). The chapter by Munro, this volume, discusses issues related to foreign accent and intelligibility in detail.

the Ontogeny Model has proved to be useful tool for L2 speech researchers, and it has been tested in a number of studies since then (see also Major 2001, as well as Chapter 3 by Major in this volume for detailed discussion of the Ontogeny Model).

Sound substitution studies

Another area of focus in L2 speech research has involved an examination of phoneme-level substitutions (rather than sub-phonemic characteristics of L2 sounds) that take place in L2 speech. Such studies usually attempt to explain why a particular substitution takes place and/or describe the conditions under which the substitution may occur. In some cases, a sound substitution occurs when the L2 target sound does not appear in the L1; the learner therefore, substitutes[6] the L2 target with another sound, usually from the L1 (e.g., the substitution of English /ɹ/ for L2 Spanish /ɾ/, as just seen). In other cases, the L2 target sound may be a phoneme in the L1 but not appear in the required L2 context. For example, an L1 speaker of German may pronounce L2 English *log* with a final [k] sound, as opposed to [g], even though /g/ is a phoneme of German. German voiced stops are devoiced in word-final position, however, and the substitution of /k/ for L2 English /g/, therefore, is the result of an L1 phonotactic constraint regarding the distribution of /g/. Eckman (1977) takes up this example and further notes that L1 English learners of German, on the other hand, generally do not have difficulty acquiring the L2 final-stop devoicing rule; he poses the question as to why this would be the case. The Contrastive Analysis Hypothesis would predict that since German and English differ with regard to the distribution of word-final stops, their realization should be difficult for L1 speakers of either language learning the other. Eckman proposes the Marked Differential Hypothesis (MDH) to explain this difference (the MDH is discussed in more detail in Chapter 4 by Eckman, this volume). In essence, the MDH modifies the CAH by attempting to take into account the role that implicational universals play in L2 acquisition. It predicts that those areas of the L2 that are different from the L1 will be difficult to acquire if they are also more marked than the L1. Aspects of the L2 that are different but less marked will not be difficult to acquire. For example, voiceless stops are typologically more common, and therefore less marked, than voiced stops. In the case of German and English, therefore, the MDH predicts that L2 German word-final stop devoicing is not difficult for learners of L1 English. However, since word-final voiced stops are more marked, this aspect of L2 English will be difficult for L1 speakers of German to acquire. Eckman's MDH, along with its more recent modifications, has had

6. The use of the term "substitution" is not entirely accurate, in that it assumes that the learner's underlying representation for the L2 target is identical to the L2 target, and that the substituted sound is therefore a phonetic variant of this underlying representation. This, however, may not be the case.

considerable influence in the literature on L2 speech production, especially with regard to the acquisition of L2 syllable types, a topic which will be addressed in more detail below.

As mentioned above, sound substitutions may also occur when the L1 does not have the corresponding L2 phone. An oft-studied substitution of this type occurs with the L2 English interdental fricative /θ/, the initial sound of the word *thing* (e.g., Altenberg & Vago 1987; Schmidt 1987; Weinberger 1990, Lombardi 2003). Lombardi (2003: 228) points out, for example, that L1 speakers of Hungarian, Russian, and Thai tend to substitute [t] for L2 English /θ/, whereas L1 speakers of Egyptian Arabic, German, and Japanese tend to substitute [s]. The substitution of /t/ for /θ/ involves a change in the manner of articulation, from a fricative to a stop. The substitution of /s/ for /θ/ involves a changes in point of articulation, from an interdental to an alveolar sound. Interestingly, all the languages mentioned by Lombardi contain both /s/ and /t/ as part of their phonemic inventories. This leads one to question why L1 speakers of one language consistently substitute one sound (/t/) for the L2 target, while L1 speakers of other languages substitute a different sound (/s/), when both sounds are available.

Lombardi (2003) uses Optimality Theory (OT) to explain this dichotomy and argues that in both cases, the particular substitution reflects the L1 hierarchy with respect to the ranking of the relevant constraints (see Chapter 5 on OT by Hancin-Bhatt, this volume). In the case of L1 speakers who substitute /t/ for L2 English /θ/, Lombardi argues that the substitution reflects typological markedness relationships and is therefore the result of Universal Grammar. That is, when given the choice of substituting either /s/ or /t/ for L2 English /θ/, the speaker chooses the lesser marked segment, /t/.[7] In the case of speakers who substitute more marked /s/ for L2 English /θ/, Lombardi finds that their L1 exhibits certain alternations (or sound substitutions) that require that the manner of articulation of the original segment be maintained in the alternation – that is, if the underlying phoneme is a fricative, then the phonetic variant must also be fricative. If the L1, therefore, has a constraint ranking that requires faithfulness to individual manner features, this ranking will be reflected in the L2, which in turn will result in the substitution of /s/ for L2 English /θ/, since both /s/ and /θ/ are fricatives.

To summarize this section, sound substitutions in L2 speech may occur when the L1 does not contain the L2 target phone, or when the L2 target appears in a position prohibited by the L1. Studies of sound substitutions have attempted

7. The nature of the markedness relationship between /t/ and /s/ is derived from a number of considerations. First, stops are more common cross-linguistically, and a language that contains fricatives will also necessarily contain stops; hence, /t/ is typologically less marked than /s/. In addition, the substitution of /t/ for /θ/ reflects L1 acquisition behavior: stops tend to be acquired first and often appear in child speech as substitutions for target fricatives.

to describe the types of substitutions that occur and explain their source. While an appeal to Universal Grammar or markedness is often made to explain some substitutions, many can be traced to the L1 phonemic inventory and phonology.

Suprasegmental aspects of L2 speech production

Syllables

The sound substitutions of the type discussed by Lombardi (2003) are generally not context-driven – that is, the substitution occurs regardless of the phonetic context that surrounds the L2 target. In other cases, a substitution may occur because of the L2 target sound's position within the word or utterance, as seen in the case of the L1 German speaker who substitutes a voiceless stop for L2 English word-final voiced stops. As indicated above, this latter example involves the transfer of an L1 allophonic rule. This particular case may also be viewed, however, as an example of the transfer of L1 syllable constraints to the L2, in which German may be said to have a constraint against voiced stops in word-final coda position. Indeed, the production and acquisition of L2 syllables is an area of research that has received considerable attention in the literature, and a number of important contributions have been made.

Most studies of L2 syllable productions entail an error analysis of the mistakes that learners make in the production of various L2 syllable types, especially those which contain complex onsets or codas. Of particular interest is how learners acquire an L2 system that allows a broader range (i.e., more phonemes) and/or more complex range of onsets and codas (i.e., sequences of two or more consonants) than the L1. Common errors in the production of L2 syllables include epenthesis (e.g., inserting a vowel sound in the middle of a consonant cluster or after a word-final consonant), deletion (especially of a consonant in a complex onset or coda), and modification of the L2 target, or feature change (such as the devoicing of a stop consonant in coda position). To give an example, Broselow, Chen and Wang (1998) found evidence for all three processes in L1 Mandarin speakers' productions of L2 English codas. Their discussion focused on the production of the English stop consonants /p t k b d g/ in final position, and the authors found that when faced with a word-final stop, their subjects inserted a schwa after the stop to create disyllabic word more than a third of the time (e.g., pronouncing the nonsense word [vɪg] as [vɪ.gə], where the period represents a syllable division, p. 263). In addition, they deleted the stop consonant more than 40% of the time, and 19% of the L2 English word-final voiced stops were produced as voiceless stops (Broselow et al. 1998: 264).

In addition to identifying error types, research on L2 syllable productions has attempted to explain the source of the errors. Two of the most commonly studied factors include L1 transfer and universal aspects of syllable structure. Hansen

(2004: 85–89) provides a thorough summary of the overall findings of much of this research. She notes (p. 86), for example, that L1 transfer appears to play an influential role in the acquisition and production of individual L2 segments within the syllable. Universal factors, on the other hand, seem to have a greater impact on the acquisition of complex syllable types. Studies have shown, for example, that longer, more marked, syllables (e.g., syllables with complex onsets and/or codas) are generally acquired later than shorter, less marked, syllables. Moreover, the acquisition of longer syllables may imply acquisition of shorter ones. In addition, learners tend to produce shorter syllables more accurately than longer ones, and they modify longer syllables in favor of less marked syllable types. Learners also tend to delete consonants in structures that violate the Universal Canonical Syllable Structure (UCSS), a principle and universal tendency that requires that syllables exhibit a continuous rise in sonority from the margins to the nucleus (Carlisle 1997: 334).

The role of sonority in L2 syllables has received much attention in the literature. In addition to the UCSS, several studies have examined the Minimal Sonority Distance (MSD) parameter (Selkirk 1982) in L2 production (e.g., Broselow & Finer 1991; Hancin-Bhatt & Bhatt 1997; Broselow, Chen & Wang 1998; Hancin-Bhatt 2000). Broselow and Finer (1991), for example, hypothesize that the MSD can be used to determine difficulties in L2 acquisition. Each language sets the MSD according to the minimal sonority distance that must occur between adjacent segments in a syllable onset. Consonants are assigned a value based on their relative sonority, where the more sonorous segments have a higher numerical ranking, as follows: stops (1) < fricatives (2) < nasals (3) < liquids (4) < glides (5). A language with an MSD setting of "three," therefore, will require that the adjacent segments in an onset be at least that far apart on the sonority scale; hence, an onset may consist of a stop + liquid or a stop + glide, for example, but not a stop + nasal. The higher the MSD value, the more restrictive the language will be with respect to the types of onsets allowed. On the basis of the MSD, Broselow and Finer (1991) predict that complex onsets with a smaller sonority distance will be harder to acquire than those with a greater sonority distance, and the results of their study on L1 speakers of Japanese and Korean learning L2 English support their claims. Hancin-Bhatt and Bhatt (1997), on the other hand, found that the MSD does not always accurately predict difficulty in acquisition (see also Eckman & Iverson 1993). Hancin-Bhatt and Bhatt found that L1 speakers of Spanish learning L2 English, for example, made more errors in the production of stop + glide onsets (with a sonority distance of 4 between the two segments) than in stop + liquid onsets (with a sonority distance of only 3). They postulated that the MSD alone is insufficient to account for the data and that one must also take into account L1 transfer, since, in this case, L1 Spanish allows onsets of the type stop + liquid, but not stop + glide. The transfer of L1 syllable structure constraints to L2 may override the impact of universal, or developmental, factors in L2 acquisition (Hancin-Bhatt & Bhatt 1997: 342).

A number of studies have tested other, more general, hypotheses regarding the role of universals and L1 transfer in the production of L2 syllables, including Major's Ontogeny Model and Eckman's Markedness Differential Hypothesis (both described above and in the chapters by Eckman and Major, respectively, this volume), as well as the Interlanguage Structural Conformity Hypothesis (ISCH), which posits that the universal generalizations that hold for primary languages also hold for interlanguages (Eckman 1991: 24; see also the chapter by Eckman, this volume). Carlisle (1998), for example, tested the claims of the ISCH in an investigation of the production of L2 English two- and three-member onsets by L1 Spanish learners. The results of his study showed that the subjects modified three-member onsets significantly more frequently than two-member onsets, a finding that supports the ISCH. Carlisle (1998: 344) further notes that his results supplement similar findings reported by Anderson (1987) and Eckman (1991). In another study, Major (1994) tested the Ontogeny Model through an investigation of initial and final L2 English consonant clusters produced by L1 Brazilian Portuguese learners. He examined their production of two-member onsets and codas in three separate sessions at approximately four-week intervals. He did find some support for the Ontogeny Model, in that errors due to L1 transfer (e.g., the pronunciation of English /ɹ/ as the flap /ɾ/) declined over time. In addition, the presence of developmental errors (e.g., the deletion of one member of a consonant cluster or word-final obstruent devoicing) remained steady over the course of the investigation.

Finally, Hansen (2004: 87–89) discusses research on a number of other factors that may influence L2 syllable production and help explain the variation found in production, including the preceding and following linguistic environment (e.g., Hansen 2001, 2004) and grammatical conditioning (e.g., Saunders 1987; Osburne 1996; Hansen 2004) (for a more detailed discussion of this, see Chapter 9 by Hansen Edwards, this volume). In her own study on the acquisition of L2 English codas by L1 Vietnamese speakers, Hansen (2004) found that both linguistic environment and grammatical conditioning significantly impacted L2 production. For example, the subjects were less likely to produce a coda consonant that was preceded by a diphthong or the mid back vowel /ɔ/. In addition, a pause after the coda promoted epenthesis, while a vowel or consonant after the coda disfavored epenthesis (p. 118). As for grammatical conditioning, the learners produced bimorphemic /d/ (the past tense ending) more accurately than monomorphemic /d/ in singleton codas; the absence of bimorphemic /d/ was also disfavored in two-member codas.

To summarize this section, studies of L2 syllable acquisition have focused primarily on a description and explanation of the types of errors that learners produce. Many studies have shown that L1 transfer plays an influential role in the production of L2 syllables. In addition, L2 syllables may exhibit effects of various universal constraints on syllable structure, especially with respect to longer, more complex, onsets and codas. Consequently, investigations of L2 syllables have often been used to test specific hypotheses regarding the role of universals in L2 acquisition, and as such, have important implications for more general models of L2 phonology. Finally, linguistic environment and grammatical conditioning may also impact the production of L2 syllables and may help explain variation found in production.

Prosodic domains

In addition to the syllable, larger prosodic domains – such as the word, phonological or syntactic phrase, and intonational phrase (see Nespor & Vogel 1986) – may also play a role in L2 speech. Young-Scholten (1994:203) points out that the acquisition of an L2 allophonic rule, such as the flapping of intervocalic English /t, d/ (as in *latter* and *ladder*), involves not only the acquisition of the required allophonic variant (a flap) and context (between vowels), but also the prosodic domain within which the rule applies. English flapping, for example, may occur both within the word and across word boundaries. The prosodic domain for flapping, therefore, is necessarily larger than the word itself. In fact, Young-Scholten reports that the domain for intervocalic flapping of English /t, d/ is the intonational phrase – that is, flapping will occur whenever the necessary context is met with an intonational grouping (spoken without pause) and will thus include all prosodic domains smaller than the intonational phrase as well. This is important, because languages may exhibit similar phonetic alternations but differ with regard to the prosodic domain within which they occur.

Young-Scholten (1994) proposes the Asymmetry Hypothesis to predict success in L2 phonological acquisition when differences of prosodic domain and rule application between the L1 and L2 occur. Consider first a situation in which the L1 and L2 share a similar phonetic alternation with the exception that the alternation occurs within a smaller prosodic domain in the L1 (e.g., word level) than the L2 (e.g., intonational phrase). In this case, Young-Scholten predicts that acquisition will be possible, because the learner will have positive evidence for the appearance of the alternation in larger prosodic domains. Consider now the opposite situation, however, whereby the domain for a given alternation is larger in the L1 (e.g., intonational phrase) than the L2 (e.g., word). In this case, Young-Scholten argues that the learner will need negative evidence in order to acquire the proper domain setting for the L2 alternation and that in this case acquisition will not be possible. Vogel (1991) also proposes that prosodic information may transfer from L1 to L2, just like other aspects of the learner's L1 phonology, and moreover, that prosodic structure may be more susceptible to transfer because of its abstract nature (p. 55).

Young-Scholten (1994) provides tentative support for the Asymmetry Hypothesis on the basis of L1 German speakers' acquisition of the L2 English intervocalic flapping rule, as well as the acquisition of certain L2 German pronominal forms by L1 speakers of English. Zampini (1998a) also found that L1 English learners of Spanish appear to acquire the L2 spirantized phones [β ð ɣ], which alternate with the corresponding stops [b d g], in stages that follow the prosodic hierarchy – that is, they are more likely to spirantize the stops in smaller prosodic domains than larger ones, and she suggests that the L2 acquisition of a phonetic alternation within a given domain may imply acquisition in lower level domains as well (see also James 1987).

Stress

In addition to learning to produce L2 segments accurately, L2 learners must master the stress patterns of the L2 in question. A number of studies have addressed this issue with regard to the acquisition of primary word stress in L2. As with other studies of L2 speech, researchers tend to examine the types of errors made in the placement of primary stress and analyze the source of those errors, whether they be due to L1 transfer or other effects. Archibald (1993, 1994, and references therein) has contributed a number of influential studies that examine L2 stress within the generative framework of metrical phonology. Metrical phonology posits a number of universal parameters to account for the possible stress patterns of the world's languages. While the parameters are universal, the particular setting for each parameter is language-specific. For example, syllables are grouped into larger prosodic structures called feet, and languages may vary with respect to the size of those feet – they may be binary (two syllables per metrical foot) or unbounded (allowing more than two syllables). In addition, binary feet may be either trochaic (left-headed, meaning stress placement occurs on the syllable to the left) or iambic (right-headed). Feet are then grouped into a word-level prosodic structure, which may also be either left-headed or right-headed. The syllable that constitutes the head of a word-level structure will be the syllable with primary word stress.

Archibald (1994) posits that one can examine the differences in the metrical parameter settings between the L1 and L2 and from them infer potential areas of transfer (p. 221). He reports on his research on the pronunciation of L2 English primary word stress by L1 speakers of Hungarian, Polish, and Spanish and argues that the data reveal that the learners' interlanguages reflect a combination of principles of Universal Grammar (in that the learners do not violate universal principles of metrical theory), the correct L2 parameter settings for stress placement (from resetting the L1 setting to the required L2 setting), and the transfer of L1 parameter settings to L2, resulting in incorrect stress placement (Archibald 1994: 230). In similar work, Pater (1997) argues that L1 speakers of French learning L2 English may also "misset" L2 parameter settings during acquisition; that is, their production of primary word stress suggest a parameter setting that is unlike either L1 French or L2 English. Archibald (1997), on the other hand, found that L1 speakers of nonaccentual languages, such as Chinese (a tone language) and Japanese (a pitch-accent language), do not appear to assign L2 stress via the application of metrical parameters. Rather, Archibald found that the L1 speakers of Chinese and Japanese appear to treat L2 English stress as a lexical phenomenon and memorize the position of stress for each individual word.

Finally, a few studies have examined the production of L2 primary word stress from a different standpoint. Guion, Harada, and Clark (2004), for example, tested the influence of syllable structure, lexical class (nouns vs. verbs), and the stress placement of phonologically similar words on the placement of stress in English nonsense words by native speakers of English, early Spanish-English bilinguals, and late Spanish-English bilinguals. They argue (p. 208) that one of the weaknesses of studies conducted within a metrical framework, like those discussed above, is that they assume that the metrical parameters and algorithms for stress placement are at work in native speakers, but do not provide empirical support for this assumption. They further argue that the learner behavior with respect to L2 stress placement could also be explained by factors other than the resetting or missetting of metrical parameters. In their study, Guion et al. examined the placement of primary word stress in L2 English nonsense words by L1 speakers of Spanish, along with a control group of native English speakers. They found that lexical class had an independent effect on stress placement for all three subject groups. The authors argue that these results indicate that "...both early and late bilinguals are able to employ relatively simple distributional patterns such as the statistical distribution of stress across nouns and verbs" (p. 223). The stress placement of phonologically similar words also had an effect on stress placement in nonsense words for all three subject groups, although the late bilinguals appeared to rely on this factor more heavily than the other two groups (Guion et al., p. 224). Lastly, they found that syllable structure had an independent effect on native English speakers' placement of stress in nonsense words, in that long vowels tended to attract stress more often that short vowels, and syllables with complex codas tended to attract stress more than syllables with simple codas. Both bilingual groups, however, had different outcomes. The early bilinguals showed similar results to the native speakers, but they did not show an effect of vowel length on stress placement. As for the late bilinguals, syllable structure did not prove to be a significant predictor of stress placement (p. 224). Results such as these present an intriguing outlook on the complexity of L2 stress acquisition, and more work is needed to more fully understand the processes by which learners master the patterns of L2 stress.

Methodological approaches to studying L2 speech production

The most common methodological approach to studying L2 speech entails the elicitation and recording of L2 speech, along with a subsequent error description and analysis. Within this general approach, however, there are number of methodological options. First, the recording of the data may take place within or outside a laboratory setting, depending upon the type of data analysis to be done (e.g., acoustic measurements of particular aspects of L2 speech vs. a broad phonetic transcription). In order to examine detailed phonetic or acoustic characteristics of speech, for example, it is advisable to record the speech sample in a laboratory setting that has the sophisticated equipment and sound attenuation necessary to achieve a robust and clean speech signal; this facilitates acoustic measurement and enhances the overall integrity of the data and reliability of the results. When the aim is to examine characteristics of speech that do not require a detailed phonetic analysis, on the other hand, the researcher may find it sufficient, or perhaps preferable, to record their subjects in a quiet, but more informal setting.

One of the primary drawbacks to laboratory-elicited speech is that it is rarely naturalistic. Indeed, some studies elicit the production of isolated syllables, words, or short phrases out of context, and the results of such studies may not be generalizable to speech in a natural setting. Some even question whether or not conversations conducted in a laboratory setting can be considered truly spontaneous or naturalistic. A setting in which subjects are acutely aware of being recorded may cause them to try to articulate "better" or more clearly; in other cases, it may raise anxiety levels, especially for novice learners who may feel uncertain and reticent with respect to their L2 speaking abilities.

As with all studies of L2 acquisition, the researcher must also carefully consider the subject pool to be examined and attempt to control for a number of potentially confounding variables. Many researchers employ pre-testing questionnaires to gather certain demographic, biographical, and experiential information of potential subjects. Given the inherent variability of speech and of L2 speech learning, researchers such as Flege (1987a) stress the importance of having subject groups that are as homogeneous as possible. Grosjean (1998: 132) also notes that studying bilinguals can prove particularly problematic, since bilingual populations tend to vary greatly as a function of fluency, quality and quantity of L2 input, reason for acquiring L2, and the need for maintaining particular language skills. He argues that studies of bilingual speech should recruit subjects that use both languages on a daily basis in order to ensure frequent use of both the L1 and L2 and to reduce the possibility of language attrition.

In addition, it is important to employ native speaker controls of both the L1 and L2 when examining learners' or bilinguals' speech. While many studies employ native speakers of the L2 as a control group, fewer employ native speakers of the learners' L1 as a control. However, one cannot assume that a learner's L1 speech will be like that of a monolingual speaker, because the acquisition of L2 may impact the L1 (Flege 1987a, 1987b; Major 1992). Another consideration involves the manipulation of *language mode*. Grosjean (1998: 136) proposes that bilinguals operate along a continuum ranging from a monolingual mode to a bilingual mode. At the monolingual end of the continuum, the bilingual interacts with monolingual speakers and has only one language activated. In bilingual mode, the speaker interacts with other bilingual speakers, so that both languages are activated (although one will still be used as the primary, or base, language), and one may find examples of codeswitching or borrowing. Bilingual speakers vary along this continuum as a function of the situation, person to whom they are speaking or listening, topic, and purpose of the conversation. The degree to which bilinguals are placed into monolingual or bilingual mode may therefore affect their language processing. For example, if bilingual subjects think that both languages are of interest in a particular study, they may keep both activated throughout the experimental tasks. Unfortunately, many published studies of bilingual speech production have made their subjects aware of an interest in bilingualism and, in doing so, may have compromised language mode (see Magloire & Green 1999 for techniques on controlling language mode). Learners (or non-fluent bilinguals) may also vary according to language mode, and researchers should take such factors into account when designing studies of L2 speech production. Lastly, given the variability of L2 speech, it is essential to collect a speech sample with several elicitations of the target phone(s) under study in order to maximize the chance that observable trends and significant results will emerge in the data. In similar fashion, it is necessary to take into account factors that may increase acoustic and phonetic variability, such as the linguistic environment surrounding the target phone(s) of the study.

With regard to the analysis of the L2 speech corpus, a variety of methods have been employed as well. As seen in the discussion on the acoustic characteristics of L2 speech above, many researchers measure the relevant acoustic components of the target phones in question and compare the learner/bilingual mean values to those of monolingual speakers from the L2 and/or L1. The researcher may also compare the learners/bilinguals' mean L1 and L2 values to determine the extent to which such speakers differentiate the two languages. While some interpretation of the acoustic data may be necessary, and baseline criteria for measuring acoustic cues must be established, this can be a relatively objective way to examine the characteristics of L2 speech. There are also a number of readily available computer programs that make acoustic analysis a viable option. However, some researchers, such as Leather (1999: 32) caution that "...large-scale acoustic and statistical analyses may yield results that are all the more difficult to interpret..." especially because of the difficulty in generalizing laboratory speech to spontaneous speech. Acoustic analyses can nevertheless furnish researchers with important information regarding conceivable parameters for the characteristics of natural speech and elucidate likely areas of differences between learners and native speakers. Such information can then be used as a basis for formulating research questions and designs for speech in a natural setting.

In other cases (e.g., in many of the studies on sound substitutions or syllable structure mentioned above), researchers may transcribe the data as a means of determining the number and kinds of allophonic or phonemic errors in L2 speech. The primary drawback to data transcription is that it tends to be more subjective than the computer-based acoustic analyses described above. Researchers may increase the reliability of transcribed data, however, by having the corpus, or a subset of it, transcribed independently by two or more trained individuals and comparing them for discrepancies. Few discrepancies should be found in the transcriptions; in addition, the discrepancies may be resolved through consultation with the transcribers in some cases, or the researcher may decide to eliminate them altogether from the analyses.

A third method that has been used to determine the "nativeness" of L2 speech production involves native speaker judgments regarding the overall (global) foreign-accentedness, comprehensibility and/or intelligibility of L2 speech (e.g., Flege, Munro, & MacKay 1995; Munro & Derwing 1995; Riney & Flege 1998; Munro 1998; Piske, MacKay & Flege 2001) (this method is also discussed in length in Chapter 7 by Munro, this volume). Native speakers are asked to judge the L2 speech, usually on an equal-appearing interval scale. Piske et al. (2001: 194) report that a 5-point scale is most commonly used, although other scales reported in the literature have ranged from three to nine points. Global accentedness ratings are an important measure of L2 speech, since, as Munro (1998) points out, "any comprehensive account of human speech perception must take into consideration the fact that listeners are able to understand. . .speech that deviates notably from typical native-speaker utterances" (p. 139). In addition, Munro notes that several studies have shown that listeners can reliably judge degrees of foreign accent that correlate with other more objective measures, such as error counts. For more detailed discussion of these issues, see the chapter by Munro in this volume.

Implications of L2 speech research

The studies presented in the preceding sections reveal some of the complexities and challenges involved in the production of L2 speech sounds. While much work remains before an adequate and comprehensive model of interlanguage phonology can be put forward, important contributions have been made. Proposals regarding the production of L2 speech sounds, such as Flege's Speech Learning Model, Major's Ontogeny Model, Eckman's Markedness Differential Hypothesis and the

Interlanguage Structural Conformity Hypothesis, have influenced the way that many researchers approach the study of L2 speech.

The findings from studies of L2 speech also have important implications for related fields of language study. Models of bilingual speech perception, for example, have much to gain from an understanding of the ways in which learners produce L2 speech sounds; indeed, such models will be incomplete unless they can account for the relationship between L2 speech perception and production (see also the chapter by Strange & Shafer, this volume). Similarly, studies of L2 speech are important for psycholinguistic models of language processing and linguistic representation, including models of word recognition and lexical access (see, for example, Grosjean & Soares 1986; Fitzpatrick & Wheeldon 2000).

Likewise, studies of L2 speech can, and should, inform models of formal phonology. Any model of phonological competence will necessarily entail the competence of the multilingual speaker and L2 learner. L2 speech data can be used to examine the validity and adequacy of the tenets of particular phonological theories and may prove useful in evaluating competing proposals. In similar fashion, theoretical phonology can provide a framework for examining interlanguage phonology. As seen above, this has been applied most often in the examination of L2 stress patterns, as well as particular sound or feature substitutions that take place in L2 speech. More recently, Optimality Theory has begun to be used to examine other aspects of L2 speech, including syllable structure and feature acquisition.

Lastly, the findings on L2 speech production and acquisition have important implications for the L2 classroom. Knowledge regarding difficulties in the acquisition of particular L2 speech sounds, as well as the identification of factors that may affect L2 speech production, can be used to help practitioners develop appropriate training techniques and pedagogical materials such that they optimize their students' chances for improvement and success in learning to produce a second language more accurately.

Oustanding issues and future directions

Outstanding issues

Along with recent advances in studies that examine L2 speech, there are a number of outstanding issues that provide continued opportunities for research. First, a number of recent studies have examined both the L2 production and perception of speech sounds; indeed, it is sometimes difficult to discuss issues of L2 speech production without also going into issues of L2 perception. One of the overriding questions concerns whether or not accurate, native-like perception of a particular

L2 contrast is necessary for accurate production of the same contrast. In addition, little is known about how the acquisition of perceptual and production capabilities progress over time, and in what ways they influence each other as they change. Any adequate model of L2 speech production must therefore prove compatible with a corresponding model of L2 speech perception (see also the chapter by Strange and Shafer in this volume).

Second, the inherent variability found in speech can make the study of L2 speech sounds more difficult, since it is necessary to determine the parameters by which speakers classify a particular instantiation of a sound as being a member of a given phonetic category. As discussed above, L2 learners may equate particular L1 and L2 sounds as being instantiations of the same category, thus inhibiting accurate L2 production. In the same vein, however, it is also important to ascertain the nature of the categories established for the L2, including the type of variability permitted and the extent to which such variability reflects native speaker norms.

In addition to the inherent variability of speech sounds, there are a number of external factors that may influence L2 speech production and acquisition (e.g., age of acquisition, motivation, personality and identity, amount of L2 input, amount of L1 use, linguistic and task variation, lexical familiarity, speech register, etc.) and it is important that researchers continue to examine those variables when conducting research on L2 speech. With regard to speech register, more research is needed on L2 speech production in casual speech. Most L2 speech production experiments are carried out in a laboratory setting, where the subjects may produce isolated syllables, words, or sentences. There are few studies of the L2 production of sounds in continuous speech, whether in a casual, conversational format or a more formal reading style.

There are also a number of areas with respect to L2 speech production that have not received adequate attention in the literature. For example (and as mentioned above), many studies of L2 speech production focus on bilinguals, as opposed to learners who are still in the process of acquiring the L2. There is a need, therefore, for studies that examine L2 speech at the beginning and intermediate stages of acquisition, as well as a need for longitudinal studies that examine changes in L2 speech production over time. In addition, most published studies focus on L2 English; relatively few studies examine the production and acquisition of L2 speech sounds in other languages. Lastly, more research is needed that explicitly addresses the implications and potential impact of the theoretical findings, such as those discussed here, for the L2 classroom. By the same token, pedagogical techniques and materials should reflect current knowledge regarding the way in which L2 learners produce and process L2 speech sounds and will need to evolve appropriately so as to optimize the potential for advancement of a native-like L2 pronunciation.

Future directions

While much meaningful work will undoubtedly continue on the topics discussed in this chapter, an examination of current work in L2 speech reveals two particularly important and exciting directions in which the field appears to be evolving. First, technological advances and the accessibility of computer programs for acoustic analysis bode well for researchers interested in the subphonemic, acoustic, gestural, and physiological properties of L2 speech. Advances in computer technology and speech analysis programs continue to facilitate data collection and analysis. In addition, neuroimaging techniques, such as fMRI (see Fiez 2001, as well as the chapters by Strange & Shafer, this volume), as well as the use of ultrasound (Gick et al., this volume) allow researchers to examine physiological aspects of speech in new ways and are beginning to shed light on components of speech that were not directly observable before.

Second, the future will likely witness a continued increase in studies of L2 speech production that look beyond a focus on the speech signal itself to include a consideration of the relationship between L2 speech and other domains, such as perception, lexical access and language processing, cognition, and the neurophysiology of speech. The relevance of the studies discussed in this chapter for other aspects of L2 acquisition and related fields of enquiry have been made throughout; and while this chapter has attempted to focus primarily on the nature of L2 speech sounds, several studies in the literature address more general issues of L2 acquisition and/or speech production. Indeed, such connections must be examined and validated if we are to arrive at a more thorough understanding of the processes by which L2 learners acquire, process, and produce L2 speech sounds.

Suggested readings

Archibald, J. A. 1994. A formal model of learning L2 prosodic phonology. *Second Language Research* 10: 215–240.

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, Baltimore: York Press, 233–277.

Flege, J. E., Bohn, O-S., & Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25: 437–470.

Grosjean, F. 1998. Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1: 131–149.

Guion, S. G., Harada, T., & Clark, J. J. 2004. Early and late Spanish-English bilinguals' acquisition of English word stress patterns. *Bilingualism: Language and Cognition* 7: 207–226.

Hansen, J. G. 2004. Developmental sequences in the acquisition of English L2 syllable codas. *Studies in Second Language Acquisition* 26: 85–124.

References

Altenberg, E., & Vago, R. M. 1987. Theoretical implications of an error analysis of second language phonology production. In G. Ioup & S. H. Weinberger (eds.), *Interlanguage phonology: The acquisition of a second language sound system*, Cambridge, MA: Newbury, 148–164.

Anderson, J. I. 1987. The Markedness Differential Hypothesis and syllable structure difficulty. In G. Ioup & S. H. Weinberger (eds.), *Interlanguage phonology: The acquisition of a second language sound system*, Cambridge, MA: Newbury, 279–291.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. 2004. Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32: 233–250.

Archibald, J. A. 1993. *Language learnability and L2 phonology: The acquisition of metrical parameters*. Boston: Kluwer Academic Publishers.

Archibald, J. A. 1994. A formal model of learning L2 prosodic phonology. *Second Language Research* 10: 215–240.

Archibald, J. A. 1997. The acquisition of English stress by speakers of nonaccentual languages: Lexical storage versus computation of stress. *Linguistics* 35: 167–181.

Bohn, O.-S., & Flege, J. E. 1992. The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition* 14: 131–158.

Bradlow, A. R. 2008. Training non-native language sound patterns: Lessons from training Japanese adults on the English /ɹ/ – /l/ contrast. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Broselow, E., & Finer, D. 1991. Parameter setting in second language phonology and syntax. *Second Language Research* 7: 35–59.

Broselow, E., Chen, S., & Wang, C. 1998. The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition* 20: 261–280.

Carlisle, R. S. 1997. The modification of onsets in a markedness relationship: Testing the Interlanguage Structural Conformity Hypothesis. *Language Learning* 47: 327–361.

Carlisle, R. S. 1998. The acquisition of onsets in a markedness relationship. *Studies in Second Language Acquisition* 20: 245–260.

Eckman, F. R. 1977. Markedness and the Contrastive Analysis Hypothesis. *Language Learning* 27: 315–330.

Eckman, F. R. 1991. The Structural Conformity Hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition* 13: 23–41.

Eckman, F. R. 2008. Typological markedness and second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Eckman, F. R., & Iverson, G. K. 1993. Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research* 9: 234–252.

Fiez, J. A. 2001. Neuroimaging studies of speech: An overview of techniques and methodological approaches. *Journal of Communication Disorders* 34: 445–454.

Fitzpatrick, J., & Wheeldon, L. R. 2000. Phonology & phonetics in psycholinguistic models of speech perception. In N. Burton-Roberts, P. Carr, & G. Docherty (eds.), *Phonological knowledge: Conceptual and empirical issues*, Oxford: Oxford University Press, 131–160.

Flege, J. E. 1981. The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly* 15: 443–455.

Flege, J. E. 1987a. The instrumental study of L2 speech production: Some methodological considerations. *Language Learning* 37: 285–296.

Flege, J. E. 1987b. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–65.

Flege, J. E. 1991. Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustic Society of America* 89: 395–411.

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, Baltimore: York Press, 233–277.

Flege, J. E., & Eefting, W. 1987. Production and perception of English stops by native Spanish speakers. *Journal of Phonetics* 15: 67–83.

Flege, J. E., & Hillenbrand, J. 1984. Limits on pronunciation accuracy in adult foreign language speech production. *Journal of the Acoustical Society of America* 76: 708–721.

Flege, J. E., Bohn, O.-S., & Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25: 437–470.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. 1995. Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 19: 90–119.

Gick, B., Bernhardt, B., Bacsfalvi, P., & Wilson, I. 2008. Ultrasound imaging applications in second language acquisition. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins, .

Green, K. P., Zampini, M. L., & Magloire, J. 1997. An examination of word-initial stop closure interval in English, Spanish and Spanish-English Bilinguals. *Journal of the Acoustical Society of America* 102: 3136.

Grosjean, F. 1989. Neurolinguists, beware: The bilingual is not two monolinguals in one person. *Brain and Language* 36: 3–15.

Grosjean, F. 1997. Processing mixed languages: Issues, findings and models. In A. de Groot & J. Kroll (eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, Mahwah, NJ: Lawrence Erlbaum, 225–254.

Grosjean, F. 1998. Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1: 131–149.

Grosjean, F., & Soares, C. 1986. Processing mixed language: Some preliminary findings. In J. Vaid (ed.), *Language processing in bilinguals: Psycholinguistic and neuropsychological perspectives*, Hillsdale, NJ: Lawrence Erlbaum, 145–179.

Guion, S. G., Flege, J. E., & Loftin, J. D. 2000. The effect of L1 use on pronunciation of Quichua-Spanish bilinguals. *Journal of Phonetics* 28: 27–42.

Guion, S. G., Harada, T., & Clark, J. J. 2004. Early and late Spanish-English bilinguals' acquisition of English word stress patterns. *Bilingualism: Language and Cognition* 7: 207–226.

Hancin-Bhatt, B. 2000. Optimality in second language phonology: Codas in Thai ESL. *Second Language Research* 16: 201–232.

Hancin-Bhatt, B. 2008. Second language phonology in optimality theory. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Hancin-Bhatt, B., & Bhatt, R. M. 1997. Optimal L2 syllables: Interactions of transfer and developmental effects. *Studies in Second Language Acquisition* 19: 331–378.

Hansen, J. G. 2001. Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics* 22: 338–365.

Hansen, J. G. 2004. Developmental sequences in the acquisition of English L2 syllable codas. *Studies in Second Language Acquisition* 26: 85–124.

Hansen Edwards, J. G. 2008. Social factors and variation in production. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Hazan, V. L., & Boulakia, G. 1993. Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech* 36: 17–38.

Ioup, G. 2008. Exploring the role of age in the acquisition of a second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

James, A. R. 1987. Prosodic structure in phonological acquisition. *Second Language Research* 3: 118–140.

Jun, S.-A., & Cowie, I. 1994. Interference for 'new' and 'similar' vowels in Korean speakers of English. *Ohio State University Working Papers* 43: 117–130.

Keating, P. A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60: 286–319.

Kessinger, R. H. & Blumstein, S. 1997. Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics* 25: 143–168.

Ladefoged, P. 1982, 2001. *A course in phonetics*. New York: Harcourt Brace Jovanovich College Publishers.

Lado, R. 1957. *Linguistics across cultures*. Ann Arbor, MI: The University of Michigan Press.

Leather, J. 1999. Second-language speech research: An introduction. *Language Learning* 49: 1–56.

Lisker, L., & Abramson, A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20: 384–422.

Lombardi, L. 2003. Second language data and constraints on Manner: Explaining substitutions for the English interdentals. *Second Language Research* 19: 225–250.

Mack, M. 1989. Consonant and vowel perception and production: Early English-French bilingual and English monolinguals. *Perception and Psychophysics* 46: 187–200.

Magloire, J., & Green, K. P. 1999. A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica* 56: 158–185.

Major, R. C. 1986. The ontogeny model: Evidence from L2 acquisition of Spanish *r*. *Language Learning* 36: 453–503.

Major, R. C. 1992. Losing English as a first language. *The Modern Language Journal* 76: 190–208.

Major, R. C. 1994. Chronological and stylistic aspects of second language acquisition of consonant clusters. *Language Learning* 44: 655–680.

Major, R. C. 2001. *Foreign accent: The ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Major, R. C. 2008. Transfer in second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

McAllister, R., Flege, J. E., & Piske, T. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30: 229–258.

Miller, J. L., Green, K. P., & Reeves, A. 1986. Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica* 43: 106–115.

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

Munro, M. J. 2008. Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

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

Munro, M. J., Flege, J. E., & MacKay, I. 1996. The effects of age of second-language learning on the production English vowels. *Applied Psycholinguistics* 17: 313–334.

Nespor, M., & Vogel, I. 1986. *Prosodic phonology*. Dordrecht: Foris.

Osburne, A. G. 1996. Final cluster reduction in L2 English speech: A case study of a Vietnamese speaker. *Applied Linguistics* 17: 164–181.

Pater, J. V. 1997. Metrical parameter missetting in second language acquisition. In S. J. Hannahs & M. Young-Scholten (eds.), *Focus on phonological acquisition*, Philadelphia: John Benjamins, 235–261.

Piske, T., MacKay, I. R. A., & Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215.

Riney, T. J., & Flege, J. E. 1998. Changes over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition* 20: 213–243.

Saunders, N. 1987. Morphophonemic variations in clusters in Japanese English. *Language Learning* 37: 247–272.

Schmidt, R.W. 1987. Sociolinguistic variation and language transfer in phonology. In G. Ioup & S. H. Weinberger (eds.), *Interlanguage phonology: The acquisition of a second language sound system*, Cambridge, MA: Newbury, 365–377.

Schmidt, A. M., & Flege, J. E. 1995. Effects of speaking rate changes on native and nonnative speech production. *Phonetica* 52: 41–54.

Schmidt, A. M., & Flege, J. E. 1996. Speaking rate effects on stops produced by Spanish and English monolinguals and Spanish/English bilinguals. *Phonetica* 53: 162–179.

Selkirk, E. 1982. The syllable. In van der Hulst & Smith (eds.), *The structure of phonological representations, Part II*, Dordrecht: Foris.

Strange, W., & Shafer, V. L. 2008. Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Thornburgh, D. F., & Ryalls, J. F. 1998. Voice onset time in Spanish-English bilinguals: Early versus late learners of English. *Journal of Communication Disorders* 31: 215–229.

Vogel, I. 1991. Prosodic phonology: Second language acquisition data as evidence in theoretical phonology. In T. Huebner & C. A. Ferguson (eds.), *Crosscurrents in second language acquisition and linguistic theories*, Philadelphia: John Benjamins, 47–65.

Volaitis, L., & Miller, J. L. 1992. Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America* 92: 723–735.

Weinberger, S. H. 1990. Minimal segments in second language phonology. In J. Leather & A. James (eds.), *New Sounds 90: Proceedings of the 1990 Amsterdam Symposium on the Acquisition of Second Language Speech*, Amsterdam: University of Amsterdam, 263–311.

Williams, L. 1977. The perception of stop consonant voicing by Spanish-English bilinguals. *Perception and Psychophysics* 21: 289–297.

Young-Scholten, M. 1994. On positive evidence and ultimate attainment in L2 phonology. *Second Language Research* 10: 193–214.

Zampini, M. L. 1998a. L2 Spanish spirantization: A prosodic analysis and pedagogical implications. *Hispanic Linguistics* 9: 154–188.

Zampini, M. L. 1998b. The relationship between the production and perception of L2 Spanish stops. *Texas Papers in Foreign Language Education* 3: 85–100.

Zampini, M. L., & Green, K. P. 2001. The voicing contrast in English and Spanish: The relationship between production and perception. In J. Nicol (ed.), *Two languages, one mind: Bilingual language processing*, Cambridge, MA: Blackwell, 23–48.

CHAPTER 9

Social factors and variation in production in L2 phonology

Jette G. Hansen Edwards
The Chinese University of Hong Kong

Introduction

This chapter focuses on two domains of research in the acquisition of L2 phonology: the effect of social factors on L2 phonology and variation in production in L2 phonology. The discussion of social factors focuses on gender, extent of L1 and L2 use, social identity, and target language variety while the discussion of variation focuses on interlocutor/speech accommodation, attention to speech/monitoring, and the effects of linguistic and social factors on production.

The research on social factors and variation is unified in the underlying theoretical framework that learners are active agents in their language use, language choices, and targets for acquisition. That is, they are not passive recipients of the target language, and variation in production is typically systematic and may be due, in part, to social marking due to gender, identity, accommodation to the interactant, and the linguistic environment, etc. As a result, differences between the target language and the language of the learner may not necessarily be errors, but may be evidence of users targeting a particular variety that is not necessarily the standard or marking their identity by using a certain variant in a specific situation with particular interactants. In other words, as Dowd, Zuengler, and Berkowitz (1990) state, performance in the L2 may be socially conditioned. This research raises issues of whether a 'deviation' from the standard target language is a lack of acquisition or social marking and of the learner's knowledge about the language and use of the knowledge to construct L2 identity. These issues will be explored in this chapter.

The structure of the chapter is as follows: First, a review of the research on social factors and variation will be presented. As theoretical frameworks vary based on the focus of research, these frameworks will be discussed under each topic area. Methodological options are also discussed briefly within each section and then synthesized in a separate section, which follows the literature review. Finally, a synthesis of the major findings on social factors and variation and suggestions for future research are presented.

Literature review

The literature review will first examine research on social factors and then research on variation, with each major topic covered in a separate section. The review will focus predominantly on recent findings, with these findings discussed in light of past research, especially in reference to classic and fundamental studies in each area.

Gender

Gender has long received attention from L2 phonology researchers. Early research (cf. Asher & Garcia 1969; Elliot 1995; Flege & Fletcher 1992; Olson & Samuels 1973; Snow & Hoefhagle-Höhle 1977; Suter 1976; Purcell & Suter 1980; Tahta, Wood, & Loewenthal 1981; Thompson 1991) defined gender biologically as 'sex', and focused on pronunciation accuracy, employing experimental data elicitation techniques such as word lists or reading passages that would be rated for accuracy and/or accent. In these studies, gender was one of a number of predictor variables, along with length of residence and age of arrival, among others.

Overall, these studies did not show gender to be a strong predictor of pronunciation accuracy; in fact, in a recent review of research on accent, Piske et al. (2001) concluded "the results obtained for gender do not lead to any strong conclusions" (p. 200). Additionally, early research on gender has been criticized on both theoretical and methodological grounds: theoretically, for confusing gender and sex and for the tendency to "exaggerate and overgeneralize differences between women and men, in addition to ignoring the social, cultural, and situational forces that shape gender categories and gender relations" (Ehrlich 1997: 426). Methodologically, the research has been criticized for employing one-time data collection techniques in which gender is conceptualized as fixed and unchanging.

Sociolinguistic research has also defined gender as a stable construct. One such study is Adamson and Regan's (1991) research on the acquisition of the {-ing} variable by Vietnamese and Cambodia immigrants to the US. As the prestige variant of {-ing}, which is [iŋ], was present in the learners' L1, the researchers wanted to investigate the learners' use of the variant [ɪn] for {-ing}, the greater the use thereof the researchers hypothesized indicated a greater integration into the L2 speech community. Native speaker controls were also employed in the research. Results indicated differences between men and women in the use of the variants:

women had a greater use of the prestige variant [iŋ] for {-ing} while men used [ɪn] more. These results were found for both the native speakers and the L2 learners in the study, leading the researchers to conclude that the men and women L2 learners were aiming for different targets, with the women targeting the variants used by the women native speakers, while the men were targeting the variant employed by the men native speakers.

Recent research (e.g., Hansen 2006; Ohara 2001) has recognized that gender is "something individuals *do* as opposed to something individuals *are* or *have*" (Ehrlich 1997: 422). These studies have employed poststructuralist theoretical frameworks (cf. Pavlenko & Piller 2001) and ethnographic and discourse-based methodologies and "...show that possibilities for comprehensible input, comprehensible output, and positive attitudes towards the target language and culture ...are determined almost exclusively by the social context of the learning environment" (Ehrlich 1997: 440).

One study that examines the social construction of a gendered identity in the L2 is Ohara (2001). In Japanese, femininity is expressed by women through the use of a high-pitched voice and "...the use of a high pitched voice is an important way of performing or 'doing' gender" (p. 234). Ohara's research examined the extent to which L2 learners of Japanese were aware of these norms and were willing to perform Japanese gender. Employing three groups of participants–five L1 American English beginning learners of Japanese; five Japanese-English bilinguals with L1 Japanese and L2 English; and five English-Japanese bilinguals with L1 English and L2 Japanese–Ohara had them perform three tasks in both English and Japanese: read isolated sentences, perform a scripted conversation with the researcher, and produce a telephone message to a professor and to a friend. The fundamental frequency (i.e. pitch) of each person's voice was then measured across the three tasks; additionally, ethnographic interviews were conducted to determine the participants' awareness of voice pitch levels in their own Japanese.

The results of the linguistic analysis found that for beginning learners, there were no significant differences in pitch between English and Japanese for the conversation and reading tasks but that there was a difference in English, and not Japanese, pitch in the telephone task. Japanese (L1) – English (L2) bilinguals had a higher pitch in Japanese across all tasks with the highest pitch to professors. Finally, for English (L1) – Japanese (L2) bilinguals the results were mixed, as two participants had similar results to beginning learners and the other three were closer to Japanese-English bilinguals.

Ohara (2001) found that the beginning learners did not have the knowledge of symbolic uses of pitch in Japanese and therefore did not vary their pitch in English and Japanese. All the bilinguals, however, were aware of the use of pitch to signal gender in Japanese. In terms of the mixed findings for the English – Japanese bilinguals, the interviews found that "it became apparent for these women that the voice pitch levels they employed correlated neatly with their attitude toward the kinds of images typically associated with Japanese women" (Ohara 2001: 242). The bilinguals who did vary pitch patterns did so in attempt to "fit into the culture" (p. 243) rather than because they were enamored of being viewed as feminine and cute; they were trying to project a Japanese identity and using pitch/voice as one way to do this. As for the two English (L1) – Japanese (L2) bilinguals who did not vary their pitch levels although they were aware of the need to do so in some social circumstances, Ohara found that they made a conscious choice not to vary their pitch as they felt it projected an identity they did not want accept (note: while not in the area of phonology, work by Siegal (1996) has found similar findings for white women studying Japanese in Japan).

Work by Hansen (2006) on Vietnamese learners of English examines how a husband and wife, recent immigrants from Vietnam to the US, had gendered access to L2 development through work places, how the participants reacted to differing types and levels of access to L1 and L2 use, as well as established and maintained this access within the family. The study also examined how these differential levels of access to L2 use impacted the participants' acquisition of English, in this case syllable final consonants and consonant clusters. Phonological data were gathered from naturalistic interviews three times during the space of one year; additionally, interview and observation data were collected for two years.

The study found that the work roles the husband, Nhi, and the wife, Anh, were able to fill were based on the constraints of both the L1 and the L2 culture. For Anh, the most viable work place – linguistically, as it required little English for training and work, and financially, as there were many jobs available – was the nail salon due to the help and support of her extensive network of Vietnamese women nail technicians. Nhi found a job more acceptable for men – an order filler in a golf factory. The workplaces offered differential opportunities for L2 use: On the surface, Anh, appeared to have more opportunities for L2 use as she needed to use English during the entire workday, while her husband had little chance to speak English during the day since his job required little interaction with other individuals. However, in reality, Anh's English language use was highly repetitive and formulaic as she only conversed briefly with her clients, many of whom were recent immigrants from Mexico and spoke very limited English. Nhi, on the other hand, had fairly limited opportunities to practice English if measured time-wise – his only chances were during short breaks and his lunch hour. However, he had a supportive English use environment at work, with four good friends at work, two American and two Mexican men. As he stated, "they teach English...if I if I speak wrong they correct for me."

The analysis of the linguistic data indicated that Anh's limited access to L2 use opportunities may have affected her acquisition of English since her production of English syllable codas was statistically significantly less accurate than Nhi's across time. Nhi also had a greater accuracy in production of CC codas and appeared to be exposed to more complex coda structures as evidenced by his greater attempts at CCC codas. This is not surprising given the greater opportunities for interaction and correction that Nhi had in comparison with Anh. In contrast, Anh had difficulty communicating with her clients, who were often non-native speakers of English, making it more difficult for her to receive the opportunities for complex language use that may aid second language acquisition.

In summary, when gender is framed and investigated as a biological construct, it does not seem to be a significant factor in L2 pronunciation accuracy. However, when gender is framed and investigated as a social construct, it does appear to impact the level of access learners have to L2 use opportunities and therefore the ability to get L2 input and negotiate meaning, which appear to affect L2 development. Finally, the perception of and willingness to adopt gender roles also appears to affect L2 production.

Extent of L1/L2 use

While early research in this area (cf. Flege & Fletcher 1992; Purcell & Suter 1980; Suter 1976; Thompson 1991) has examined the effect of the amount of L2 use on L2 pronunciation accuracy, later research studies, the majority of which were conducted by Flege and his colleagues (e.g., Flege, Frieda, & Nozawa 1997; Guion, Flege, & Loftin 2000; Piske & MacKay 1999; Piske, MacKay, & Flege 2001), have examined the effect of L1 use on L2 production (see also Chapter 2 by Ioup and Chapter 13 by Derwing, both in this volume, for related discussion). Work in this area has largely been experimental in design and employed accent ratings on words and sentences and self-reports of L1 and L2 use.

The results of the early studies indicate that amount of L2 use may not significantly affect L2 accent: While Suter (1976) found that amount of L2 conversation at work and/or school was the third best predictor of pronunciation accuracy (after native language and level of speaker's concern about her/his pronunciation), a reanalysis of this data by Purcell and Suter (1980) found that L2 use was no longer significant. Additionally, research by Thompson (1991) and Flege and Fletcher (1992) found no significant effects of L2 use.

An exception to these findings is a study by Moyer (2004) on L2 learners of German. Moyer's study focused on twenty-five immigrants to Berlin, all advanced speakers of German with varying lengths of stay and ages of arrival. Moyer found that the frequency with which the participants had spoken interaction in German with native speakers was significantly correlated with ratings of the participants' nativeness by native speakers of German. As Moyer notes, "...how effectively and consistently the learner utilizes available linguistic resources may be a deciding factor in *constraints* on attainment" (p. 98). Contact with native speakers of German –

and the resulting spoken interaction based on this contact – was also viewed by the participants themselves as a critical element to their L2 success: "…many participants say that personal contact has been the most effective and important aspect of their experience in-country for developing near-native fluency" (p. 103). The research also indicated that age and extent of L2 use may be connected as younger immigrants may have an easier time establishing and preserving native speaker friendships and contacts than older immigrants, and therefore have greater access to L2 spoken interaction:

> …contact must ultimately be welcome on both sides, and maintaining such connections may become more difficult as one gets older – a phenomenon several participants confirm. Maturation can thus be seen related to social adaptation, in mutually constitutive ways, impacting access to quality linguistic input.
>
> (Moyer 2004: 101)

Interesting, Moyer (2004) also found that the participants in her study avoided interacting with speakers of their L1 in order to develop their L2 and attain cultural assimilation, indicating that L2 learners may actively employ L1 use avoidance as a L2 linguistic and cultural acquisition strategy.

The majority of recent research (e.g., Flege, Frieda, & Nozawa 1997; Guion, Flege, & Loftin 2000; Piske & MacKay 1999; Piske, MacKay, & Flege 2001) on extent of L1/L2 use has shifted to examining the effect of L1 use on L2 accent. In their research on native speakers of Italian who immigrated to Canada, Flege, Frieda and Nozawa (1997) found that while both high and low users of Italian were rated as having a detectable foreign accent, the participants who seldom spoke Italian had a significantly lesser foreign accent in English than those who spoke Italian more often. In a replication study, Piske and MacKay (1999) added the variable of early versus late bilingual, and found that regardless of whether the participant was an early or late bilingual, the group with higher L1 accent ratings had a higher use of the L1. Piske, MacKay, and Flege (2001) also conducted a study on Italian (L1) – English (L2) bilinguals and found that while L1 use was a significant indicator of accent in the L2 for both early and late bilinguals, late bilinguals had a stronger accent overall, with age of arrival having a stronger effect on L2 accent than L1 use.

Building on previous research, Guion, Flege, and Loftin (2000) examined the effect of L1 use on both L2 and L1 production on Quichua (L1) – Spanish (L2) bilinguals in Ecuador and found that individuals with high Quichua use had the strongest accent in the Spanish and that the majority of speakers with low Quichua use received native-like accent ratings in Spanish. In a follow-up experiment, the researchers examined whether a Spanish accent could be detected in Quichua by examining two groups of Quichua speakers – those who had acquired Quichua as infants and those who acquired it 'late' (e.g., after age 15). Results show that late learners of Quichua had more of an accent than early learners, which the researchers state indicate that Spanish accent in Quichua is not result of L1 use but of age of acquisition of Quichua.

Guion et al. (2000) conducted a second study with Korean (L1) – English (L2) bilinguals and found that accent in L1 and L2 were inversely correlated: "... the subjects who had a relatively good pronunciation of English (mostly early bilinguals) tended to have poor pronunciation of Korean, whereas those who had a poor pronunciation of English (mostly late bilinguals) tended to have a good pronunciation of Korean" (p. 36–37). While they found that the low L1 use group had a significantly lesser accent in the L2 than the high L1 group did, the two groups did not differ in terms of L1 accent, indicating that L1 use did not affect L1 accent.

While Guion et al. (2000) explain their finding through the single system hypothesis (Flege 1995), they also note that, "Another plausible explanation for the asymmetrical effect of L1 use on L1 and L2 might be the greater importance of L1 production for social identity. The appearance of a Spanish accent in Quichua might well threaten individuals' identity as Quichua speakers and community members in ways which are quite different from the consequences of a Quichua accent in Spanish" (p. 40).

In summary, there is conflicting evidence on the effect of amount of L2 use on L2 acquisition although it appears that L1 use does affect L2 accent regardless of whether the L2 was acquired as a child or an adult.

Social identity

As Zuengler (1988) states, "...pronunciation is a domain within which one's identity is expressed..." (p. 34). Research on social identity has employed both sociolinguistic and social constructivist frameworks. Studies on social identity and L2 phonology (e.g., Gatbonton 1975; Lybeck 2002; Thompson 1991) that have employed sociolinguistic frameworks have primarily focused on the use and acquisition of particular sounds and their variants in terms of their role as social markers of identity. Social constructivist research (e.g., Marx 2002; Morgan 1997; see also Hansen 2006, and Ohara 2001, above) has focused on how learners construct a viable identity in the L2, as well as how identity is related to access to L2 use opportunities.

The earliest research on social identity and L2 phonology has focused on how learners use and acquire the L2 sound system and retain certain variants of the L1 sound system in the L2 as markers of identity. Two important studies in this area are Gatbonton (1975) and Thompson (1991). Gatbonton's research focused on French-Canadian learners of English and their production of interdental fricatives in light of the participants' self-identification as nationalistic, and therefore pro-French, or non-nationalistic, and therefore pro-English. She found a higher amount of English L2 dental fricative use among non-nationalistic learners as well as an awareness among the learners of how accent signaled ethnic identity. Thompson's (1991) study focused on 36 Russian born immigrants, all of whom had professional ability in Russian, and their production of the English velar nasal and interdental fricatives on a reading passage and spontaneous speech. Ratings of global accentedness (see Munro, Chapter 7, in this volume) were conducted by both inexperienced and experienced raters. Interestingly, of the 36 participants that had come to the US when they were ten years old or younger, only 2 received any perfect ratings, and none was "consistently judged to be accent-free" (p. 193). Thompson believes that this can be explained by "the mutual effect between phonetic categories of English and Russian" (cf. Flege 1987), as well as the participants' social identity since they retained strong connections to a Russian community and had extensive use of Russian.

A recent study in this area, Lybeck (2002), combined social network theory with a reformulation of Schumann's acculturation model. In social network theory (cf. Milroy & Milroy 1992), there are three types of network structures: "*exchange* networks made up of ties with family and close friends, *interactive* networks constructed of ties with acquaintances, and *passive* networks that consist of physically distant ties" (Lybeck 2002: 176). As Lybeck notes, in close-knit exchange networks, "Individuals within exchange networks are likely to use the same linguistic variants as their network members whereas interactive networks are unlikely to enforce norms and are open to variation and change" (p. 176). Lybeck combines this with Schumann's Acculturation Model to theorize that learners who have exchange networks will have less social and psychological distance and will therefore have greater L2 learning than learners who only have interactive or passive networks.

Lybeck (2002) collected speech and social data through interviews from nine American women who had been living in Norway between one and three years at the time of the study. The participants' overall pronunciation accuracy, as well as the production of a particularly salient phonological marker, /r/, were analyzed against the women's social networks, categorized into three groups: "A: supportive engagement in exchange networks helped them reduce cultural distance; B: moderate cultural distance due to some success in developing contacts who were supportive; C: had a high level of cultural distance / unable to develop supportive networks" (p. 179). Lybeck found that the two women who had been categorized in the A group had the best pronunciation accuracy overall (over 80%), followed by the women in the B group. The C group had the lowest accuracy overall. Lybeck also found that the women in the A group, "used Norwegian *r* almost exclusively, showing identification with (low distance from) Norwegian culture" (p. 183). The women in the B group had more variable but still a great deal of /r/ use while the women in the C group either exclusively used American /r/ or decreased in their use of the Norwegian /r/ across time. As Lybeck states, "Those participants who were engaged in supportive exchange networks within the target culture were provided meaningful frameworks within which they could access and acquire both linguistically and culturally appropriate behaviors, effectively reducing their cultural distance, whereas those who were left outside of these networks or whose needs were not met by target-culture networks were not" (p. 184).

A second strand of work on social identity and L2 phonology has employed sociocultural and social constructivist frameworks to explore how learners construct a viable identity in the L2, as well as how identity is related to access to L2 language use and learning opportunities. One such study is Marx (2002), who conducted a first person reflective study of a Canadian English L1 speaker who moved to Germany for three years and then returned to Canada. Focusing her analysis on issues of accent and identity, Marx found that there were six main stages in her language learning and use: 1) displacement, which was initiated by her entry into the second culture, German. At this point, her German was English-accented and others perceived that she was American. In order to reject this identity, she worked hard to learn the L2 and avoided members of the L1 culture; 2) beginning stages of loss: after four months in the second culture, she took on a French accent (her first L2) because she perceived that French students were more positively perceived by Germans than American students; 3) towards a native speaker accent in the L2: after one year in the second culture she attempted to have native-like L2 accent in order to "be judged as a competent member of the [second culture]" (p. 272). During this period, she began to have difficulties in speaking the L1; 4) construction of an L2 identity and attrition of the L1: after 2 years in the second culture, many perceived her to be German due to her accent but also because she had also adopted clothing and manners of C2; she had more difficulties with speaking and writing in the L1; 5) re-entry into the C1: after 3 years in Germany, she returned to Canada. She had a British/German accented English for 3 months, as she wanted her L2 identity to be salient and wanted to preserve the outside identity/foreigner identity in C1; 6) reconstruction and renewal of the L1: Three months after her return to Canada, she moved to the US to study and teach. The 'false' L1 accent began to diminish. As Marx stated, "I returned to being a native Canadian and moved psychologically away from the [second culture]" (p. 276).

In Moyer's (2004) study on immigrants to German, discussed above, it was also found that the concept of 'confidence' in using the L2 was a major component in the participants' ability to develop not only L2 social contacts but also a sense of self or L2 identity, and that for the majority of the participants, developing a sense of self in the L2 was a struggle. However, the more they acquired of the language, and gained confidence in their ability to use the L2, the more the participants felt that they belonged in the L2 culture and were able to develop a L2 identity. Confidence in using the L2, as Moyer points out, has not received a great deal of attention from L2 researchers but may be a critical element in how learners view and make use of their linguistic abilities. Moyer's research also focuses on the issue of 'passing' (cf. Piller 2002; Rampton 2001) and the dynamic nature of L1 and L2 identity: "Some participants describe how they 'play' with language identity, i.e. purposefully misrepresenting their national heritage for their own amusement, as they put it" (p. 112). This would occur most frequently when the participants were outside the L2 context, traveling to other countries. As Moyer states, "The fact that these stories were not unusual shows that identity represents a conscious choice, that it is flexible and that there may be some special purpose in passing for a native speaker, particularly as a temporary performance" (p. 112–113).

As these studies, as well as the research discussed in other sections of the chapter, have found, learners may be active agents in targeting which variants to use and acquire and may use the variants purposefully to mark gender, social, and ethnic identity. Learners may also resist using certain variants if they perceive that doing so creates a L2 identity that is not viable.

Target language variety

There have been a number of studies (e.g., Adamson & Regan 1991; Anisman 1975; Thompson 1976; Wolfram, Carter, & Moriello 2004), as well as a number of discussion articles and reviews (cf. Beebe 1985; Dowd et al. 1990; Zuengler 1989b), that have examined target variety selection by L2 learners. This research has typically been sociolinguistic in nature, employing sociolinguistic and ethnographic interview techniques to elicit linguistic data for analysis, as well as information on the learners' social networks and social group targets/ preferences. This work is based on the view that L2 learners are not "passive recipients of comprehensible input or incomprehensible input from native speakers (NSs) but [are] active participants in choosing the target language models they prefer and thus acquiring 'the right stuff' according to their values" (Beebe 1980: 404).

The earliest work in this area was conducted by Anisman (1975) and Thompson (1976). Anisman's (1975) work examined the effect of peer group influences on language choice among speakers of Puerto Rican English, focusing on the voiced interdental fricative, /aɪ/, and the schwa. Anisman found that Puerto Rican adolescents with Black peer group contacts had more Black English variants than Standard English or Spanish variants. In contrast, adolescents who were targeting mainstream values/norms had more Standard English variants over Black English or Spanish variants. Finally, the adolescents who had the greatest amount of contact with a Puerto Rican peer group had the most Spanish variants. In work on L2 English of Chicanos, Thompson (1976) found that social class impacted target variety: learners who were of higher socioeconomic status and felt accent was important for social mobility targeted a regional variety of English. Learners from the same social class who did not feel that accent was important for social mobility targeted non-regional variety while learners from a lower socioeconomic status used a Spanish influenced variety of English. Adamson and Regan (1991), as described above, also found that gender influenced the {-ing} variants targeted by their Vietnamese learners of English. As the researchers found that the Vietnamese learners of English had the same variant use patterns by gender, they speculated that the Vietnamese women were targeting the variant employed by native speaking women and likewise for men.

In recent research, Wolfram, Carter, and Moriello (2004) studied differences in use of the /ai/ diphthong by L2 learners in an urban (Raleigh) versus rural (Siler City) setting in North Carolina due to the pervasive nature of glide reduction in this diphthong in southern English. They collected conversational interview data with 60 L2 learners who were immigrants from Mexico, El Salvador and Central and South American countries. Results from the analysis of /ai/ indicated that the participants who lived in the rural area had some glide reduction although it was not as pervasive as it was by non-Hispanic residents of this setting. The learners who lived in the urban setting had less glide reduction although there was more glide reduction by learners who had lived in the urban setting longer. The researchers speculate that these findings indicate that, with more L2 acquisition, learners become more aligned with local norms.

In conclusion, findings from this line of research indicate that a number of factors, such as peer group, social class, gender, and the stage of L2 acquisition can affect which language variety L2 learners target.

Variation

Variationist work in L2 phonology has for the most part been based on the work of the sociolinguist William Labov (cf. Labov 1966) and social psychologists Howard Giles and colleagues (cf. Giles & Powesland 1975). The issue of variation has long been debated in the SLA literature (cf. Ellis 1990; Gregg 1990; Tarone 1990) for a number of reasons. Firstly, the issue of variability is problematic for the construct of acquisition. If variability is a feature of production, does it mean that learners have not acquired a target language form if they produced it variably, even if the variation is systematic? In other words, is variation part of 'competence'? This latter view is espoused by most variationists. As Bayley and Regan (2004) state, "Variationist sociolinguistics...has suggested, convincingly in our view, that far from being a peripheral element, knowledge of variation is part of speaker competence. The implication of this position is that, in order to become fully proficient in the target language, second language learners also need to acquire native-speaker (NS) patterns of variation..." (p. 325).

L2 research employing this framework can by and large be categorized into three strands: research on interlocutor/speech accommodation, research that examines stylistic variation based on attention to speech and monitoring, and research that examines the role of linguistic and social factors on variation. The first strand was led by Beebe and colleagues (cf. Beebe 1977; Beebe & Zuengler 1985), and is based on the work by the social psychologist Giles and colleagues (cf. Giles & Powesland 1975). The second strand was led by the work of Tarone (cf. 1979, 1982). The third strand began emerging in the 1970s, with more recent work employing variable rule (VARBRUL) analysis for data analysis. Each is discussed in turn below.

Interlocutor/Speech accommodation

Speech Accommodation Theory (SAT), developed by Giles and colleagues (cf. Giles & Powesland 1975) has had a significant impact on how variation has been theorized in L2 phonology. As Zuengler (1989a) notes, SAT has received attention "as a paradigm for explaining second language (L2) performance variation" (p. 49) although it is not suggested that it is the only explanation for L2 sociolinguistic variation. As Beebe and Giles (1984) note, "SAT was devised to explain some of the motivations underlying certain shifts in people's speech styles during social encounters and some of the social consequences arising from them. More specifically, it originated in order to elucidate the cognitive and affective processes underlying speech convergence[1] and divergence" (p. 8).

Studies in this area have employed sociolinguistic interviews or short tasks and have usually focused on accommodation to the interlocutor (e.g., Beebe 1977; Beebe & Zuengler 1985; Sawyer 1973; Young 1987) or accommodation to the standard variant in the target language (e.g., Zuengler 1982, 1989a). For example, in research on Mexican-Americans interacting with Anglo and Hispanic interlocutors, Sawyer (1973) found that Spanish words were pronounced with Spanish pronunciation with a Hispanic interlocutor and with English pronunciation with an Anglo interlocutor. Similar results were found by Beebe (1977) in her study of the Thai usage of bilingual Chinese-Thai adults in Bangkok. When the participants were interviewed by a native Thai speaker, they had a significantly higher usage of Thai variants than when interviewers were Chinese and vice versa for Chinese interviewer and Chinese variants. These findings were also found for Chinese-Thai children (Beebe & Zuengler 1985).

Other factors may affect the extent to which a speaker identifies with, as well as accommodates to, the interlocutor. In research on Chinese speakers' production of the English plural, Young (1987) found that "...if interlocutors share other

1. Beebe and Giles (1984) go on to define convergence and divergence: "Convergence has been defined as a linguistic strategy whereby individuals adapt to each other's speech by means of a wide range of linguistic features including speech rate, pause and utterance lengths, pronunciations, etc. ... whereby divergence refers to the manner by which speakers accentuate vocal differences between themselves and others." (p. 8)

characteristics such as occupation, education, or gender, these characteristics in combination may override any single effect of shared ethnicity" (p. 84). Zuengler (1989a) also found that 'dominance' may be a factor. In her study, Zuengler examined the interaction in native speaker (NS)/nonnative speaker (NNS) dyads. The research focused on how perceptions of 'expertness' by either the NS or NNS would affect the level of standardness of production of four phonological variants – voiced and voiceless dental fricatives, (r), and (oh), a mid-back rounded vowel – which had been found by Labov (1966) to be socially conditioned in New York City, where the study took place. However, an initial analysis of the data showed only limited evidence that expertness affected variant usage; other factors, such as 'dominance' (operationalized as 'amount of talk' and 'interruptions') and ability to move task along, were found to affect the interaction. As Zuengler notes, several factors may be at play in these interactions: "One is dominance and another may be accommodation. The latter could be competing with, or stifled by, the former. Consequently, to explain performance in such interactions in accommodative terms alone . . . is to risk missing an equally, or more important, dynamic underlying the subjects' language performance" (p. 65).

Zuengler (1982) also found that ethnic threat may affect the extent to which speakers accommodate to the target language form. In her study of native speakers of Spanish and Greek, she analyzed the pronunciation of English pre-vocalic /r/, /ɪ/, and word-final /z/ for both groups and the voiced interdental fricative for Spanish speakers across three questions, the last one of which was 'ethnolinguistically threatening'. Zuengler found that

> . . .some of the subjects may have identified strongly as ethnic group members, and defended their ethnic solidarity through making their IL phonologically distinctive from that of the Anglo interlocutor. The other subjects, who increased in TL correctness, thereby making their speech more like that of the Anglo interlocutor, might not have been displaying ethnic solidarity. If so, they were possibly maintaining a distinctiveness from their own ethnic group in responding to the Anglo interlocutor. (p. 85–86)

To summarize, results of this research indicates that a number of factors can influence learners' use of a particular variant. These factors include the learners' perception of ethnic identity and ethnic threat. Additionally, other factors, such as dominance, may mitigate accommodation.

Attention to speech/monitoring

Work in the area of stylistic variation based on attention to speech/monitoring has been led by Tarone (1979, 1982) and modeled on the work of Labov (1969, 1972) and his Observer's Paradox, which is "the problem of observing how people speak when they are not being observed" (Labov 1972: 256). Tarone developed the

Capability Continuum based on Labov's Observer's Paradox, and gives a number of assumptions for the Continuum. The first assumption is that "The underlying IL capability is an abstract linguistic system which is inferred to exist apart from any particular instance of its use; this system consists of a range of styles, any one of which a speaker may use, for a variety of psychological and social reasons" (p. 152). According to Tarone, the range of speech styles in an individual's capability can be placed on a continuum, from less formal and more vernacular to more formal and more target-like. Although the degree to which each style is native-like differs, with the more native-like at the more formal end of the continuum, each style is systematic. The paradigm also assumes that the speech style of the learner is related to degree of attention (monitoring) paid to speech and that different speech styles can be elicited through different types of tasks. For example, tasks such as reading word lists would be perceived as eliciting more monitoring and careful speech, and therefore a more formal speech style. In contrast, a more naturalistic conversation would elicit less monitoring and therefore a more natural, vernacular style of speech.

A number of studies (e.g., Dickerson 1974, 1974; Dickerson & Dickerson 1977; Gatbonton 1975, 1978) support the assertion that learners' language differs across speech styles, and that tasks such as reading passages elicit more target-like speech. For example, in what is most likely the earliest variation study, Dickerson (1974, 1975), in her work on the pronunciation of English /z/ by Japanese learners, found that learners were more correct in word lists, then on reading dialogues, and least accurate in free conversation; additionally, production within each style was found to be systematic. Gatbonton (1975), in her research on the production of English interdental fricatives by French-Canadian learners of English, also found that in tasks where learners were hypothesized to pay more attention to speech (e.g., reading tasks), there were more target-like variants than in less formal tasks. Dickerson and Dickerson (1977) also found more correct usage of English /r/ for Japanese learners in word lists than in free conversation.

However, a number of other studies (Beebe 1980; Moyer 2004; Sato 1985; Schmidt 1977) have conflicting findings, indicating that style alone may not be the only factor to affect degree of accuracy. In her study of the acquisition of English word-final codas by a young Vietnamese boy, Sato (1985) found that task variation may depend on the phonological variable under study, as her results indicated that the learner sometimes produced the codas more target-like in the casual than in the more formal style. A study by Beebe (1980) on the production of /r/ by Thai learners of English found that linguistic environment had an effect on production based on the transfer of sociolinguistic patterns from Thai: while /r/ in final position had more target-like production in the careful style, initial /r/ was more correct in the vernacular style and had more L1 variants in the careful style. In his study of the production of English dental fricatives by Egyptian Arabic speakers, Schmidt (1977) found that production of the dental fricatives was influenced not only by task variation, but also by social class and educational background.

Moyer (2004), in her study of immigrants to Germany, found that there was no significant effect for task type for the 4 tasks in her study (word list, reading passage, spontaneous speech, and reciting proverbs) in the ratings of nativeness of her participants by native speakers of German. However, Moyer found that spontaneous speech was rated closer to native speech than any other task and that speech rated as the most non-native was elicited in the word list and reading passage tasks. As Moyer states, "This indicates that informal speech, perhaps reflecting a more natural rhythm and individual style, brings out the best performance" (p. 73). The formality of word list and reading passage tasks may also not foster the use of stress and rhythm, which may make speech sound more natural and thus perhaps more native. As Moyer concludes:

> … the presumed formality of a task may not be the salient factor in performance accuracy. It is far more likely that native-like delivery is a matter of suprasegmental and even pragmatic features, such as tempo, rhythm and style as well as linguistic control, or accuracy. The extent of contextual isolation, or even text type itself, may evoke varying degrees of naturalness in style, and therefore fluency. (p. 73)

In sum, there does not appear to be as direct a relationship between variation and task formality as Tarone's (1979) Capability Continuum suggests. While some research has suggested that learners are more target-like on more formal tasks, the extent to which task production can be linked to monitoring is unclear (cf. Brown & Fraser 1979; Giles 1973); additionally, other factors, such as linguistic environment, type of phonological variable under investigation, and social class and educational background may affect production.

Social and linguistic factors

This research has examined how linguistic factors (sometimes called 'internal' factors) such as preceding and following linguistic environment and extralinguistic and/or social factors (sometimes called 'external' factors) such as gender and social class, affect variable production. As Preston (1996) states, "The central claim of this approach is that the alternative forms of linguistic elements do not occur randomly. The frequency of their occurrences is predicted by 1) the shape and identity of the element itself and its linguistic context, 2) stylistic level (defined operationally), 3) social identity, and 4) 'historical' position (i.e., an assumption that, in much variation, one form is on the way in, the other on the way out" (p. 2). Early research typically employed descriptive statistics (e.g., percentages) while later research has employed variable rule (VARBRUL) analysis to develop probabilistic rules. VARBRUL employs loglinear regression to quantitatively model the effect (i.e., weight) of a particular factor (e.g., preceding linguistic environment) on the use by a learner of a particular variant.[2] Not all studies examine both linguistic and extralinguistic factors, so the interactions of these constraints are only discussed when they have been employed in a study and found to be significant.

Some of the work in L2 variation has focused on morphophonemics, such as past tense marking (e.g., /t d/ deletion), plural marking (e.g., /s z/ production), and {-ing}. For example, influenced by the work on /t d/ deletion in native varieties of English, L2 researchers have examined the extent to which the constraints operating on /t d/ deletion in nonnative varieties of English are similar to those for native varieties of English. In research on Vietnamese speakers of English, Wolfram (1985) found that both extralinguistic and linguistic factors constrained /t d/ deletion: participants who had a longer length of residence (4–7 years vs. 1–3 years) had a higher rate of past tense deletion in consonant clusters followed by a consonant, as well as more deletion on monomorphemic rather than past tense clusters, both patterns being similar to those found in native varieties of English.

In work on Chinese learners of English, Bayley (1996) found both divergence and convergence with target language patterns for /t d/ deletion. Findings on the effect of phonological environment, including preceding environment, following environment, and voicing agreement, were overall similar to findings for native speakers of English. However, the L2 learners in this study were more likely to reduce inflectional than lexical /t d/ clusters, which is the opposite of the pattern for native speakers of English, but confirms research by Wolfram and Hatfield (1984) on other non-native speakers of English, in this case, Vietnamese learners of English, who also had higher /t d/ deletion rates on inflectional rather than lexical /t d/ clusters. In terms of the effect of social factors, Bayley divided the participants into two groups: one that had a mixed social network, which included both Chinese and Americans, and another that had a primarily Chinese social network. He also examined the effect of language proficiency, rated either high or lower. Both of these extralinguistic factors were significant, with participants with a mixed social network being more likely to have /t d/ deletion; lower proficiency participants were more likely to delete the /t d/ than those participants labeled as having a higher proficiency. As Bayley explains, the lesser likelihood of lower proficiency learners to mark /t d/ appears to be a lack of acquisition of past tense as well as consonant clusters. However, the higher level of /t d/ deletion by learners who have a mixed social network may appear puzzling although, as Bayley asserts, this may be due to them acquiring more native-like patterns of /t d/ deletion as they are exposed to native speakers' variation patterns, more so than participants with

2. See Paolillo (2002) and Young and Bayley (1996) for detailed discussions of how to employ VARBRUL in linguistic analysis.

primarily Chinese social networks, who may speak more careful English as their primary English input may be in formal classroom settings.

Hansen (2005) also researched the /t d/ deletion patterns of Chinese learners of English, and focused on the acquisition of target language patterns by learners in the study. She found that four constraints operated on the deletion of /t d/ for the participants, with the following order of greatest to least effect: following linguistic environment, preceding linguistic environment, voicing agreement, and grammatical conditioning. The patterns overall indicated a process of acquisition of target language patterns of /t d/ deletion, though some individual differences existed. However, there were a great number of similarities across speakers and between the participants of this study and those of native speakers of English, indicating that the learners were in the process of acquiring the native speaker linguistic variation patterns.

Another area of research has been the {-s} morpheme. For example, Saunders (1987) conducted research on the production of voiceless stop + sibilant clusters in the third person singular on verbs or the plural morpheme on nouns. His participants were Japanese learners of English. Saunders found grammatical category had an effect on production, as learners had a higher rate of errors on third person singular (45%) than on plural nouns (32%). Preceding linguistic context, in this case type of voiceless stop, also had an effect on production: across both verbs and nouns, error rates were highest on /ts/ clusters, followed by /ps/ and least on /ks/.

Young (1988), in research on {-s} inflection on plural nouns by Chinese learners of English, also found that preceding linguistic environment affected plural marking, with preceding non-sibilant fricatives, vowels and stops promoting marking and preceding sibilants, nasals, and laterals inhibiting marking. The extralinguistic factor of proficiency was also found to affect plural marking, with participants with high proficiency favoring plural marking over those with low proficiency. Other factors, such as position of the noun in the noun phrase, function of nouns in noun phrases, and following linguistic environment also affected plural marking.

Wolfram, Christian, and Hatfield (1986) investigated four grammatical structures – plural absence, agreement marking, negation, and tense marking – along with age and years in the US for Vietnamese immigrants to the US. The researchers found that native-like variation was conditioned by years in the US and age, with adolescents (versus adults aged 20 and over) more likely to conform to native-like patterns if they had been in the US for over four years, while the other groups did not conform nearly as well.

In their research on the variable {-ing}, Adamson and Regan (1991) also found that both linguistic and social factors affected whether the participants employed [ɪn] or [iŋ], with gender, style, and grammatical category all being significant. Specifically, the researchers found that women tended to use [iŋ] more than men, that this variable is used by both groups more often in monitored over unmonitored tasks, and that nouns favored [iŋ] while verbs, particularly the progressive and periphrastic future, did not.

There have also been a number of L2 phonology variation studies. For example, Dickerson's (1975) research on the production of /z/ by Japanese learners of English (discussed in more detail under "Attention to speech/monitoring" above) found that in addition to task, phonetic environment also affected /z/ production, with a following vowel promoting accurate production of /z/, while a following pause or following consonant promoting the deletion of /z/ or production of /z/ as [s] or [dʒ], for example.

Ross (1994) focused on paragoge (final vowel insertion) and apocope (final vowel deletion) in Japanese English, and found that three factors affected paragoge while two factors affected apocope. For the former, intonation of utterance, ultimate syllable of the word, and following segment were significant, with a low-falling tone promoting paragoge while a rising tone inhibited it; [-son] in the final syllable motivated paragoge as obstruents promoted paragoge while nasals, glides, and laterals inhibited it; and paragoge was promoted when the following segment was a consonant and inhibited when the following segment was a pause or a vowel. For apocope, word stress and syllable final consonant were significant. Stressed syllables had more cases of apocope than unstressed syllables, and apocope occurred more often with final affricates than with continuants.

Hansen (2001), in research on the acquisition of English L2 syllable codas by native speakers of Chinese, found that both grammatical conditioning and linguistic environment affected the production of codas. Specifically, she found that the participants of this study deleted final /t d/ on lexical over inflectional clusters, which contrasted to the patterns found for other non-native speakers of English (e.g., Bayley 1996; Wolfram & Hatfield 1984), but was similar to patterns for native speakers of English (Labov 1989). Both preceding and following linguistic environment were found to have an effect on coda production. The study also found that homovoicing of segments favored absence while heterovoicing favored retention.

Hansen (2004) found that different factors had an effect on different types of production. She analyzed the production of English syllable codas by Vietnamese learners of English across five types of production (or lack thereof): target-like production, production with epenthesis, production with feature change, deletion, and two types of production modifications (e.g., in a two-member cluster, deletion of one member and epenthesis of the other). For both target-like production and deletion, both coda length (one, two, or three member coda) and preceding linguistic environment had an effect, while for epenthesis, these two factors along with following linguistic environment, syllable stress, and time (data were collected three times over the duration of one year) were significant. For feature change, following linguistic environment, length, and stress were significant while for two types of production modifications, length had a significant effect. Finally, individual difference were also found for three of the production types: target-like production, absence, and two-types of production.

While not variation studies, other studies on L2 phonology have also found that linguistic environment affects L2 production. For example, both Gatbonton (1978) and Major (1996) also found that a following vowel may facilitate production (vs. deletion) of a given segment while other researchers (Edge 1991; Major 1987; Tarone 1980) have found that a following pause may facilitate devoicing and/or epenthesis. Benson (1988), Osburne (1996), and Yavaş (1997) also found that the preceding linguistic environment had an effect on production, with a preceding diphthong promoting absence of the following coda for Vietnamese speakers, a finding that was confirmed by Hansen (2004) as well, and a high vowel promoting devoicing.

Additionally, a number of non-linguistic factors have been found to affect variation. Flege, Munro, and MacKay (1996) examined the voice onset time (VOT) values of English stops by native speakers of Italian as well as the production of interdental fricatives, and found that for production of interdental fricatives, age of L2 learning, home use, integrative motivation, and work use were significant while for VOT in stop consonants, age of L2 learning, social use, home use, and work use were significant (see Zampini, Chapter 8 of this volume, for a complete description of VOT and related studies).

As these studies show, there have been consistent findings indicating that a number of linguistic and non-linguistic factors constrain production. Linguistic factors such as voicing agreement, preceding linguistic environment, following linguistic environment, stress, intonation, coda length, and grammatical category, as well as non-linguistic/social factors such as gender, proficiency level, task, use of L2 at home, work, and socially, age of L2 learning, motivation, and length of stay affect L2 variation.

Methodological choices

There have been a number of approaches to the study of social factors and variation in L2 phonology: experimental approaches that typically entail recording word list and/or reading passage data that is then rated by native speaking judges; sociolinguistic approaches that involve sociolinguistic interviews and either using variable rule analysis or other inferential or descriptive statistics; and the use of multiple techniques, such as self-reports, observations, and interviews along with more experimental data. Each of these approaches is discussed below.

The earliest studies on the effect of social factors on L2 phonology as well as more recent studies focusing on extent of L1 use have employed experimental research methods for both data collection and analysis. The focus of these studies (cf. early studies such as Asher & Garcia, Purcell & Suter 1980; Suter 1976; Thompson 1991 as well as more recent research on extent of L1 use such as Flege, Frieda, & Nozawa 1997; Guion, Flege & Loftin 2000; Piske & MacKay 1999; Piske, MacKay, & Flege 2001) has typically been the rating of pronunciation accuracy as measured against a number of predictor variables such as age of arrival, length of stay, extent of L1/L2 use, gender, etc. Data are commonly gathered via word list and/or reading passages, and questionnaires may be used to elicit background data about L1/L2 use, etc. Accent and intelligibility ratings are conducted on the phonological data and the questionnaire data is quantified; data are then analyzed via a variety a statistical procedures such as correlations, ANOVAs, and/or multiple regression to determine the strength and nature of the relationships between the predictor variables and the pronunciation accuracy rating (see Chapter 7 this volume by Munro for a further discussion of accent and intelligibility ratings).

One strength of this approach is that it offers researchers statistical power to support their findings. However, there are also a number of weaknesses with this approach: firstly, it is not clear whether ratings of pronunciation accuracy as based on highly controlled tasks such as word lists and reading passages accurately reflects the learners' abilities in the L2. Secondly, self-report on questionnaires and not interviews and/or observations are employed to solicit information on social factors such as L1 use – it may be the case that the participants over or underestimate their L2/L1 use. Finally, as will be discussed in more detail below, a number of social factors (e.g., gender and identity) are confounded, and using one-time research (i.e., gathering data only one time) that conceptualizes these concepts as stable and unchanging rather than dynamic may not fully portray the complex social context of the language learner.

Sociolinguistic research methods have also commonly been employed in L2 phonology research, particularly for research on variation and interlocutor/speech accommodation (cf. Adamson & Regan 1991; Bayley 1996; Beebe 1980; Beebe & Zuengler 1985; Dickerson 1974; Hansen 2005; Young 1987; Zuengler 1989a, 1989b). In this methodology, data is most commonly gathered through sociolinguistic interviews; in these interviews, the interlocutor may ask the participant to talk about emotional subjects such as dangerous experiences in the belief that these topics makes the participant less focused on how they are speaking and more on what they are saying. The interviews are then coded for the use of the variant under study in order to determine patterns in the use of the specific variants; this may be analyzed via descriptive statistics or through loglinear regression programs such as VARBRUL (see Paolillo 2002; and Young & Bayley 1996, for detailed discussions of VARBRUL). In a VARBRUL analysis, data are also coded for a number of linguistic and social factors, such as preceding linguistic environment, following linguistic environment, gender, task, social networks, etc.; VARBRUL then models the variation through a series of loglinear regressions in order to determine the model that best fits the data. The effects are given as weights from 0 to 1.00, with weights below .50 perceived to inhibit the production or deletion of the phonological variant under study (e.g., in /t d/ deletion research, the presence or deletion of the /t d/) and weights above .50 said to promote the production or deletion of the variant. For example, the researcher may focus on /t d/ deletion patterns of L2 learners of English and analyze deletion patterns against such factors as preceding and following linguistic environment, length of the coda (CC or CCC), grammatical category (monomorphemic or bimorphemic), gender of the participant, time (if more than one data set is collected), etc. Typically, only a number of these factors may be found to best explain the variation patterns in /t d/ deletion.

This approach also has a number of strengths. For example, like experimental research, it offers the researchers statistical power to support the findings. Another strength is that it allows the researcher to explore multiple factors, including both linguistic and social factors. However, there are also a number of shortcomings to this approach. One criticism of this line of research is the nature of the interviews; it is questionable whether 'emotional' topics elicit a more vernacular (and less monitored) style of speech than other topics. Another criticism is that is it treats social variables such as gender and social networks as reductionist, and codes "...aspects of social identity as categorical and invariant across contexts" (Ehrlich 1997:421). As Eckert (1991) notes, an additional problem with this research is that "general sociological factors are applied without attempting to identify community-specific factors that might also be relevant" (p. 7). Without employing ethnographic data collection techniques in order to examine the community and the participants' lives and interactions in more depth, it is not possible to determine whether:

> the social factors traditionally used in studies of majority sound change, such as age, sex and social class, are sufficient for an explanation of sociolinguistic variation in this community.... The use of ethnography in the study of variation allows the researcher to discover the social groups, categories and divisions particular to the community in question, and to explore their relation to linguistic form.
> (Eckert 1991:7)

Recent variationist research acknowledges this problem and instead incorporates ethnographic research into the design to first determine variables that can then be examined through variable rule analysis. While there have been a number of research studies in this area in sociolinguistics, such as Eckert (1988, 1991) and Fought (1999), there has been no research on L2 phonology to date, as far as this researcher is aware, that employs this approach although this direction of variationist research provides a way of integrating both qualitative and quantitative research methods, which enables both deeper and wider analyses of issues under investigation.

Recently, L2 phonology researchers have begun employing a wider range of approaches to explore L2 phonology. For example, Marx (2002) employed self-report and self-observation in her research on her phonology use and acquisition across different social contexts. Hansen (2006), Lybeck (2002), Moyer (2004) and Ohara (2001) all employed both statistical analysis of data collected via interviews and/or controlled tasks as well as interviews and observations that probed the participants' social networks, social identities, and other social factors. The use of multiple data collection and analysis tools is the most promising direction for future research as it provides us with a deeper, broader, and more robust insight into the phenomena under study.

Synthesis of major findings

Two major finding emerge from the research on social factors and variation in L2 phonology. One finding is that learners are active agents in choosing not only what and how they use their L2, but also in choosing the L2 target, and therefore what they acquire of the L2. Another finding is that certain factors such as access to L2 use and linguistic environment, factors that may be beyond the learner's control, also impact L2 learning.

Much of the research on social factors, especially the work on gender, target language variety, interlocutor/speech accommodation, and identity, has shown that learners are sophisticated L2 users and L2 learners, and they are active agents in what elements of the L2 they target for acquisition and/or use in different contexts. For example, research has indicated that learners are able to accommodate their speech to their interlocutor based on perceived similarities such as ethnic identification (Beebe 1977; Beebe & Zuengler 1985; Sawyer 1973) and occupation, education, and gender (Young 1987). Additionally, learners may be aware of how certain variants are used by speakers in different contexts/communities. Therefore, they may actively use (or avoid using) some variants or linguistic features over others based on gender, ethnic, national identities (cf. Adamson & Regan 1991; Gatbonton 1975; Ohara 2001) and peer group identifications (Anisman 1975; Thompson 1976).

At the same time, both social and linguistic factors also limit/affect L2 use and production. As variation studies have shown, a number of linguistic constraints, such as following and preceding linguistic environment, grammatical conditioning, voicing agreement, etc., affect the production of a particular phonological variant. These linguistic constraints may be connected to the acquisition of a particular structure (e.g., see the research on the acquisition of /t d/ deletion patterns in Hansen 2005). Additionally, learners' abilities to gain access to L2 use opportunities and the density of this access, as well as attitudes to the L1 and L2 community (cf. Hansen 2006; Lybeck 2002; Marx 2002; Moyer 2004) may affect not only the learners' use of L2 but also their perceptions of their own L1 and L2 identities, and therefore, their willingness – or lack thereof – to acquire and/or use the appropriate speech markers to signal belongingness in that community (Lybeck 2002; Ohara 2001).

The nexus of these two phenomena – both having control over the use and acquisition of the L2 while at the same time lacking control[3] is what makes language learning highly individual. While linguistic (and task) constraints will always affect L2 phonological production, and therefore in a sense always be beyond the learner's explicit control, the acquisition of native speaker linguistic constraint patterns is probably connected to the extent to which the learner has access to L2 communities and L2 use opportunities. Access – or lack thereof – to various communities may affect what elements of the L2 are targeted for acquisition and use, as well as the extent to which L2 learners use or avoid using (or avoid acquiring) certain features of the L2, which they perceive would, if used, create an L2 identity that they do not find viable or conflicts with their L1 identity (cf. Hansen 2006).

There are a number of implications of these findings. Firstly, as Cook (2002) suggests, we need to shift our view of learners to 'users' of language rather than 'learners.' What we perceive as L2 learners' 'deviations' from the standard target language may not be mistakes or errors; instead, this usage could be purposeful. In other words, learners may know that they are deviating from standard L2 usage but choose to do so for a number of reasons. A second implication has to do with research methodology: ethnographic techniques such as long-term observations and interviews need to be employed along with experimental approaches to determine whether the speech that we analyze is in fact representative of the speech of the participant and under what social conditions. In particular, it is important to determine whether use of a non-standard variant or incorrect pronunciation is indicative of a lack of acquisition or avoidance of use, e.g., whether they are forms retained from the L1 as identity markers, are used to avoid a L2 marker that the participant finds stigmatizing, and/or is the form the participant is targeting due to her/his social group. Additionally, data should be subjected to a more complex linguistic analysis, such as those conducted in variation studies that examine the effect of linguistic and task constraints, since use of the L2 will always be variable

3. See chapters on acquisition by Eckman and Major, this volume, for a discussion on how other factors such as transfer and markedness affect acquisition.

across different social and linguistic contexts. We need to understand how these contexts affect how learners acquire and use the L2.

Future directions

A number of issues, given below, need investigation:

- Research needs to be conducted to investigate the interface between variation and acquisition. For example, do learners acquire the variation patterns found in the target language as they acquire the variant in question, or do learners need to be proficient to a certain extent in order to acquire these patterns?
- Research needs to be conducted on suprasegmentals and variation/social factors, especially in relation to gender, culture, and identity, as well as variation;
- Research needs to incorporate more ethnolinguistic research techniques in order to determine which social factors are relevant in a given community or for the participants in the study rather than assigning social factors apriori;
- Gender and social identity research needs to be expanded to research on men; as yet, there have been only a few research studies that have focused on men in L2 phonology;
- Research needs to view language learners as 'users' of the L2, who construct their own identities, instead of comparing them, typically negatively, against standard target language models. We need to understand the use of certain variants against social context to determine whether not acquired or a marker that functions specifically in a context (e.g., /t d/ deletion);
- Research on "passing" for L2 users should be conducted. As Rampton (2001) states, "crossing's defining interest [is] in the use of a language that doesn't obviously belong to the speaker" (Rampton 2001: 50). Research by Piller (2002) on German L1 users' use of different German regional dialect markers indicate "L2 users may strategically employ stereotypical features characteristic of a particular variety in order to pass" (p. 193). As she states, some want to hide L1 background "Thus, successful L2 users do not necessarily aim to pass for native speakers. Rather, they just don't want to be perceived as members of a particular national group right away" (p. 194) to avoid being stereotyped. Research by Marx (2002) and Moyer (2004) gives some insight into this phenomenon; however, further research in this new direction is necessary;
- Finally, in light of recent research in SLA on how the L1 and L2 community may constrain the access L2 learners have to linguistic resources (cf. Blackledge 2001; Cumming & Gill 1992) and the findings on the effects of L1 use discussed above, it appears that research adding a phonological analysis component to this focus would be promising.

Suggested readings

Bayley, R., and Preston, R. 1996. *Second language acquisition and linguistic variation*. Amsterdam: John Benjamins Publishing Company.

Dowd, J., Zuengler, J., & Berkowitz, D. 1990. L2 social marking: Research issues. *Applied Linguistics* 11: 16–29.

Fought, C. 1999. A majority sound change in a minority community: /u/-fronting in Chicano English. *Journal of Sociolinguistics 3: 5–23.*

Paolillo, J. C. 2002. *Analyzing linguistic variation: Statistical models and methods.* Stanford, California: CSLI Publications.

Piske, T., MacKay, I. R. A., & Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215.

References

Adamson, H. D., & Regan, V. M. 1991. The acquisition of community speech norms by Asian immigrants learning English as a second language. *Studies in Second Language Acquisition* 13: 1–22.

Asher, J. J., & Garcia, R. 1969. The optimal age to learn a foreign language. *The Modern Language Journal* 53: 334–341.

Anisman, P. H. 1975. Some aspects of code switching in New York Puerto Rican English. *Bilingual Review* 2: 56–85.

Bayley, R. 1996. Competing constraints on variation in the speech of adult Chinese learners of English. In R. Bayley and R. Preston (eds.), *Second language acquisition and linguistic variation*, Amsterdam: John Benjamins Publishing Company, 97–120.

Bayley, R., & Regan, V. 2004. Introduction: The acquisition of sociolinguistic competence. *Journal of Sociolinguistics* 8: 323–338.

Beebe, L. M. 1977. The influence of the listening on code-switching. *Language Learning* 27: 331–339.

Beebe, L. M. 1980. Sociolinguistic variation and style shifting in second language acquisition. *Language Learning* 30: 433–447.

Beebe, L. M. 1985. Input: Choosing the right stuff. In S. M. Gass & C. G. Madden (eds.), *Input in second language acquisition,* Cambridge: Newbury House, 404–414.

Beebe, L. M., & Giles, H. 1984. Speech-accommodation theories: A discussion in terms of second-language acquisition. *International Journal of the Sociology of Language* 46: 5–32.

Beebe, L. M., & Zuengler, J. 1985. Accommodation theory: An explanation for style shifting in second language dialects. In N. Wolfson and E. Judd (eds.), *Sociolinguistics and language acquisition,* Rowley, Mass.: Newbury House Publishers, Inc., 195–213.

Benson, B. 1988. Universal preference for the open syllable as an independent process in interlanguage phonology. *Language Learning* 38: 221–235.

Blackledge. A. 2001. Complex positionings: Women negotiating identity and power in a minority urban setting. In A. Pavlenko, A. Blackledge, I. Piller, & M. Teutsch-Dwyer, (eds.), *Multilingualism, second language learning, and gender*, Berlin: Mouton de Gruyter, 53–75.

Brown, P. & Fraser, C. 1979. Speech as a marker of situation. In K. R. Scherer and H. Giles (eds.), *Social markers in speech,* Cambridge: Cambridge University Press.

Cook, V. 2002. *Portraits of the L2 user.* Clevedon: Multilingual Matters.

Cumming, A., & Gill, J. 1992. Motivation or accessibility? Factors permitting Indo-Canadian women to pursue ESL literacy instruction. In B. Burnaby & A. Cumming (eds.), *Socio-political aspects of ESL education in Canada*, Toronto: OISE Press, 241–252.

Derwing, T. M. 2008. Curriculum issues in teaching pronunciation to second language learners. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Dickerson, L. 1974. *Internal and external patterning of phonological variability in the speech of Japanese learners of English: Toward a theory of second-language acquisition*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Dickerson, L. 1975. The learner's interlanguage as a system of variable rules. *TESOL Quarterly* 9: 401–407.

Dickerson, L., & Dickerson, W. 1977. Interlanguage phonology: Current research and future directions. In S. P. Corder, & E. Roulet (eds.), *The notions of simplification, interlanguages and pidgins and their relation to second language pedagogy*, Faculté des Lettres, Neufchâtel, and Librairie Droz, Genève.

Dowd, J., Zuengler, J., & Berkowitz, D. 1990. L2 social marking: Research issues. *Applied Linguistics* 11: 16–29.

Eckert, P. 1988. Sound change and adolescent social structure. *Language in Society* 17: 183–207.

Eckert, P. 1991. *New ways of analyzing sound change.* New York: Academic Press.

Edge, B. A. 1991. The production of word-final voiced obstruents in English by L1 speakers of Japanese and Cantonese. *Studies in Second Language Acquisition* 13: 377–393.

Ehrlich, S. 1997. Gender as social practice: Implications for second language acquisition. *Studies in Second Language Acquisition* 19: 421–446.

Ellis, R. 1990. A response to Gregg. *Applied Linguistics* 11: 384–391.

Elliott, A. R. 1995. Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *The Modern Language Journal* 79: 356–371.

Flege, J. E. 1987. The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 16: 47–65.

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues*, Timonium, MD: York Press, 229–273.

Flege, J. E., & Fletcher, K. L. 1992. Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America* 91: 370–389.

Flege, J. E., Frieda, E. M., & Nozawa, T. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics* 25: 169–186.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. 1996. Factors affecting the production of word-initial consonants in a second language. In R. Bayley and R. Preston (eds.), *Second language acquisition and linguistic variation,* Amsterdam: John Benjamins Publishing Company, 47–73.

Fought, C. 1999. A majority sound change in a minority community: /u/-fronting in Chicano English. *Journal of Sociolinguistics* 3: 5–23.

Gatbonton, E. 1975. *Systematic variations in second language speech: A sociolinguistic study.* Unpublished doctoral dissertation, McGill University.

Gatbonton, E. 1978. Patterned phonetic variability in second-language speech: A gradual diffusion model. *Canadian Modern Language Review* 34: 335–347.

Giles, H. 1973. Accent mobility: A model and some data. *Anthropological Linguistics* 15: 87–105.

Giles, H., & Powesland, P. F. 1975. *Speech style and social evaluation.* London: Academic Press.

Gregg, K. R. 1990. The variable competence model of second language acquisition, and why it isn't. *Applied Linguistics* 11: 364–383.

Guion, S. G., Flege, J. E., & Loftin, J. D. 2000. The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. *Journal of Phonetics* 28: 27–42.

Hansen, J. G. 2001. Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics* 22: 338–365.

Hansen, J. G. 2004. Developmental sequences in the acquisition of L2 syllable codas: A preliminary study. *Studies in Second Language Acquisition* 26: 85–124.

Hansen, J. G. 2005. -t/d absence and the acquisition of linguistic variation by second language learners of English. Unpublished manuscript.

Hansen, J. G. 2006. *Acquiring a non-native phonology: Linguistic constraints and social barriers.* London: Continuum Publishers.

Ioup, G. 2008. Exploring the role of age in the acquisition of a second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Labov, W. 1966. *The social stratification of English in New York City.* Washington, D.C.: Center for Applied Linguistics.

Labov, W. 1969. The study of language in its social context. *Studium Generale* 23: 30–87.

Labov, W. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. 1989. The child as linguistic historian. *Language Variation and Change* 1: 85–97.

Lybeck, K. 2002. Cultural identification and second language pronunciation of Americans in Norway. *The Modern Language Journal* 86: 174–191.

Major, R. C. 1987. Variation in second language phonology. *ERIC reproduction service document*.

Major, R. C. 1996. Markedness in second language acquisition of consonant clusters. In R. Bayley and R. Preston (eds.), *Second language acquisition and linguistic variation,* Amsterdam: John Benjamins Publishing Company, 75–96.

Marx, N. 2002. Never quite a 'native speaker': Accent and identity in the L2 – and the L1. *The Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes* 59: 264–281.

Milroy, L., & Milroy, J. 1992. Social network and social class: Toward an integrated sociolinguistic model. *Language in Society* 21: 1–26.

Moyer, A. 2004. *Age, accent and experience in second language acquisition: An integrated approach to critical period inquiry.* Clevedon: Multilingual Matters.

Munro, M. J. 2008. Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Ohara, Y. 2001. Finding one's voice in Japanese: A study of the pitch levels of L2 users. In A. Pavlenko, A. Blackledge, I. Piller, and M. Teutsch-Dwyer (eds.), *Multilingualism, second language learning, and gender*, Berlin: Mouton de Gruyter, 231–256.

Olson, L. L., & Samuels, S. J. 1973. The relationship between age and accuracy of foreign language pronunciation. *Journal of Educational Research* 66: 263–268.

Osburne, A. G. 1996. Final cluster reduction in English L2 speech: A case study of a Vietnamese speaker. *Applied Linguistics* 17: 164–181.

Paolillo, J. C. 2002. *Analyzing linguistic variation: Statistical models and methods.* Stanford, California: CSLI Publications.

Pavlenko, A., & Piller, I. 2001. New directions in the study of multilingualism, second language learning, and gender. In A. Pavlenko, A. Blackledge, I. Piller, and M. Teutsch-Dwyer (eds.), *Multilingualism, second language learning, and gender*, Berlin: Mouton de Gruyter, 17–52.

Piller, I. 2002. Passing for a native speaker: Identity and success in second language learning. *Journal of Sociolinguistics* 6: 179–206.

Piske, T., & MacKay, I. 1999. Age and L1 use affects on degree of foreign accent in English. *Proceedings of the 14th international congress of phonetic sciences,* San Francisco, 1–7 August 1999, 1433–1436.

Piske, T., MacKay, I. R. A., & Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215.

Preston, D. (1996). Variationist perspectives on second language acquisition. In R. Bayley and R. Preston (eds.), *Second language acquisition and linguistic variation*, Amsterdam: John Benjamins Publishing Company, 1–45.

Purcell, E. T., & Suter, R. W. 1980. Predictors of pronunciation accuracy: A reexamination. *Language Learning,* 30: 271–287.

Rampton, B. 2001. Crossing. In A. Duranti (ed.), *Key terms in language and culture*, Oxford: Blackwell, 49–51.

Ross, S. 1994. The ins and outs of paragoge and apocope in Japanese-English interphonology. *Second Language Research* 10: 1–24.

Sato, C. J. 1985. Task variation in interlanguage phonology. In S. M. Gass and C. G. Madden (eds.), *Input in second language acquisition*, Cambridge, Mass., Newbury House.

Saunders, N. J. 1987. Morphophonemic variation in clusters in Japanese English. *Language Learning* 37: 247–270.

Sawyer, J. 1973. Social aspects of bilingualism in San Antonio, Texas. In R. W. Bailey and J. L. Robinson (eds.), *Varieties of present-day English,* New York: Macmillan Publishing Co., Inc.

Schmidt, R. W. 1977. Sociolinguistic variation and language transfer in phonology. *ERIC Reproduction Service Document ED 135 235.*

Siegal, M. 1996. The role of learner subjectivity in second language sociolinguistic competency: Western women learning Japanese. *Applied Linguistics* 17: 356–382.

Snow, C. E., & Hoefnagel-Höhle, M. 1977. Age differences in the pronunciation of foreign sounds. *Language & Speech* 20: 357–365.

Suter, R. W. 1976. Predictors of pronunciation accuracy in second language learning. *Language Learning* 26: 233–253.

Tahta, S., Wood, M., & Loewenthal, K. 1981. Foreign accents: Factors relating to transfer of accent from the first language to a second language. *Language & Speech* 24: 265–272.

Tarone, E. E. 1979. Interlanguage as chameleon. *Language Learning* 29: 181–191.

Tarone, E. E. 1980. Some influences on the syllable structure of interlanguage phonology. *IRAL* 18: 139–151.

Tarone, E. E. 1982. Systematicity and attention in interlanguage. *Language Learning* 32: 69–84.

Tarone, E. E. 1990. On variation in interlanguage: A response to Gregg. *Applied Linguistics* 11: 392–400.

Thompson, R. 1976. Mexican-American English: Social correlates of regional pronunciation. *American Speech* 50: 18–24.

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

Wolfram, W. 1985. Variability in tense marking: A case for the obvious. *Language Learning* 35: 229–253.

Wolfram, W., Carter, P., & Moriello, B. 2004. Emerging Hispanic English: New dialect formation in the American South. *Journal of Sociolinguistics* 8: 339–358.

Wolfram, W., Christian, D., & Hatfield, D. 1986. The English of adolescent and young adult Vietnamese refugees in the United States. *World Englishes* 5: 47–60.

Wolfram, W., & Hatfield, D. 1984. *Tense marking in second language learning: Patterns of spoken and written English in a Vietnamese community.* Washington, DC: Center for Applied Linguistics [ERIC document ED 25 960].

Yavaş, M. 1997. The effects of vowel height and place of articulation in interlanguage final stop devoicing. *IRAL* 35: 115–125.

Young, R. 1987. Variation and the interlanguage hypothesis. *Studies in Second Language Acquisition* 10: 281–302.

Young, R., & Bayley, R. 1996. VARBRUL analysis for second language acquisition research. In R. Bayley and R. Preston (eds.), *Second language acquisition and linguistic variation*, Amsterdam: John Benjamins Publishing Company, 253–306.

Zampini, M. L. 2008. L2 speech research production research: Findings, issues, and advances. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Zuengler, J. 1982. Applying accommodation theory to variable performance data in L2. *Studies in Second Language Acquisition* 4: 181–192.

Zuengler, J. 1988. Identity markers and L2 pronunciation. *Studies in Second Language Acquisition* 10: 33–49.

Zuengler, J. 1989a. Assessing an interaction-based paradigm: How accommodative should we be? In M. R. Eisenstein (ed.), *The dynamic of interlanguage: Empirical studies in second language variation*, New York: Plenum Press, 49–67.

Zuengler, J. 1989b. Identity and IL development and use. *Applied Linguistics* 10: 80–96.

PART III

Technology, training, and curriculum

Preface

Parts I and II of this book have focused on theoretical issues related to L2 phonology, as well as studies on the production and perception of L2 speech sounds. In much of the work surveyed to this point, researchers have examined particular aspects of L2 speech in an attempt to gain a better understanding of the nature of the learner's interlanguage phonology – that is, the intermediate and incomplete knowledge of the L2 sound system that the learner has at any given point in time during the acquisition process. Through an examination of L2 speech production and perception, researchers hope to gain insight into the ways in which learners organize, process, and realize L2 speech sounds; identify factors that intervene and affect the formation and evolution of an L2 phonology; and outline ways in which the learner's internal representation of the L2 sound system may change over the course of acquisition. As research findings converge, and through an analysis of the ways in which findings differ, it may be possible to begin to develop a model for L2 phonology that adequately reflects the nature of the acquisition process. In order to be useful, however, such a model must also inform more practical and applied domains of L2 speech: the teaching and training of L2 sound patterns and pronunciation. In the same way, the results from research on training have important implications for, and may lend insights to, the development of adequate models of L2 phonology. A discussion of the state of the art in L2 phonological acquisition, therefore, would be incomplete without an overview of applied L2 phonetics and phonology. To that end, the chapters in Part III reflect a more applied focus and examine one or more of the following themes: the use of technology for training and pedagogy in L2 phonological development, effective training practices, and issues related to curriculum and materials development. The first two chapters of this section deal with training in the articulation of individual speech sounds. Chapter 10 examines both traditional and current laboratory training methods in the perception and production of L2 speech sounds and highlights the most significant findings and outstanding problems in this area of research. Chapter 11 describes promising new research on the use of one particular device – ultrasound technology – as a tool for articulatory training. These two chapters, like much of the research surveyed in earlier sections of this book, focus on individual speech sounds. Chapter 12, on the other hand, focuses on theoretical and pedagogical issues related to the training of suprasegmentals in L2 speech, including intonation and discourse prosody. In addition, it discusses important research on the development and use of instructional software for teaching and training prosody in L2. Finally, Chapter 13 surveys recent research related to the teaching of pronunciation and discusses issues of curriculum and materials development. This final chapter thus provides a fitting conclusion to the book by examining ways in which research findings may be used to improve and shape curricular decisions regarding the teaching of pronunciation in the L2 classroom.

In Chapter 10 ("Training non-native language sound patterns: Lessons from training Japanese adults on the English /ɹ/–/l/ contrast"), Ann R. Bradlow examines research on the effects of training on L2 speech production and perception. In her approach to this topic, she employs the acquisition of the English /ɹ/–/l/ contrast by native speakers of Japanese as a focal point of departure and illustrative case study for more general issues of L2 training. Since this contrast has proved particularly difficult for L1 speakers of Japanese, it has been well-studied in training research (in fact, it is probably the most commonly studied contrast), and as such, it exemplifies many of the issues and problems involved in the development of adequate training techniques. Bradlow first provides an overview of what Japanese learners of English need to learn in order to accurately perceive and produce the English /ɹ/–/l/ contrast through a detailed review of the relevant research. She then outlines procedures for perceptive training of this contrast; in doing so, she also incorporates a broader discussion of methodological approaches to training. Bradlow concludes her chapter with an appraisal of lessons to be derived from the findings of the research on the training of the /ɹ/–/l/ contrast, not only for /ɹ/–/l/ training in particular, but for non-native speech sound training in general, as well. She also identifies a number of areas for future research based on these lessons.

In Chapter 11 ("Ultrasound imaging applications in second language acquisition"), Bryan Gick, Barbara Bernhardt, Penelope Bacsfalvi, and Ian Wilson present cutting-edge research on the use of ultrasound to study and observe the articulation of particular L2 speech sounds. They first introduce ultrasound applications for speech training and research and compare it against other electronic methods, such as spectrograms and electropalatography. The authors state that ultrasound provides a relatively affordable, non-invasive, and versatile option to other methods, although it does have limitations with respect to what articulatory information can be displayed. They then provide an illustration of the use of this technology for L2 phonology training through a discussion of a study on the training of liquids for Japanese learners of English; this discussion also includes a detailed overview of methodological options for ultrasound research, including participants, equipment, stimuli, and evaluation. Gick et al., conclude the chapter by outlining the limitations of ultrasound research thus far as well as directions for research in the future.

In Chapter 12 ("Technologies for prosody in context: Past and future of L2 research and practice"), Dorothy M. Chun, Debra M. Hardison, and Martha C. Pennington discuss the training of L2 discourse prosody through computer-based technologies. Chun et al. first outline the history of approaches to the study of L2 discourse intonation and prosody, followed by a discussion of current and future approaches to research and teaching of this aspect of the L2 phonology. In the next section of the chapter, the authors outline various methodological approaches to both research and teaching, including an overview of technological tools that have been used in L2 prosody instruction, and they synthesize research findings on both perception and production/perception-based studies of L2 discourse prosody. Future directions for research and training are also outlined, including research that is multimodal (e.g., includes both auditory and visual feedback) and expanding the focus of research/teaching to include gestures and movements and their correlations with discourse prosody. The authors conclude the chapter by outlining challenges in technology-based teaching and research and by describing directions for technology development.

Finally, in Chapter 13 ("Curriculum issues in teaching pronunciation to second language learners"), Tracey M. Derwing examines a number of concerns with regard to course design and implementation. She first outlines considerations that one must take into account before beginning curricular planning and identifies factors that affect success in L2 pronunciation training. Derwing argues that intelligibility, rather than accent reduction, should be the primary aim of L2 pronunciation courses, and she surveys several research studies that examine factors that influence pronunciation and intelligibility. Based on findings that indicate that suprasegmental aspects of speech affect intelligibility to a greater extent than the articulation of individual speech sounds, Derwing advocates for a more central focus on suprasegmentals in the classroom. She then examines a number of instructional issues, including student background, the integration of pronunciation training in the general L2 curriculum, textbooks and technology, and ways of measuring improvement. She also discusses social factors that may affect communication between native speakers and L2 speakers and proposes that a discussion of such factors should be a part of the pronunciation curriculum. She concludes her chapter with a discussion of issues related to teacher preparation and an assessment areas for future research.

CHAPTER 10

Training non-native language sound patterns

Lessons from training Japanese adults on the English /ɹ/–/l/ contrast

Ann R. Bradlow
Northwestern University

Introduction

During native language acquisition the infant progresses from a language-general to a language-specific state. The task of the native language learner can be characterized as a "tuning" of the learner's phonetic system to the distributional patterns of sounds in the ambient language resulting in a self-reinforcing match between native talkers and native listeners (for extensive and up-to-date discussions of native language phonetic and phonological acquisition, see Peperkamp 2003, and accompanying articles). In contrast, during non-native language acquisition the learner must progress from a monolingual to a bilingual state. The task of an adult non-native language learner can be characterized as a shift from a system that is tuned uniquely to the sound structure of the native language (and therefore mis-tuned to the sound structure of the to-be-acquired non-native language) to a flexible system that can be tuned to the sound structure of both the native and the non-native languages (Iverson et al. 2003). While the tuning required for native language speech perception and production acquisition develops spontaneously in response to exposure to the ambient language, the flexibility and "re-tuning" required for the acquisition of non-native language perception and production is usually rather effortful and could presumably benefit from explicit instruction. Accordingly, the goal of non-native training programs is to identify the conditions under which the most general and linguistically functional phonetic and phonological learning can be achieved by adult second language learners.

An important premise of the entire non-native language sound structure training enterprise is that the monolingual adult speech perception and production capabilities are sufficiently plastic to support the acquisition of non-native language sound patterns. Indeed, a major goal of early training studies was to test the hypothesis that sensitivity to acoustic features that are not reinforced by linguistic experience is permanently lost over the course of normal language development (e.g., Pisoni, Aslin, Perey, & Hennessy 1982; Tees & Werker 1984; Werker & Tees 1984; see also Chapter 6 by Strange & Shafer, this volume). Within this theoretical context, numerous non-native language sound structure training studies were conducted on various sound contrasts with listeners from various native language backgrounds. These relatively early training studies tended to adopt auditory training methods that had been developed in the speech and hearing sciences and which focused on increasing sensitivity to fine-grained acoustic differences. Examples of these studies include training English listeners to perceive an "extra", nonphonemic category along a voice onset time continuum (Pisoni, Aslin, Perey & Hennessy 1982), training Canadian French speakers on the English /θ/–/ð/ contrast (Jamieson & Morosan 1986, 1989; Morosan & Jamieson 1989) and training Chinese speakers on word-final /t/ and /d/ in English (Flege 1989). These studies achieved some success in modifying the listeners' responses to the trained stimuli and, in some cases, to untrained stimuli that differed minimally from the trained stimuli thereby providing evidence against a strong interpretation of the hypothesis that the adult speech perception system is no longer plastic. However, at the same time, they began to reveal some limitations on adult abilities to acquire non-native speech sound contrasts. Most noteworthy in this regard is the exceptional difficulty encountered by studies that attempted to train Japanese listeners on the English /ɹ/–/l/ contrast (e.g. Strange & Dittman 1984) using the auditory training techniques that had proved successful in the training studies described above.

Due to its unusual resistance to acquisition, the case of English /ɹ/–/l/ contrast learning by adult Japanese speakers has been particularly well-studied and has effectively served as a testing ground for different non-native speech sound training approaches. Therefore, the goal of this chapter is to compare and contrast training approaches to this notoriously difficult case. This examination of a well-studied case of non-native contrast acquisition will serve as a base from which we will attempt to derive some general principles of non-native speech sound training that can be applied to a wide range of cases.

The chapter will begin by considering the nature of the problem that the English /ɹ/–/l/ contrast poses for adult Japanese speakers. Although the focus of this discussion will be on a particular non-native contrast for learners from a particular native language background, it will serve as a convenient vehicle for pointing out the parameters that need to be considered when developing an adequate description of the learners' "initial state," prior to any training, for all cases of non-native language sound structure learning. Studies that tested various approaches to training Japanese speakers to perceive the English /ɹ/–/l/ contrast will then be presented in the next section. This particular case serves as an effective means of comparing and contrasting training approaches since it has been the subject of investigation for multiple training studies using different approaches, thereby setting the stage for an unusually well-controlled evaluation of different training procedures. Finally, the last section will present some general lessons that we can extract from this case and raise some additional questions that future research should address.

The object of training: What needs to be learned?

Over the course of the past two decades numerous empirical and theoretical developments have made it possible for us to describe in detail the nature and extent of the difficulties encountered by native speakers of one language in response to speech sounds from another language. These advances have consequently made it possible for us to provide adequate descriptions of the task of any second-language learner trying to acquire the sound structure of any non-native language, thereby allowing us to clarify the object of training and to understand exactly what needs to be learned in any particular case. Here we consider in detail the case of perception and production of the English /ɹ/–/l/ contrast by Japanese speakers. This case provides a convenient illustration of the parameters of cross-language comparison that need to be considered in order to understand the nature of non-native speech sound learning.

It has long been noted that native speakers of Japanese have extreme difficulty perceiving and producing the English /ɹ/–/l/ contrast. Several studies have provided experimental data that identify the precise conditions under which the perceptual difficulty of Japanese speakers with English /ɹ/ and /l/ is manifested. The broad conclusion to be drawn from these empirical studies is that under controlled laboratory conditions, Japanese listeners generally exhibit great difficulty identifying and/or discriminating stimuli that exemplify the English /ɹ/–/l/ contrast, but there is considerable variability in perceptual accuracy across individual listeners and across stimulus types. For example, Miyawaki, Strange, Verbrugge, Liberman, Jenkins, and Fujimura (1975) showed that American English listeners exhibited categorical perception along a synthetic /ɹɑ/–/lɑ/ continuum in which only the third formant varied; but Japanese listeners showed continuous perception along this speech continuum.[1] That is, the Americans exhibited a peak in

1. Categorical perception is the phenomenon according to which listeners perceive sounds that differ from each other in terms of equal steps along a continuum as belonging to either one or another category. In contrast, continuous perception of sounds along a continuum is observed when listeners' conscious perception of the sounds is analogous to their physical difference, that is, all differences are perceived and the sounds are not "forced" into one or another category. Categorical perception is typically assessed by testing (a) the consistency with which subjects

discrimination accuracy for stimulus pairs that straddled the category boundary as determined from an identification test, but the Japanese showed uniformly poor (though above chance) discrimination for all pairs along the synthetic speech continuum. In this same study, the American and Japanese listeners performed virtually identically on a discrimination task with non-speech stimuli that consisted of the isolated third formant (F3)[2] component, suggesting that the effect of language background is limited to complex, synthetic speech stimuli and does not extend to relatively simple, non-speech stimuli. Similarly, Iverson et al. (2003) demonstrated that Japanese listeners have some degree of sensitivity to the acoustic differences between English /ɹ/ and /l/ exemplars even if they tend to classify them all as members of a single phoneme category. This pattern of results demonstrates that, rather than having "lost" sensitivity to the acoustic features that cue the English /ɹ/–/l/ contrast at a basic auditory perceptual level, Japanese listeners have instead learned to effectively "ignore" this difference during speech perception resulting in a perceptual space that is "mis-tuned" to the English /ɹ/–/l/ contrast (Iverson et al. 2003).

In response to naturally produced words exemplifying the English /ɹ/–/l/ contrast, Mochizuki (1981) found varying levels of identification accuracy by Japanese listeners depending on the position of the /ɹ/ or /l/ in the words. Identification accuracy ranged from greater than 95% for /ɹ/ and /l/ in word-final position to less than 65% for /ɹ/ in a word-initial consonant cluster. Although the American English listeners also showed some variability in performance as a function of position in the word, the native listeners showed consistently more accurate /ɹ/ and /l/ word identification than the non-native Japanese listeners. Finally, substantial individual listener differences in /ɹ/–/l/ contrast perception have been observed even across native Japanese listeners with apparently comparable language backgrounds (Yamada & Tohkura 1992; MacKain, Best & Strange 1981). Together, these findings indicate that under certain circumstances Japanese listeners are sensitive to the

label stimuli along a continuum as members of two contrasting categories and (b) the accuracy with which subjects discriminate stimuli that are identified as belong to the same category versus stimuli that straddle a category boundary. Greater categorical perception is indicated by (a) consistent labeling of stimuli as belonging to one or the other category even for potentially ambiguous stimuli and (b) relatively good discrimination for stimuli that are identified as belonging to two different categories but relatively poor discrimination for stimuli that are identified as belonging to the same category (even if their acoustic characteristics are quite different).

2. Formants are amplitude peaks in the spectra of vowel and other sonorant sounds, including /r/ and /l/. Formant frequencies are directly related to the articulatory configuration of the vocal tract during speech production. The third formant frequency (F3) is a major cue for the /r/–/l/ distinction with a low F3 frequency providing a strong indicator of the presence of an /r/ articulation.

acoustic differences between English /ɹ/ and /l/; however, in general, their perceptual responses to this linguistic contrast of English are substantially less accurate from a linguistic functional point of view than the responses of native English listeners.

The most obvious source of the Japanese listeners' trouble English /ɹ/ and /l/ perception is at the level of phoneme inventory structure. Whereas English has four contrasting approximant categories (/ɹ, j, w, l/), Japanese has just two contrasting approximants, /j/ and /w/ (Handbook of the IPA 1999; Vance 1987). Thus, when a native Japanese speaker is presented with the English system of sounds, English /j/ and /w/ can be quite well mapped onto Japanese /j/ and /w/, respectively. However, the two English alveolar approximants, /ɹ/ and /l/, do not map well onto any contrasting Japanese approximant pair. Instead, by virtue of similarity on other features (voicing, place of articulation), both of these English phonemes are identified by Japanese listeners rather unsystematically as the Japanese apico-alveolar tap /ɾ/, the Japanese labio-velar approximant /w/ or the Japanese high back unrounded vowel /ɯ̜/ (Best & Strange 1992; Yamada & Tohkura 1992; Mochizuki 1981; Guion, Flege, Akahane-Yamada & Pruitt 2000). Thus, in order to acquire the sound structure of English, a Japanese speaker must learn to organize a poorly distinguished pair of sounds into two contrasting phoneme categories.

Current models of non-native language perception (Perceptual Assimilation Model (PAM): Best 1994, 1995; Best et al. 1988, 2001; Native Language Magnet (NLM) model: Grieser & Kuhl 1989; Kuhl 1991, 1992; Kuhl & Iverson 1995; Speech Learning Model (SLM): Flege 1995, 1999, 2002, 2003) all offer formalizations of this basic conceptualization of the English-Japanese alveolar approximant mapping at the level of phoneme inventory structure (for additional discussion of these models, see also Chapter 2 by Ioup, Chapter 6 by Strange & Shafer, and Chapter 8 by Zampini [the latter for Flege's SLM]). In particular, these models capture the important insight that non-native contrasts are not uniformly poorly perceived. Instead, the difficulty with which a particular non-native contrast is perceived by listeners from a particular native language background depends on the relationship between the phoneme inventories of the two languages in question. All three models agree that the case of Japanese speakers and the English /ɹ/–/l/ contrast is an example of the most difficult kind of non-native contrast to acquire due to the fact that the organizing perceptual framework of the native language (Japanese) results in both English /ɹ/ and English /l/ being identified with the same Japanese category (or categories). Best's Perceptual Assimilation Model (PAM) is explicit in identifying this kind of contrast, a "Single Category" (SC) contrast, as the most difficult kind of contrast for non-native listeners to acquire. According to PAM, SC contrasts are predicted to be more difficult than "Two Category" (TC) or "Category Goodness (CG) contrasts in which the members of a contrasting pair are assimilated by non-native listeners into two separate native categories or into a single native category with different degrees of goodness-of-fit, respectively.

Furthermore, in the production of English /ɹ/ and /l/, the primary acoustic difference between the realization of these phonemes is in the higher formants. For /ɹ/, the third formant frequency can dip below 2000 Hz; whereas, for /l/ the third formant frequency is in the neighborhood of 2400 Hz. Additionally, for /l/ (but not for /ɹ/), the higher formants are substantially reduced in intensity. The exceptionally low F3 frequency for /ɹ/ is related to simultaneous constrictions in the pharyngeal and velar regions of the vocal tract as well as lip rounding.[3] (For additional information regarding the acoustic properties of English liquids see Stevens 1998; Johnson 2003; Ladefoged 2003). In Japanese, the phonemes that are closest to the English liquids, /ɹ/ and /l/, in terms of their acoustic features are the apico-alveolar tap, /ɾ/, the palatal approximant, /j/, the velar approximant, /w/, and the high back unrounded vowel, /ɯ̜/. None of the contrasts represented by this group of phoneme categories (or for that matter, any of the Japanese phonemes) requires auditory attention to the combination of frequency and intensity features that cues the English /ɹ/–/l/ contrast. Therefore, as a consequence of Japanese listeners' lack of experience attending to this particular combination of acoustic-phonetic features, Japanese listeners can be expected to have great difficulty in tasks that require sensitivity to the distinguishing acoustic features of English /ɹ/ and /l/.

It is important to note here that not all novel phoneme contrasts require the same degree of modification at the auditory-perceptual level as the case of Japanese listeners acquiring the English /ɹ/–/l/ contrast. For example, Best et al. (2001) report that the plosive versus implosive voiced bilabial stop contrast of Zulu was treated by the majority of American English listeners in their study as a clear single category (SC) contrast: the American English listeners generally classified both members of the contrasting pair as belonging to the single English /b/ category and showed poor discrimination of the two phones. However, acoustic analyses showed that the primary acoustic differences between these contrasting phonemes in Zulu were that the implosives had higher pitch and F1 frequencies in the early part of the following vowel, higher-amplitude bursts, and substantial pre-voicing in contrast to the small positive VOT for the plosives. This combination of acoustic cues is not entirely unfamiliar to American English listeners and the acquisition of this contrast would require modifications to the category boundary locations along a constellation of dimensions that are already functionally significant for the American English listeners. This situation stands in contrast to the required atten-

3. Note that the lip rounding feature of English /r/ production can be a useful characteristic to stress when teaching English pronunciation.

tion to a new constellation of speech signal dimensions for the acquisition of the English /ɹ/–/l/ contrast by native Japanese speakers.

In terms of speech production, native Japanese speakers have little or no experience with the precise articulatory configurations required for English /ɹ/ and /l/ production (see Gick et al., Chapter 11 of this volume, for discussion of these articulatory configurations). While the separate articulatory gestures involved may be represented in the inventory of native Japanese sounds, including retroflexion, lip rounding and even lateralization, the exact constellation of gestures for English /ɹ/ and /l/ are likely to be novel for native Japanese speakers. Indeed, several experimental studies have demonstrated that native Japanese speakers generally have difficulty producing /ɹ/–/l/ minimal pairs accurately enough for native American English listeners to identify them with a high degree of accuracy (Goto 1971; Sheldon & Strange 1982; Mochizuki 1981). However, somewhat surprisingly, it appears that for many Japanese learners of English, their ability to produce the English /ɹ/–/l/ contrast exceeds their ability to perceive the contrast, particularly in the early stages of acquisition (Yamada, Strange, Magnuson, Pruitt, & Clarke 1994).

In summary, the English /ɹ/–/l/ contrast presents great difficulty for native Japanese speakers due to extensive mismatches between the underlying systems of contrasting approximant categories of the two languages, the particular acoustic-phonetic features that listeners of the two languages have learned to attend to, and the articulatory configurations that talkers of the two languages have learned to produce. The available data on Japanese speakers' perception and production of English /ɹ/ and /l/ clearly demonstrate that this difficulty is general across individuals and is apparent in a range of speech perception and production tasks. However, despite this rather stark contrast between American English and Japanese listeners, Japanese listeners exhibit some sensitivity to the English /ɹ/–/l/ contrast in terms of both perception and production. That is, Japanese listeners are apparently not entirely insensitive to the acoustic and articulatory dimensions that English speakers use to cue this contrast. Thus, the task of learning this novel contrast for a Japanese second language learner is a matter of developing a new organizational framework along existing phonetic dimensions rather than a matter of (re)acquiring sensitivity along acoustic and/or articulatory dimensions that were previously completely unattended to or ignored.

For the reasons described above, it is not surprising that the case of training Japanese speakers to acquire the English /ɹ/–/l/ contrast has been met with remarkable resistance. Few other cases are likely to be as difficult to train since few other cases are likely to require such extensive modification by the learners. Indeed, the early successes of training studies all involved cases that differed from the /ɹ/–/l/ case in some significant way. For example, the early training studies that focused on introducing an extra voicing category (e.g. Pisoni, Aslin, Perey & Hennessy 1982) were probably quite successful with very relatively little training due to the fact that the listeners already had experience with categorization along the relevant acoustic-phonetic dimension (i.e. voice onset time). Similarly, Canadian French speakers being trained on the English /θ/–/ð/ contrast could potentially take advantage of their native language experience with a voicing contrast for other fricatives (Jamieson & Morosan 1986, 1989; Morosan & Jamieson 1989) and Chinese speakers being trained on word-final /t/ and /d/ in English could potentially take advantage of their native language experience with this contrast in other word positions (Flege 1989). Thus, in general, when designing or evaluating a training procedure it is important to first consider the nature of the learners' task in terms of the relevant aspects of the phonetic and phonological structures of the native and the target languages.

Approaches to training: What can be learned?

The first indication that learning should be possible for this difficult case came from reports that Japanese listeners with extended immersion in an English speaking environment generally performed better on English /ɹ/–/l/ perception and production tasks than inexperienced Japanese listeners (MacKain et al. 1981; Flege, Takagi & Mann 1995, 1996; Best & Strange 1992; Yamada et al. 1994; Yamada 1995). Although there are virtually no reports of native-like performance, the fact that performance varies with amount of exposure to English even amongst individuals whose first exposure is at a relatively late stage (beyond childhood) is strong evidence that experience-dependent learning is possible even for this difficult case. Indeed, because of the well-documented difficulty of this particular case in both laboratory and natural settings, it has been upheld as the "gold standard" for proposed training approaches, and has served as a productive testing ground for general principles of learning and claims about adult neural plasticity. The focus here is on perception training procedures; however, it should be noted that the development of production training procedures is an active area of research as well (e.g., Catford & Pisoni 1970; Akahane-Yamada, Adachi & Kawahara 1995; Kewley-Port & Watson 1994; Dalby, Kewley-Port, & Sillings 1998; Dalby & Kewley-Port 1999).

In a seminal study that laid the groundwork for future non-native phoneme contrast training, Strange and Dittman (1984) attempted to train Japanese speakers on the English /ɹ/–/l/ contrast using a training procedure that had proved successful in auditory training studies that aimed to increase listeners' sensitivity to small differences between sounds. In particular, Strange and Dittmann (1984) adopted a training strategy that was used in a study demonstrating that American English listeners could be trained to discriminate within-category differences along a voice onset time continuum (Carney, Widin & Viemeister 1977). The objective of this general training approach is to explicitly draw attention to the acoustic parameters that vary from one end of a synthetic speech continuum to the other and in so doing to enhance discrimination between items along the continuum.

A crucial feature of the overall design of this study was that following the discrimination training (with stimuli along a rock-lock continuum), subjects were tested on a different synthetic continuum (rake-lake), as well as on a minimal pair identification task using naturally produced /ɹ/ and /l/ words. For example, subjects heard "rock", and identified it as either "rock" or "lock." Thus, this training study assessed the extent of any learning on the trained stimuli and task as well as the generalization of this learning to novel stimuli (i.e. stimuli not included in the training set) and a novel task (i.e., a task that was different from the training task).

Subjects were native speakers of Japanese who were recruited from an English as a Second Language program at the University of Minnesota. The subjects ranged in age from 25 to 33 years and had lived in the USA from 5 to 30 months. Although their levels of English proficiency varied widely at the pretest phase, all subjects reported difficulty with English /ɹ/ and /l/ and all were highly motivated to improve their English skills. At the pretest and posttest phases, the subjects performed a minimal pair identification test with naturally produced stimuli (16 pairs of words produced by an adult male native speaker of American English), as well as identification and discrimination tests with the rock-lock stimuli (from the training phase) and a novel (i.e., untrained) rake-lake stimulus continuum. The training task was a same-different discrimination task in which subjects were presented with pairs of stimuli from the synthetic rock-lock continuum and required to respond by labeling a pair as either S (same) or D (different). Immediate feedback after each trial was provided during training. Subjects completed 14–18 training sessions conducted over the course of 3 weeks.

During training the Japanese subjects generally improved in their ability to discriminate stimuli along the synthetic rock-lock continuum. This improvement during training was evident in the posttest phase by a change towards greater categorical perception along the rock-lock continuum for seven of the eight Japanese subjects. However, it is important to note, that the Japanese subjects still differed from American English listeners in terms of their identification consistency and discrimination accuracy for stimuli along this rock-lock continuum. The Japanese subjects also showed more categorical perception along the rake-lake continuum at posttest than at pretest; however, the Japanese subjects exhibited considerably less categorical perception along this untrained rake-lake continuum than along the trained rock-lock continuum.[4] In contrast to this move towards greater cat-

4. See Note 1 above.

egorical perception in response to discrimination training, the Japanese subjects showed no improvement in their ability to identify naturally produced /ɹ/–/l/ minimal pairs from pretest to posttest. In other words, while the discrimination training in this study modified the Japanese subjects' responses to synthetic stimuli, this change did not generalize to naturally produced words.

The Strange and Dittmann (1984) study is an example of a "low variability" training approach since training involved the presentation of stimuli representing only one /ɹ/–/l/ minimal pair as produced by only one synthetic "talker." In a further test of this general, low variability training approach, a recent study investigated whether Japanese listeners would acquire the English /ɹ/–/l/ contrast through initial exposure to maximally differentiated, or exaggerated, category exemplars (i.e. exemplars of the English /ɹ/–/l/ contrast in which the acoustic difference between /ɹ/ and /l/ is maximized) followed by exposure to increasingly natural exemplars (McCandliss, Fiez, Protopapas, Conway, & McClelland 2002). The rationale behind this training procedure is as follows: provided that the exaggerated exemplars are discriminable at the start of training, and that the discrimination of exaggerated exemplars generalizes to less exaggerated exemplars, then by slowly decreasing the acoustic distance between the training stimuli, listeners should eventually be able to discriminate natural exemplars.

The stimuli for this study came from a synthetic /ɹ/–/l/ continuum (road-load or rock-lock) that was created by editing samples of the words as produced by a male native speaker of American English. The continuum was constructed by calculating the spectral distance between the members of the minimal pair (based on a linear predictive coding (LPC) analysis at intervals of approximately 10 msec) and then adjusting the LPC coefficients to interpolate between and extrapolate beyond the two endpoints, yielding a well-sampled, extended /ɹ/–/l/ continuum. The training task was an identification task in which the subject had to identify the initial segment of the test word as /ɹ/ or /l/. The general design of this study tested the effects of two training variables: "adaptive" (i.e. training that begins with exaggerated stimuli and ends with more typical stimuli) versus "fixed" (i.e. training with typical stimuli only), and with feedback versus without feedback during training.

Following training, subjects in the adaptive training group showed more native-like identification and discrimination functions along the trained continuum than subjects in the fixed training group or subjects in the untrained control group. However, there were no significant differences between the two groups of trained subjects (adaptive vs. fixed) nor between either of these trained groups and the untrained control group when tested on a novel continuum (rock-lock for subject trained on road-load, or road-load for subjects trained on rock-lock.) The most dramatic effect revealed by this study was that, regardless of whether the subjects were initially exposed to exaggerated stimuli (adaptive vs. fixed training procedures), subjects who were provided with feedback during training made substantial gains towards establishing distinct /ɹ/ and /l/ categories along both the trained and the generalization continua. Unfortunately, this test of generalization was severely limited in that it did not test generalization to a novel (i.e. untrained) talker or to a novel phonetic environment. It is therefore impossible to determine at this point whether the learning that results from this type of low-variability identification training with feedback is stimulus-specific or stimulus-general.

Although the results of the low-variability, discrimination training procedure of Strange and Dittmann (1984) and those of the low-variability, identification training procedure of McCandliss et al. (2002) showed some success in modifying the Japanese learners' responses to synthetic /ɹ/–/l/ continua, neither provided evidence that laboratory-based training could induce improved recognition of novel, naturally-produced English /ɹ/ and /l/ words. Other studies have tested an alternative, "high variability" training approach that attempts to achieve this goal by exposing subjects to the full range of stimulus variability within each of the contrasting categories that the learner can expect to encounter in the real world. The first attempt at implementing this training approach (Logan, Lively & Pisoni 1991) began as a follow-up to the suggestion of Strange and Dittmann (1984) to expand the training procedure to cover a wider range of training stimuli (see also Jamieson & Morosan 1986 for a similar suggestion). Logan et al. (1991) also noted that this suggestion was consistent with work in visual stimulus classification demonstrating that training on highly variable stimuli promoted more accurate classification of novel, untrained stimuli than training with a low variability stimulus set (Posner & Keele 1968). Moreover, in a departure from the focus on categorical versus continuous perception of /ɹ/–/l/ continua for American English versus Japanese listeners, respectively, the high variability training procedure involves a training task that more closely matches the task of word recognition that occurs in real-world spoken language processing. Specifically, the training task and stimuli require the listeners to classify a wide range of naturally produced and highly variable words exemplifying English /ɹ/ and /l/ into broadly defined categories.

The overall design of the first test of the high variability approach (Logan et al. 1991) included pretest, training and posttest phases. In all phases, the subjects performed a minimal pair identification task in which they heard a single word and had to identify it as either the /ɹ/ or /l/ word from an /ɹ/–/l/ minimal pair. For all tests, the stimuli were naturally produced words by native American English talkers that placed /ɹ/ or /l/ in various positions in the word (e.g. right-light, pray-play, bear-bell, bard-bald). At pretest, the subjects performed the minimal pair identification test without feedback using the word list from Strange and Dittmann (1984) as produced by one male talker. During the training phase the subjects performed the minimal pair identification task with stimuli produced by 5 talkers (3 males and 2 females). The training stimulus set included 68 minimal pairs (a total of 136 stimuli) none of which were included in the pretest. During training, the subjects were provided with immediate feedback. At posttest, all subjects performed the same minimal pair identification test as at the pretest phase. In addition, a subset of the subjects also performed two tests of generalization: the first presented a novel set of words produced by one of the talkers that produced the training stimuli, while the second presented a novel set of words produced by a novel talker. Subjects were 6 native speakers of Japanese who were students at Indiana University. They had lived in the USA from 6 to 36 months.

The results of this training study showed significant minimal pair identification improvement from the pretest phase to the posttest phase for all subjects, and for those subjects who performed the generalization tests, this learning showed some generalization to novel, untrained stimuli and a novel, untrained talker. In a follow-up study Lively et al. (1993) demonstrated similar learning when the extent of the phonetic context variability of the training stimuli was reduced to include only words that place /ɹ/ and /l/ in the most difficult positions in the word (i.e., in pre-vocalic positions where pretest performance is poorest.) Stimuli with /ɹ/ and /l/ in post-vocalic positions were eliminated from the training stimulus set due to high identification accuracy for such words even at pretest. However, this same study demonstrated that a reduction in the extent of talker variability in the training stimulus set to just one talker (instead of five talkers) did not lead to substantial improvements in identification accuracy (however see Magnuson et al. 1995 for evidence that training on some individual talkers can be as effective as multiple talker training). These results suggest that exposure to multiple talkers during training is effective for achieving general, rather than stimulus-specific, learning. However, training may be optimized by focusing only on phonetic environments that are known to be difficult at the pretest phase.

This pattern of results was replicated and extended to monolingual Japanese subjects who had never lived in an English-speaking country, and the perceptual learning that resulted from the high variability training procedure was shown to be retained for a period of at least 6 months with no additional training (Lively et al. 1994). A subsequent study showed that if the training period continued to the point were the average learning curve "leveled off," (i.e. 45 rather than 15 sessions of approximately 30 minutes each), essentially perfect generalization of the perceptual learning to novel, untrained words and to a novel, untrained talker could be attained (Yamada 1993). Finally, the generalized perceptual learning that was induced by the extended (45 session) high variability training procedure transferred from the perceptual domain to improvements in /ɹ/–/l/ contrast production by the Japanese trainees (Bradlow et al. 1997, 1999). While all of the above-mentioned high variability training studies used the same stimulus set and training procedure, a separate study using new stimuli but the same overall high-variability approach, replicated and extended these learning patterns for the English /ɹ/–/l/ contrast to training in combined audio and visual modalities and to both Japanese and Korean speaking adults (Hardison 2003).

Taken together, the series of studies on training Japanese speakers to identify English /ɹ/ and /l/ using a high-variability training approach proved conclusively that robust, linguistically-functional learning can be achieved under laboratory training conditions even for this unusually difficult case. Provided that the training phase continued to the point of saturation, the learning demonstrated by these studies was not specific to the training items, was resistant to decay over time, and extended beyond the perception domain to the production domain. This high degree of success of the high variability training approach stands in contrast to the limited success of the low variability training approach which, as far as could be determined, was stimulus and task specific (i.e., did not generalize to stimuli and task that were not part of the training procedure).[5] It is important to note here that the high variability approach could not have been devised without the groundwork laid by prior low variability training studies. In particular, the overall design of Strange and Dittmann (1984), which emphasized the importance of testing the generalization of training-induced learning beyond the specific stimuli and task used in training, was a critical step in the development of non-native contrast training approaches.

Following the success of the high-variability training approach with the difficult case of training Japanese speakers on the English /ɹ/–/l/contrast, several other non-native contrast training studies adopted the high variability approach and showed similar learning patterns. These studies include training of English listeners on Chinese lexical tone contrasts (Wang, Spence, Jongman & Sereno 1999; Wang, Jongman & Sereno 2003), training of English and Japanese listeners on Hindi dental and retroflex stops (Pruitt 1995), training English listeners on Japanese vowel length contrasts (Yamada, Yamada & Strange 1996), training Chinese listeners on English word-final /t/ and /d/ (Flege 1995), and training English listeners on various German vowel contrasts (Kingston 2003). These studies have all demonstrated substantial learning in response to high variability training and, in those cases that tested generalization, results showed good to excellent generalization to novel, untrained talkers and stimuli. While all of these studies involved a high variability approach with respect to the stimuli used during training (multiple words produced by multiple talkers), they differed somewhat with respect to the training task (identification versus discrimination) and sequence of stimulus presentation during training (unstructured versus gradual introduction of more various and challenging stimuli). In particular, in the training of Chinese

5. For a more recent and direct comparison of training methods for this particular case, see Iverson, Hazan & Bannister 2005.

speakers on English word-final /t/ and /d/, Flege (1995) directly compared two training tasks, identification and categorial discrimination (in which the stimuli presented for discrimination are always members of contrasting categories rather than members of the same phoneme category). Both training tasks resulted in significant learning and generalization to novel stimuli. In the training of English and Japanese listeners on Hindi dental and retroflex stops, Pruitt (1995) adopted a “fading” stimulus presentation scheme, in which training began with a limited set of easily identified stimuli and, as the subject’s performance improved, additional and more challenging stimuli were gradually introduced. In this study, while the listeners from the two native language backgrounds showed different levels of performance at all stages (the Japanese listeners always performed better than the English listeners), both showed significant improvements in response to training and this perceptual learning generalized to novel, untrained stimuli.

General lessons and future directions

As in the study of many physical and psychological systems, it is often highly instructive to consider the extreme cases. The case of Japanese speakers’ difficulties with English /ɹ/ and /l/ has in many respects served this purpose in the development of non-native speech sound training programs and has therefore been the major focus of this chapter. We conclude by identifying three general lessons to be learned from the rich history of research on training Japanese speakers on the English /ɹ/–/l/ contrast.

Lesson 1

Laboratory-based training can lead to successful non-native contrast learning even for the most difficult cases. Even though native-like performance may be an unattainable goal for non-native language sound structure training programs, robust and highly generalized improvements in speech perception and production can be attained by adult learners with extremely limited prior exposure to the target language. This now well-established fact contributes an important line of evidence against a strict interpretation of the hypothesis that, in the absence of early language exposure, certain sensorineural sensitivities are permanently lost. Instead, it appears that the ability to modify speech perception and production patterns is retained well into adulthood. This claim is made explicit in the Speech Learning Model (Flege 1995) and argued for extensively in much of Flege’s recent writings (Flege 1995, 1999, 2002, 2003).

The open questions that current research on the issue of neural plasticity for speech learning should continue to address are: (1) What levels of processing and representation are shaped by early language exposure? That is, where along the pathway from lower level sensorineural encoding to higher level, linguistic processing does the effect of linguistic experience become evident? (See for example, Iverson et al. 2003, Cheour et al. 1998; Bent, Bradlow & Wright 2006) (2) Is early exposure necessary and sufficient to induce native-like speech perception and production of a non-native language, and how does continued native language exposure and use interact with non-native language acquisition? (See for example, Flege & MacKay 2004; Flege, Frieda & Nozawa 1997; Pallier, Bosch & Sebastián-Gallés 1997; Pallier, Colome & Sebastián-Gallés 2001; Mayo, Florentine & Buus 1997) (3) What is the relationship between initial non-native language speech perception abilities and training-induced learning? This question pertains both to the causes and consequences of individual differences across learners from the same native language background, as well as across learners from different native language backgrounds in response to a given non-native language. In the case of the former, variables such as age and conditions of initial exposure are of interest. In the case of the latter, learning patterns generated from models of non-native contrast perception (Flege's Speech Learning Model, Kuhls' Native Language Magnet Model, and Best's Perceptual Assimilation Model) can be identified and tested (see for example, Polka & Bohn 1996; Guion et al. 2000; Bohn & Polka 2001; Kingston 2003).

Lesson 2

The essential goal of non-native language contrast acquisition is accurate recognition of words that exemplify the contrast in the target language rather than native-like patterns of categorization along acoustic-phonetic continua. Although the effect of training with naturally produced words on categorization along the relevant acoustic dimensions has not been examined, it is possible (even probable) that non-native listeners develop functional, non-native language category representations for the purposes of word recognition in the absence of native-like sensitivity to specific acoustic features of the speech signal. Conversely, as demonstrated by the low-variability training reviewed earlier in this chapter, non-native listeners may develop more native-like patterns of categorization but still show highly inaccurate word recognition (e.g. Strange & Dittmann 1984). It is very likely that a key to successful non-native sound structure acquisition is to focus on the perception of even more contextualized speech samples than isolated words, such as full sentences and larger discourse units (e.g. Hirata 2003, 2004). However, the cost of introducing greater processing requirements must be examined in relation to the benefit of presenting the non-native contrast under acquisition in the context of a meaningful linguistic unit instead of in isolation.

Lesson 3

Exposure to highly variable training stimuli promotes, rather than interferes with, non-native contrast acquisition. In particular, exposure to multiple talkers appears to be a highly effective means of ensuring that perceptual learning generalizes to novel talkers. This general principle of the high variability training approach has received further support from studies of speech learning at a more global level than the level of phoneme category contrasts. For example, native English listeners exposed to multiple talkers of Chinese-accented English during training were able to generalize their learning of this particular accent to a novel (i.e. never before heard) talker of Chinese-accented English; whereas, listeners exposed to a single talker during training showed only talker-specific learning (Bradlow & Bent 2003, in press). Similarly, in an American English dialect classification task (in which listeners are asked to identify the region in the USA from which the talker comes) a group of native English listeners who had been exposed to multiple talkers from each region were better able to categorize a set of novel talkers than a comparable group of listeners who had been exposed to just one representative talker from each region (Clopper & Pisoni 2004). This positive effect of the high variability training approach on speech category learning is consistent with exemplar-based models of speech perception (Goldinger 1996; Johnson 1997; Pierrehumbert 2001, 2002, 2003a, 2003b) in which item-specific acoustic-phonetic variability is encoded in the cognitive representation of experienced speech samples. Moreover, the patterns of learning and generalization revealed by dialect/accent and non-native phoneme contrast training studies such as those discussed in this chapter provide crucial information regarding the dimensions over which linguistic and paralinguistic generalizations are formed, and about the structure of an exemplar-based phonetic category system. Thus, speech training studies represent an area of research with unusual importance in both theoretical and practical arenas.

Suggested readings

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tohkura, Y. 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

Flege, J. E. 2002. Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (eds.), *An integrated view of language development: Papers in honor of Henning Wode,* Trier, Germany: Wissenschaflicher Verlag Trier, 217–244.

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

Strange, W. (ed.). 1995. *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press.

Strange, W., & Dittmann, S. 1984. Effects of discrimination training on the perception of /r–l/ by Japanese adults learning English. *Perception and Psychophysics* 36: 131–145.

References

Akahane-Yamada, R. 1996. Learning non-native speech contrasts: What laboratory training studies tell us. In *Proceedings of the third joint meeting of the Acoustical Society of America and Acoustical Society of Japan*, Honolulu, Hawaii.

Akahane-Yamada, R., Adachi, T., & Kawahara, H. 1997. Second language production training using spectrographic representations as feedback. *Journal of the Acoustical Society of Japan* 18: 341–343.

Bent, T., Bradlow, A. R., & Wright, B. A. 2006. The influence of linguistic experience on pitch perception in speech and non-speech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 32(1), 97–103.

Best, C. T. 1995. A direct realist perspective on cross-language speech perception. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, Timonium, MD: York Press, 171–206.

Best, C. T. & Strange, W. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20: 305–350.

Best, C. T., McRoberts, G. W., & Goodell, E. 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.

Best, C. T. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. Goodman & H. C. Nusbaum (eds.), *The development of speech perception: The transition from speech sounds to spoken words*, Cambridge, MA, MIT Press, 167–224.

Best, C. T., McRoberts, G. W., & Sithole, N. M. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14: 345–360.

Bohn, O.-S., & Polka, L. 2001. Target spectral, dynamic spectral and duration cues in infant perception of German vowels. *Journal of the Acoustical Society of America* 110: 504–515.

Bradlow, A. R., & Bent, T. 2003. Listener adaptation to foreign-accented English. In M. J. Sole, D. Recasens, & J. Romero (eds.), *Proceedings of the XVth International Congress of Phonetic Sciences*, Barcelona, Spain, 2881–2884.

Bradlow, A. R., & Bent, T. (in press). Perceptual adaptation to non-native speech. *Cognition*.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. 1999. Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics* 61: 977–985.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tohkura, Y. 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.

Carney, A. E., Widin, G. P., & Viemeister, N. F. 1977. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America* 62: 961–970.

Catford, J. C., & Pisoni, D. B. 1970. Auditory vs. articulatory training in exotic sounds. *The Modern Language Journal* 54: 477–481.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A, Allik, J., Alho, K., & Näätänen, R. 1998. Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience* 1: 351–353.

Clopper, C. G., & Pisoni, D. B. 2004. Effects of talker variability on perceptual learning of dialects. *Language and Speech* 47: 207–239.

Dalby, J., & Kewley-Port, D. 1999. Explicit pronunciation training using automatic speech recognition. In M. Holland (ed.), *Tutors that listen: Speech recognition for language training, Special issue of the Journal of the Computer Assisted Language Learning Consortium (CALICO)* 16(5): 425–445.

Dalby, J., Kewley-Port, D., & Sillings, R. 1998. Language- specific pronunciation training using automatic speech recognition technology. In *Proceedings of the European Speech Communication Association Conference on Speech Technology in Language Learning*, 25–28.

Flege, J. E. 1995. Second-language speech learning: Theory, findings and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, Timonium, MD: York Press, 233–272.

Flege, J. E. 1995. Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics* 16: 425–442.

Flege, J. E. 1999. Age of learning and second-language speech. In D. Birdsong (ed.) *New perspectives on the critical period hypothesis for second language acquisition*, Mahwah, NJ, Lawrence Erlbaum, 101–132.

Flege, J. E. 2002. Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (eds.), *An integrated view of language development: Papers in honor of Henning Wode*, Trier, Germany: Wissenschaflicher Verlag Trier, 217–244.

Flege, J. E. 2003. Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities*, Berlin: Mouton de Gruyter, 319–355.

Flege, J. E., & MacKay, I. R. A. 2004. Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26: 1–34.

Flege, J. E., Frieda, E. M., & Nozawa, T. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics* 25: 169–186.

Flege, J. E., Takagi, N., & Mann, V. 1995. Japanese adults can learn to produce English /ɹ/ and /l/ accurately. *Language and Speech* 38: 25–55.

Flege, J. E., Takagi, N., and Mann, V. 1996. Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *Journal of the Acoustical Society of America* 99: 1161–1173.

Gick, B., Bernhardt, B., Bacsfalvi, P., & Wilson, I. 2008. Ultrasound imaging applications in second language acquisition. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

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.

Goto, H. 1971. Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia* 9: 317–323.

Grieser, D., & Kuhl, P. 1989. Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology* 25: 577–588.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. 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.

Handbook of the International Phonetic Association. (1999). Cambridge, UK: Cambridge University Press.

Hardison, D. 2003. Acquisition of second-language speech: Effects of visual cues, context and talker variability. *Applied Psycholinguistics* 24: 495–522.

Harnesberger, J. D. 2001. On the relationship between identification and discrimination of non-native nasal consonants. *Journal of the Acoustical Society of America* 110: 489–503.

Hirata, Y. 2003. Learning to form new L2 phonetic categories in sentence contexts. In M. J. Sole, D. Recasens, & J. Romero (eds.), *Proceedings of the XVth International Congress of Phonetic Sciences*, Barcelona, Spain, 515–518.

Hirata, Y. 2004. Training native English speakers to perceive Japanese length contrast in word versus sentence contexts. *Journal of the Acoustical Society of America* 116: 2384–2394.

Ioup, G. 2008. Exploring the role of age in the acquisition of a second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87: B47–B57.

Iverson, P., Hazan, V., & Bannister K. 2005. Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/–/l/ to Japanese adults. *Journal of the Acoustical Society of America* 118: 3267–3278.

Jamieson, D. G. 1995. Techniques for training difficult non-native speech contrasts. *Proceedings XIIIth Congress of Phonetic Sciences*, Stockholm.

Jamieson, D. G., & Moore, AE 1991. Generalization of new speech contrasts trained using the fading technique. *Proceedings XIIth International Congress of Phonetic Sciences*, Aix-en-Provence.

Jamieson, D. G., & Morosan, D. E. 1986. Training non-native speech contrasts in aduts: Acquisition of the English /θ/–/ð/ contrast by francophones. *Perception and Psychophysics* 40: 205–215.

Jamieson, D. G. & Morosan, D. E. 1989. Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology* 43: 88–96.

Jamieson, D. G., & Rvachew, S. 1992. Remediating speech production errors with sound identification training. *JSLPA* 16: 201–210.

Jamieson, D. G., & Rvachew, S. 1994. Perception, production and training of new consonant contrasts in children with articulation disorders. *Proceedings International Conference on Spoken Language Processing*, Yokohama.

Johnson, K. 1997. Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (eds.), *Talker variability in speech processing*, San Diego: Academic Press, 145–166.

Johnson, K. 2003. *Acoustic and auditory phonetics*, 2nd Edition. Malden, MA: Blackwell Publishing.

Kewley-Port, D. & Watson, C. S. 1994. Computer assisted speech training: Practical considerations. In A. Syrdal, R. Bennett & S. Greenspan (eds.), *Applied speech technology*, Boca Raton, FL: CRC Press, 565–582.

Kingston, J. 2003. Learning foreign vowels. *Language and speech* 46: 295–349.

Kuhl, P. K., & Iverson, P. 1995. Linguistic experience and the "perceptual magnet effect." In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, Timonium, MD: York Press, 121–154.

Kuhl, P. K. 1991. Human adults and human infants show a ”perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* 50: 93–107.

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

Ladefoged, P. 1996. *Elements of acoustic phonetics*, 2nd Edition. Chicago: The University of Chicago Press.

Lively, S. E., Logan, J. S., & Pisoni, D. B. 1993. Training Japanese listeners to identify English /r/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* 94: 1242–1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tokhura, Y., & Yamada, T. 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., & Pruitt, J. S. 1995. Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, Timonium, MD: York Press, 351–378.

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

MacKain, K. S., Best, C. T., & Strange, W. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics* 2: 369–390.

Magnuson, J. S., Yamada, R. A., Tohkura, Y., & Bradlow, A. R. 1995. Testing the importance of talker variability in non-native speech contrast training. *Journal of the Acoustical Society of America* 97(5), Pt. 2: 3417.

Mayo, L. H., Florentine, M., & Buus, S. 1997. Age of second-language acquisition and perception of speech in noise. *Journal of Speech and Hearing Research* 40: 686–693.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. 2002. Success and failure in teaching the [r]–[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience* 2: 89–108.

McClaskey, C. L., Pisoni, D. B., & Carrell, T. D. 1983. Transfer of training of a new linguistic contrast in voicing. *Perception and Psychophysics* 34: 323–330.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics* 18: 331–340.

Mochizuki, M. 1981. The identification of /ɹ/ and /l/ in natural and synthesized speech. *Journal of Phonetics* 9: 283–303.

Morosan, D. E., & Jamieson, D. G. 1989. Evaluation of a technique for training new speech contrasts: Generalization across voices, but not word-position or task. *Journal of Speech and Hearing Research* 32: 501–511.

Pallier, C., Bosch, L., & Sebastian-Galles, N. 1997. A limit on behavioral plasticity in speech perception. *Cognition* 64: B9–B17.

Pallier, C., Colome, A., & Sebastian-Galles, N. 2001. The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science* 12: 445–449.

Peperkamp, S. 2003. Phonological acquisition: Recent attainments and new challenges. *Language and Speech* 46: 87–113.

Pierrehumbert, J. B. 2002. Word specific phonetics. In C. Gussenhoven & N. Warner (eds.), *Laboratory Phonology 7*, Berlin & New York: Mouton de Gruyter, 101–140.

Pierrehumbert, J. B. 2003. Probabilistic theories of phonology. In R. Bod, J. B. Hay, & S. Jannedy (eds.), *Probability theory in linguistics*, Cambridge, MA: The MIT Press, 177–228.
Pierrehumbert, J. B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (eds.), *Frequency effects and the emergence of linguistic structure*, Amersterdam: John Benjamins, 137–157.
Pierrehumbert, J. B. 2003. Phonetic diversity, statistical learning and acquisition of phonology. *Language and Speech* 46: 115–154.
Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. 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. 1991. Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of Acoustical Society of America* 89: 2961–2977.
Polka, L. 1992. Characterizing the influence of native experience on adults speech perception. *Perception and Psychophysics* 52: 37–52.
Polka, L., & Bohn, O.-S. 1996. A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America* 100: 577–592.
Posner, M., & Keele, S. 1968. On the genesis of abstract ideas. *Journal of Experimental Psychology* 77: 353–363.
Pruitt, J. S. 1995. Perceptual training on Hindi dental and retroflex consonants by native English and Japanese speakers. *Journal of the Acoustical Society of America* 97: 3417.
Sheldon, A., & Strange, W. 1982. The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3: 243–261.
Stevens, K. N. 1998. *Acoustic phonetics*. Cambridge, MA: The MIT Press.
Strange, W., & Dittmann, S. 1984. Effects of discrimination training on the perception of /r–l/ by Japanese adults learning English. *Perception and Psychophysics* 36: 131–145.
Strange, W., & Shafer, V. L. 2008. Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.
Tees, R. C., & Werker, J. F. 1984. Perceptual flexibility: Maintenance or recovery of the ability to discriminate nonnative speech sounds. *Canadian Journal of Psychology* 38: 579–590.
Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. 1997. Central auditory system plasticity: Generalization to novel stimuli following listening training. *Journal of the Acoustical Society of America* 102: 3762–3773.
Vance, T. 1987. *An introduction to Japanese phonology*. Albany, NY: State University of New York Press.
Wang, Y., Jongman, A., & Sereno, J. A. 2003) Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America* 113: 1033–1043.
Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. 1999. Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America* 106: 3649–3658.
Werker, J. F., & Tees, R. C. 1984. Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America* 75: 1866–1878.
Yamada, R. A. 1995. Age and acquisition of second language speech sounds: Perception of American English /ɹ/ and /l/ by native speakers of Japanese. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, Timonium, MD: York Press, 305–320.

Yamada, R. A. 1993. Effect of extended training on /ɹ/ and /l/ identification by native speakers of Japanese. *Journal of the Acoustical Society of America* 93: 2391.

Yamada, R. A., & Tokhura, Y. 1992. The effects of experimental variables on the perception of American English /ɹ/ and /l/ by Japanese listeners. *Perception and Psychophysics* 52: 376–392.

Yamada, R. A., Strange, W., Magnuson, J. S., Pruitt, J. S., & Clarke, W. D. 1994. The intelligibility of Japanese speakers' production of American English /r/, /l/, and /w/ as evaluated by native speakers of American English, ATR Human Information Processing Research Laboratories.

Yamada, T., Yamada, R. A., & Strange, W. 1996. Perceptual learning of Japanese mora syllables by native speakers of American English. *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, Japan.

Zampini, M. L. 2008. L2 speech production research: Findings, issues, and advances. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

CHAPTER 11

Ultrasound imaging applications in second language acquisition

Bryan Gick,[1] Barbara May Bernhardt,[2]
Penelope Bacsfalvi,[3] and Ian Wilson[4]
[1]University of British Columbia, Haskins Laboratories /
[2,3]University of British Columbia / [4]University of Aizu

Introduction

Ultrasound imaging has been used for decades as a tool for direct measurement of the tongue for speech research (e.g., Kelsey, Woodhouse & Minifie 1969; Skolnick, Zagzebski & Watkin 1975; Zagzebski 1975). However, with recent improvements in the image quality and affordability of ultrasound systems, possible applications of ultrasound to second language (L2) acquisition are only now beginning to be explored. This chapter discusses current directions in applying ultrasound to both research and pedagogical issues in L2 acquisition and is organized as follows. First, a brief description of ultrasound imaging, along with examples of its application for speech research, are given. The next section provides an overview on the use of technology in pronunciation training and instruction and identifies major research contributions in this area. Methods for conducting speech research using ultrasound imaging are then explained in detail, and several examples of recent and current studies are described. The chapter concludes with a discussion of some of the limitations of ultrasound research and a consideration of promising avenues for future research.

Background

An ultrasound machine emits ultra-high frequency sound through a transducer or "probe" containing piezoelectric crystals. When this transducer is held against the skin of the neck, the sound travels through the tongue and is reflected back to the transducer, resulting in echo patterns from which 2-dimensional images of the tongue surface are reproduced, as shown in Figure 1. These images can be viewed continuously on the machine itself for visual feedback, or recorded to video for later analysis. Because ultrasound is not able to image through bone or air, it can only allow visualization of the tongue and not, for example, the palate, jaw or rear pharyngeal wall. However, it is able to image the entire length of the moving tongue (sagittally, or along any 2-dimensional axis), and to do so at high temporal resolution (30 frames/sec or more), and with little or no discomfort or danger to the subject.

Perhaps the most obvious application of ultrasound in the pedagogical realm is to provide visual biofeedback in the teaching of challenging speech sounds. Other methods of articulatory visual feedback training have been shown to be effective in previous studies of L2 teaching (Catford & Pisoni 1970). However, tools providing direct visual biofeedback of articulation have traditionally been too expensive, slow, hard to use, or invasive for pedagogical purposes. With the cost of ultrasound systems coming within reach of many laboratories and practitioner groups, and an increase in portability and image quality, ultrasound has become a feasible tool for L2 applications. Recent speech therapy studies with hearing-impaired speakers (Bernhardt, Gick, Bacsfalvi & Ashdown 2003) and with speakers who have delayed acquisition of /r/ (Adler-Bock, Bernhardt, Gick & Bacsfalvi

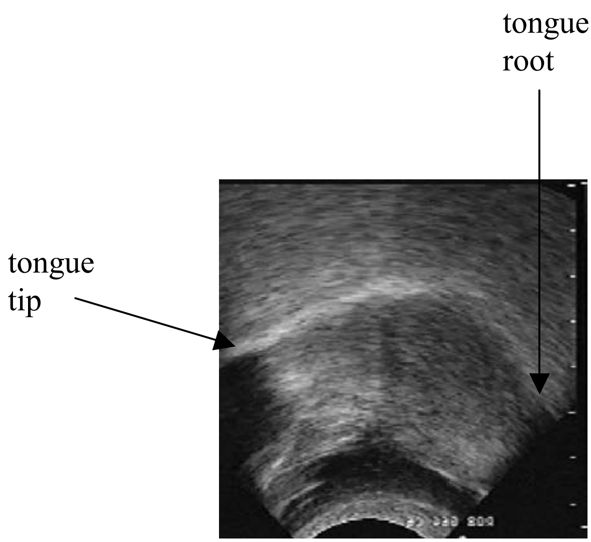


Figure 1. Example of a midsagittal ultrasound image of the tongue, showing the location of the tongue tip and root, the "shadow" of the jaw or sublingual cavity (below the tip), the "shadow" of the hyoid bone (below the root), and the arc at the bottom of the image indicating the location where the head of the transducer contacts the skin of the subject's neck.

2007) have shown that visual feedback therapy using ultrasound can facilitate the acquisition of articulatory targets across a wide range of speech sounds. Similar techniques described below are currently being applied to L2 learners.

In the research realm, beyond evaluation of the pedagogical efficacy of ultrasound as a learning tool, ultrasound provides the ability to measure articulator positions directly, allowing a finer-grained view of speech production and control. One area where this is of obvious interest for L2 acquisition is in describing the physical details of difficult or unusual sounds in specific languages to help facilitate in their learning (e.g., English /r/ as discussed above). Another area of particular relevance to L2 acquisition is that of language-specific "articulatory settings" (Honikman 1964). While these settings have long been discussed in the pedagogical literature (see Collins & Mees 1995), they have proven elusive to measurement. Recent imaging studies have, however, uncovered these settings through measuring language-specific postures held during non-speech segments between utterances (Gick, Wilson, Koch & Cook 2004; Wilson 2006). Ultrasound imaging will allow further study of this phenomenon across speakers of different languages, and will help to feed pedagogical programs advocating the direct teaching of articulatory setting (Mompeán González 2003).

Review of previous literature

The methods and status of pronunciation teaching have fluctuated greatly in the last 50 years (see Morley 1991, and Celce-Murcia, Brinton & Goodwin 1996, for excellent reviews, as well as Chun, Hardison & Pennington, this volume). In the 1940s to the early 1960s, when the audiolingual method of language teaching was the primary one in North America, the pronunciation component was a high priority, with a bottom-up focus (i.e., a focus on sound segments as the building blocks). From the late 1960s to the mid 1980s, when communicative competence and task-based methodologies were heavily promoted, pronunciation teaching was overshadowed by a focus on other areas. From the mid 1980s through the 1990s, pronunciation teaching was revitalized, especially with the realization of the salience of teaching suprasegmentals (i.e., stress, rhythm, and intonation), a top-down approach, and a call for the teaching of articulatory setting (including voice quality or voice-setting). A major development in the 1990s was the increasing popularity of computer-aided pronunciation (CAP) pedagogy (see Chapter 12, this volume, by Chun, Hardison & Pennington for a detailed discussion of CAP).

Electronic methods of teaching pronunciation have been used at least as far back as the early 1950s, shortly after the first commercially available sound spectrograph, the "Sona-Graph", was produced in 1951. Locke (1954: 420) reports that Pierre Delattre was already using spectrograms to teach pronunciation of French vowels, and Locke himself used spectrograms to teach timing, diphthongs, and aspiration. At that time, however, a real-time spectrograph had not yet been designed and so these methods simply provided a record of a student's speech, not online feedback. More CAP methods exist now enabling the pronunciation student to receive visual information, either dynamic or static, about his/her pronunciation. This visual information can take the form of after-the-fact analyses of one's pronunciation, e.g. formants, intonation contours, VOT, etc., or it can be instant biofeedback, either articulatory or acoustic. Anderson-Hsieh (1996) refers to the latter as *electronic visual feedback* (EVF). Most means of EVF provide acoustic information, as opposed to direct articulatory information. It is left to the student and/or teacher to interpret the mapping from the acoustic information provided to the articulatory adjustments that are demanded. In some cases this is not difficult, e.g. it is usually a simple matter to adjust the duration of a segment or the pitch of one's voice, but in other cases the mapping is not very transparent due to the non-linear relationship between vocal tract configurations and acoustic output, e.g. learning what to do to lower the third formant for production of /r/ in English (Guenther, Espy-Wilson, Boyce, Matthies, Zandipour & Perkell 1999; Lambacher 1999).

Articulatory information and feedback have often been used effectively in L2 teaching and learning. Commonly applied methods include the use of direct articulatory instruction and textbook figures of the vocal tract (e.g., Catford & Pisoni 1970; Kelly 2000), the use of a mirror for immediate articulatory feedback (e.g. Clawson 1907: 51; Dale & Poms 1994), encouraging students to concentrate on tactile and proprioceptive feedback (e.g. Acton 1984; Catford 1987; Celce-Murcia, Brinton & Goodwin 1996), and even using a ruler to monitor lip aperture (Odisho 2003: 89). Catford and Pisoni (1970) found that when teaching subjects new sounds, giving the subjects articulatory instruction and having them silently practice was more effective than simply having them listen and mimic. This advantage also carried over to the realm of speech perception as subjects given articulatory training also showed more proficiency at identifying the new sounds they were learning to produce. The results from Yule and Macdonald's (1994) study of 23 Chinese speakers emphasize the great degree of variability in learners' results after different types of pronunciation teaching (for a detailed discussion of L2 pronunciation teaching, see Chapter 13 by Derwing, this volume). One of the few methods of EVF that provides direct and immediate feedback of articulatory information is electropalatography (EPG), a method that has the subject speak with a prosthetic palate in place in his/her mouth. The palate has sensors that monitor the place of contact of the tongue with the palate and this information is displayed on a computer in real-time. This has been used successfully with hearing-impaired subjects and in other clinical applications (see Bernhardt, Gick, Bacsfalvi & Ashdown 2003). However, primarily because of the high cost and time investment required to have custom pseudopalates made for each subject or student, EPG has not been widely used to teach pronunciation to normally hearing L2 learners.

Ultrasound imaging addresses many of the shortcomings of previous EVF methods for L2 applications, being relatively affordable, non-invasive, safe, portable, quick, and versatile, while offering high-dimensional continuous data to be viewed and/or collected. This method has the potential to contribute to the teaching of pronunciation through both a top-down method (i.e., by shedding more light on underlying articulatory setting) and a bottom-up method (i.e., by enabling learners to view real-time images of their tongues as they produce individual sounds).

One example of a typical application of ultrasound imaging to pronunciation teaching involves English /r/. The /r/ sound can be particularly difficult to teach because it involves multiple constrictions (pharyngeal, palatal and labial; Delattre & Freeman 1968) and, as Lambacher (1999) points out, because the labial constriction hides the tongue from view. In a recent intervention study using ultrasound to provide visual articulatory feedback to adolescent English speakers with delayed mastery of /r/, Adler-Bock, Bernhardt, Gick & Bacsfalvi (2007) found that ultrasound allowed this complex sound to be broken down into its individual component movements, enabling learners to experience success at various componential levels on their way to mastering production of the /r/ without having to master the entire sound. In the end, this technique helped learners to make dramatic progress with a challenging speech target in a very short time. Techniques and issues for research and pedagogical applications will be discussed in detail in the following section.

Research methods for ultrasound imaging in L2 acquisition

Increased access to ultrasound imaging will enable advances in certain aspects of sound acquisition and production in L2 research and pedagogy. Aspects of production that were previously inferred from partial or indirect data can now be viewed directly. Because it is non-invasive and portable, and provides an easily interpretable signal, ultrasound technology lends itself well to use in the clinic or classroom (for a description of some field applications of ultrasound, see Gick 2002). While there are many possible applications for ultrasound imaging in L2 research and pedagogy, the present section focuses on describing the details of experiment design for L2 intervention studies, and briefly describes the methods used in a pilot study of Japanese learners of English.

Single participant design

Researchers interested in outcomes measures will find the single participant design merges nicely with the goals of ultrasound intervention studies. Single participant design allows for more focus on individual data, individual variation, and more detail, all of which are often applicable to L2 learning situations. In a group design, individual variation may be lost, and participants often need to be perfectly matched for a large number of criteria (e.g., age, education, language background and experience with the second language). Single participant research uses an approach that repeatedly measures the dependent variables from individual participants (Morgan & Morgan 2001). The dependent variables in such a design would consist of the targets to be learned (e.g., vowels, consonants, or suprasegmentals). The aspects of speech production to be analyzed would include: (a) articulator position and accuracy of segments, and (b) speech intelligibility and accuracy of production. Articulatory accuracy can be measured using graphical analysis software, such as NIH Image (http://rsb.info.nih.gov/nih-image/Default.html), ImageJ (http://rsb.info.nih.gov/ij/) or more specialized ultrasound-specific software such as Ultrax (developed at UBC by S. Rahemtulla and B. Gick; see http://www.linguistics.ubc.ca/isrl; see Figure 2), while intelligibility can be measured by listener judgments (e.g., Bernhardt, Bacsfalvi, Gick, Radanov & Williams 2005; see also Chapter 7 by Munro, this volume). A changing-criterion design with replication across targets using a multiple probe strategy is a powerful design for this type of study (Richards, Taylor, Ramasamy & Richards 1999). A changing criterion design allows the clinician to change the criterion gradually in a step-wise fashion, demonstrating learning at each step of the intervention. In this way, there is no question that the success is due to the intervention.

The design of a typical single-participant intervention study has three phases: (a) a baseline, (b) the intervention, and (c) a follow-up. The functional relationship between the independent variable and the dependent variables will be documented through step-wise improvement in speech production that matches the phases and sub-phases of the research design. Criteria during the intervention phase will be changed when the participant meets the criteria for three consecutive sessions. Intelligibility will be measured at each session. This will occur two-thirds of the way through the session after the client has "warmed-up" and before fatigue begins. Criteria will be met when the participant produces seven out of ten target productions during a sub-phase. Reliability is addressed through repetition of the experiment over many participants. In general, one needs to assure that the data are consistent across participants (Huck 2000). Aside from repetition, interobserver agreement ensures that the process has been fair, ethical and rigorous (Richards, Taylor, Ramasamy, & Richards 1999).

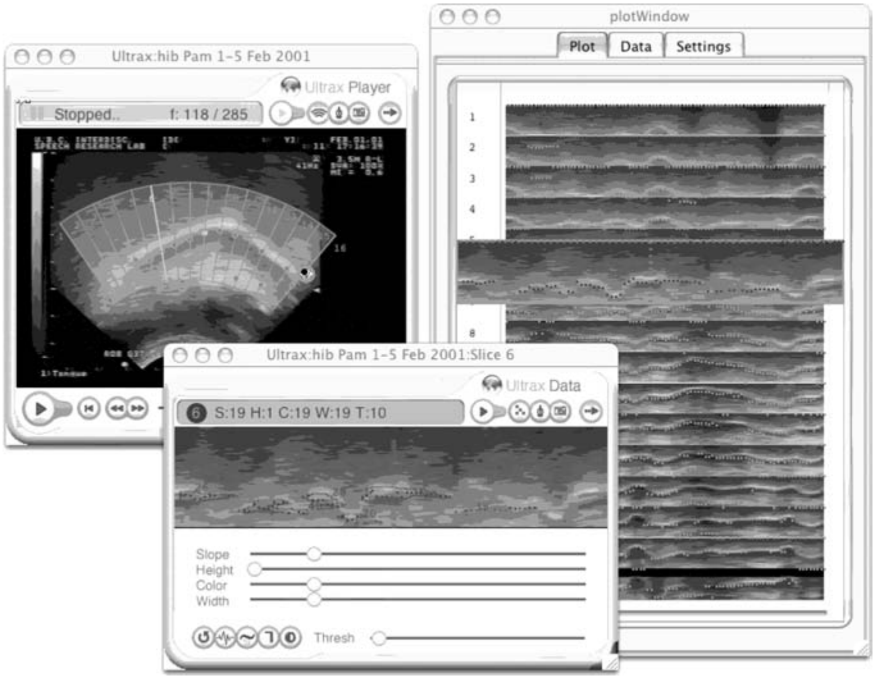


Figure 2. Example of Ultrax ultrasound analysis software. The left image shows a mid-saggital image of the tongue overlaid with an array of measurement lines; the center image shows the control window for edge detection settings; the right image tracks the movement of the tongue along each measurement line over time throughout an utterance.

Equipment

The primary piece of equipment needed is the ultrasound machine. For laboratory applications, any large hospital machine will do, though more recent models tend to have superior image quality. For portable applications see, for example, http://www.sonosite.com for an example of a very small portable unit. Other personal computer-based units can be adapted to field use using a laptop computer. All ultrasound machines require a transducer, and it is important to choose one that is appropriate for imaging the tongue. Our group has obtained the best results using endo-vaginal or pediatric intercostal transducers. These transducers have very small heads with sharp convex angles (120–180 degrees). This allows for a small contact area near the bend in the neck, avoiding several problems such as interference with jaw movement, excessive transducer displacement because of lingual floor muscles, and obscuring of the tongue tip from "shadows" cast by the jaw or the sublingual cavity. The one drawback of the endo-vaginal transducer is that the handle tends to be quite long, which can become awkward in space-limited situations (especially with small individuals). A chair with supports for the arms and head is also needed, although in field conditions, a wall can be used effectively for support and reduction of head movement (Gick, Bird & Wilson 2005). If articulatory data will be subject to quantification and/or measurement, in addition to stabilizing the head, a device should be used to hold the transducer (e.g., a tabletop or floor-mounted microphone stand, a mechanical arm such as that of a dental or ophthalmic chair, or a specially designed helmet). If images are only being used for biofeedback, the transducer may be hand-held by the subject or the investigator (see Gick, Bird & Wilson 2005 for further details regarding field applications and controls). Be aware that participants will often fatigue after 20 to 30 minutes from maintaining a relatively constant position and will need breaks during long sessions for rest and hydration. Finally, recording equipment is needed, including acoustic recording equipment and possibly video equipment.

Stimuli

1. For pre-chosen targets:
   If the target sound has been pre-determined then stimuli lists can be created based on those sounds. The target sound should occur word initially, medially, and finally in different phonetic contexts. Each context should be repeated at least ten times, distributing like tokens across the recording session to avoid list effects, and minimize any movement effects of both transducer and head.
2. For unknown goals:
   If the participant is unknown to the researcher, a broader set of data should be collected. Once again tokens should be distributed across the recording session. A list of words may then be created that gives a wide range of L2 consonants and vowels in a variety of contexts. In addition, the investigator should be aware of phonetic contexts that may influence the shape or position of the target sound.

Evaluation

A rating scale is effective in quantifying how much a participant's speech intelligibility has improved over the period of the experiment. Target sounds may be measured as individual sounds or in words in word-initial, medial and final positions (including in clusters). Productions may be judged by the investigators or by everyday native-speaking listeners using, e.g., a four-point Likert-type scale: 1 (exactly on target), 2 (in category), 3 (somewhat), 4 (not at all). While points 1 and 4 are clear, points 2 and 3 need further explanation. Point 2 (in category) indicates that a speech sound, for example /r/, would have most of the components of /r/

but may lack a crucial component or have a component of another sound, e.g., a raised tongue body. Point 3 (somewhat) would indicate that there is some rhotic quality present in the sound, but that all components of the /r/ are inaccurate, e.g., the tongue is too retracted in the pharynx, there is excessive lip rounding, and there is no retroflexed or bunched anterior gesture. Therapists, L2 teachers and investigators can use ultrasound outside of the experiment or training session to train themselves in the perception of such mismatches with the target. If the experimenters are the ones evaluating the productions, steps should be taken to ensure that sufficient inter-observer agreement is attained.

Criteria should be determined prior to evaluation, each of which represents a step in the changing criterion design. In the case of English /r/, for example, four criteria (C1, C2, C3, and C4) may be set and measured by the researchers: C1 (tongue *root retraction*), C2 (tongue *grooving*), C3 (*palatal constriction*), and C4 (*S-shape configuration* for tongue).

Pilot experiment: Using ultrasound in L2 speech sound training

In order to test the potential utility of ultrasound in L2 speech sound training, a preliminary single-session investigation was conducted with three Japanese linguistics student participants who had recently arrived in North America, facilitated by the four authors of this paper (three native speakers of Canadian English, and one native speaker of American English). Each subject participated in a single one-hour-long session with the investigator team for assessment, training, and post-assessment of their production of the English approximants /l/ and /r/ (for a detailed discussion of other training methods for this contrast, see Chapter 10 by Bradlow, this volume).

Pre- and post-training ultrasound recordings of /r/ and /l/ were made using an Aloka ProSound SSD-5000 ultrasound machine with a UST-9118 endo-vaginal 180-degree convex array transducer held in position using a fixed mechanical arm. Target sounds were elicited in word-initial, word-medial, and word-final positions in six vowel contexts (a variety of front, back, low and high vowels). Word-initial and word-final stimuli consisted of CV or CVC syllables; word-medial stimuli consisted of CVCV words. The randomized word list was repeated ten times pre- and post-training, with each word uttered in the carrier phrase "See X be". During the initial assessment, two of the authors phonetically transcribed on-line to identify contexts in which the participants' pronunciations of the two English approximants needed the most improvement.

The initial assessment showed that all three participants could already produce an English-sounding /l/ or /r/ in at least some phonetic context, with variability among the speakers in degree of proficiency with these targets. One speaker's /r/ was at 100% accuracy; however, this speaker showed neutralization of back low and mid vowels in the context of post-vocalic /l/. Thus, these contexts for /l/ became the training targets. Another speaker's /l/ was 100% accurate, but this speaker showed inconsistent production of /r/ across all word positions, with medial position showing the greatest difference from English. Medial context was the primary focus of training for this second speaker, although /r/ was targeted in all word positions. The third speaker produced /r/ only in post-vocalic position after /ɑ/ and /ɔ/, and /l/ only pre-vocalically. For the third speaker, the /ɑr/ and /ɔr/ productions were used as anchors to address other postvocalic and word-initial /r/ productions. The /l/ was not targeted during ultrasound training, but the participant was given verbal instructions for self-correction at the end of the session using verbal cues only.

For the training part of the session (about 30 minutes), the participants were first shown their best and most troublesome productions from the ultrasound video-recordings. They were asked to compare their productions (both in drawings and verbally) with images produced by the authors in terms of (a) general shape of the tongue, and (b) specific shapes and movements of various parts of the tongue – tip, blade, body, dorsum, and root. In other work in the Interdisciplinary Speech Research Laboratory with adolescents with speech impairments, it has been found effective to have the participant engage intellectually in the treatment process, reflecting on the details of the articulation, and sub-dividing the tongue into relevant areas for shape and movement (Adler-Bock, Bernhardt, Gick & Bacsfalvi 2007; Bernhardt, Gick, Bacsfalvi & Ashdown 2003; Bacsfalvi, Adler-Bock, Bernhardt & Gick 2004). Because these L2 participants were linguistics students, they already had some knowledge of phonetics that they could apply to the training session, making the extremely short training period feasible. Further, all of the participants had had years of English training, including pronunciation training. However, none of the participants had previously examined images of their productions of /l/ and /r/. Syllable- and word-lists were created on the spot for practice in the session and post-training.

The particular components identified for English productions of /l/ and /r/ on ultrasound were as follows: The /l/ has two major lingual constrictions – a tip constriction at the alveolar ridge, and a dorsum retraction toward the uvula or into the upper pharynx. The 'stretching' of the tongue allows for the lateral release that is characteristic of the /l/. Pre-vocalic /l/ before non-back vowels shows simultaneous production of the two constrictions; post-vocalic /l/ and /l/ before back vowels shows sequential timing of the constrictions, with the post-vocalic constriction preceding the pre-vocalic constriction. For the /r/, there are two major variant shapes: bunched and retroflex. Both have two primary lingual constrictions: an anterior constriction in the palatal region, and a root retraction into the pharyngeal cavity. For the retroflexed /r/, the anterior constriction shows a curling back and raising of the tongue tip, with the body gently sloping downwards towards the pharynx. In the bunched /r/, the anterior constriction shows the tip down and the blade/body raised toward the palate, with a fairly steep downwards slope towards the tongue root. In both cases, the sides of the tongue body contact the back teeth and palate, bracing the anterior sections of the tongue.

At the end of the 30-minute session, all three participants were able to produce their target approximant successfully in the problem contexts. In the pre-training assessment, participants varied in which and how many articulatory components of /l/ or /r/ were missing or incorrectly produced. Post-assessment showed generalization of the changes made to the word-list for the assessment, although least for the third speaker, who had the most changes to make. The success of ultrasound in facilitating change for these participants who had persistent difficulties with specific L2 pronunciation targets exemplifies how visual feedback technologies can have exciting potential for L2 training in speech production, and helps to illustrate the goals and methods of an intervention study using ultrasound imaging. More in-depth training studies are ongoing using speakers from a variety of backgrounds and ages.

Discussion

While the potential benefits of articulatory feedback for speech training have long been acknowledged, it is only recently that the technology has reached a point where implementation in typical L2 research and pedagogy has become feasible. Previous findings using ultrasound imaging in pronunciation training, as well as the pilot experiment outlined in the present paper, show strong promise for ultrasound imaging in the future of these areas.

Other areas of L2 research where ultrasound imaging has clear implications – such as in the description of poorly described speech sounds and a deeper understanding of articulatory settings - have been described elsewhere (see above), and may be considered equally promising for the future of L2 pronunciation research.

Limitations and future directions

Because ultrasound applications in speech research are still relatively new, there remain a number of core issues in ultrasound research that have not been thoroughly worked through, mainly concerning quantification. First, because ultrasound provides a large amount of information (full spatial 2-dimensional images of the tongue at standard video rate), there has been little standardization, with different researchers using different methods of measuring the tongue, some focusing on “depth” or distance from the transducer, others reconstructing absolute spatial positions, and still others approximating and quantifying the shape of the tongue surface. Second, any quantification technique except those that depend only on shape (e.g., Stone, Morish, Sonies, & Shawker 1987; Iskarous 2004) requires location of the tongue surface in space. Because ultrasound does not image bone or other fixed anatomical structures, head and transducer stabilization or tracking is a vital part of determining best practices for ultrasound tongue measurements. Third, although temporal resolution is high compared to some other available imaging techniques (e.g., MRI), 30 frames per second is still too slow to capture some types of movement adequately. All of these issues are the subjects of ongoing investigation. However, it is important to note that none of these are limiting factors in using ultrasound for its most powerful L2 application: the imaging of tongue positions for visual feedback in learning to produce novel speech sounds.

Conclusions

While applications for ultrasound are still new in speech research, this is even more the case in L2 research. Even so, the potential value of this tool for pronunciation teaching is already being realized, and the implications of such a powerful tool for the advancement of knowledge and theory in L2 acquisition are extensive. One of the more important fundamental contributions of ultrasound in pedagogy to date has been in allowing teachers and learners to break down complex articulatory tasks into their practical components. However, whether in visualization, description, or experimental investigation, ultrasound provides an easy-to-use, non-invasive technique available to L2 researchers.

Suggested readings

Computer-aided pronunciation (CAP) in pedagogy:

Anderson-Hsieh, J. 1996. A guide to computer assisted aids for pronunciation improvement. *Speak Out!* 19: 28–31.

Hincks, R. 2003. Speech technologies for pronunciation feedback and evaluation. *ReCALL* 15(1): 3–20.

Pennington, M. C. 1999. Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning* 12(5): 427–440.

Ultrasound applications in speech:

Bernhardt, B., Gick, B., Bacsfalvi, P., & Adler-Bock, M. 2005. Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics and Phonetics*. 19(6/7): 605–617.

Stone, M. 1997. Laboratory techniques for investigating speech articulation. In W. J. Hardcastle & J. Laver (eds.), *Handbook of Phonetic Sciences*, Oxford: Blackwell, 11–32.

References

Acton, W. 1984. Changing fossilized pronunciation. *TESOL Quarterly* 18(1): 69–83.

Adler-Bock, M., Bernhardt, B., Gick, B., & Bacsfalvi, P. 2007. The use of ultrasound in remediation of English /r/ in two adolescents. *American Journal of Speech-Language Pathology* 16(2): 128–139.

Anderson-Hsieh, J. 1996. Teaching suprasegmentals to Japanese ESL learners through electronic visual feedback. *Japan Association for Language Teaching (JALT) Journal* 18(2): 315–325.

Bacsfalvi, P., Adler-Bock, M., Bernhardt, B., & Gick, B. 2004. Technology breakthroughs for /r/ and /l/ treatment. Poster presented at the CASLPA Annual Conference. Ottawa. May 2004.

Bernhardt, B., Gick, B., Bacsfalvi, P., & Ashdown, J. 2003. Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics & Phonetics*. 17(3): 199–217.

Bernhardt, B., Bacsfalvi, P., Gick, B., Radanov, B. & Williams, R. 2005. Exploring electropalatography and ultrasound in speech habilitation. *Journal of Speech Language Pathology and Audiology* 29(4): 169–182.

Bradlow, A. R. 2008. Training non-native language sound patterns: Lessons from training Japanese adults on the English /ɹ/–/l/ contrast. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Catford, J. C. 1987. Phonetics and the teaching of pronunciation: A systemic description of English phonology. In J. Morley, (ed.), *Current perspectives on pronunciation: Practices anchored in theory*, Washington, D.C.: TESOL, 87–100.

Catford, J. C., & Pisoni, D. B. 1970. Auditory vs. articulatory training in exotic sounds. *Modern Language Journal* 54(7): 477–481.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. 1996. *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge: Cambridge University Press.

Chun, D. M., Hardison, D. M., & Pennington, M. C. 2008. Technologies for prosody in context: Past and future of L2 research and practice. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Clawson, E. 1907. Value of phonetics from the standpoint of the student. *The School Review* 15(1): 48–54.

Collins, B., & Mees, I. M. 1995. Approaches to articulatory setting in foreign-language teaching. In J. W. Lewis, (ed.), *Studies in general and English phonetics: Essays in honour of Professor J. D. O'Connor*, London: Routledge, 415–424.

Dale, P. W., & Poms, L. 1994. *English pronunciation for international students*. Englewood Cliffs, NJ: Prentice Hall.

Delattre, P. C., & Freeman, D. C. 1968. A dialect study of American *r*'s by x-ray motion picture. *Linguistics* 44: 29–68.

Derwing, T. M. 2008. Curriculum issues in teaching pronunciation to second language learners. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and Second Language Acquisition*, Philadelphia: John Benjamins.

Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., & Perkell, J. S. 1999. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America* 105(5): 2854–2865.

Gick, B. 2002. The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association.* 32(2): 113–122.

Gick, B., Bird, S., & Wilson, I. 2005. Techniques for field application of lingual ultrasound imaging. *Clinical Linguistics and Phonetics*. 19(6/7): 503–514.

Gick, B., Wilson, I., Koch, K., & Cook, C. 2004. Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica* 61(4): 220–233.

Honikman, B. 1964. Articulatory settings. In Abercrombie, Fry, MacCarthy, Scott and Trim (eds.), *In honour of Daniel Jones*, London: Longman, 73–84.

Huck, S. 2000. *Reading statistics and research*. New York: Longman.

Iskarous, K. 2004. Edge detection and shape measurement of the edge of the tongue. Ms., Haskins Laboratories, New Haven, CT.

Kelly, G. 2000. *How to teach pronunciation*. Harlow, U.K.: Longman.

Kelsey, C. A., Woodhouse, R. J., & Minifie, F. D. 1969. Ultrasonic observations of coarticulation in the pharynx. *Journal of the Acoustical Society of America* 46(4): 1016–1018.

Lambacher, S. 1999. A CALL tool for improving second language acquisition of English consonants by Japanese learners. *Computer Assisted Language Learning* 12(2): 137–156.

Locke, W. N. 1954. III. Speech analysis and synthesis: Development and prospects. Part (pp. 416–423) of Girard, D.P., Borglum, G., Locke, W.N., & Dostert, L.E., Teaching aids: Problems and possibilities. *The French Review* 27(6): 406–427. Originally presented at the 68th Annual Meeting of the MLA, Chicago, Dec. 29, 1953.

Mompeán González, J. A. 2003. Pedagogical tools for teaching articulatory setting. In M. J. Solé, D. Recansens, & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences Barcelona 2003* (ICPhS), Adelaide, South Australia: Causal, 1603–1606.

Morgan, D. L., & Morgan, R. K. 2001. Single-participant research design: Bringing science to managed care. *American Psychologist* 56: 119–127.

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

Munro, M. J. 2008. Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Odisho, E. Y. 2003. *Techniques of teaching pronunciation in ESL, bilingual & foreign language classes*. Munich: LINCOM Europa.

Richards, S. B., Taylor, R. L., Ramasamy, R., & Richards, R. 1999. *Single subject research: Applications in educational and clinical settings*. San Diego, CA: Singular Publishing Group.

Skolnick, M. L., Zagzebski, J. A., & Watkin, K. L. 1975. Two dimensional ultrasonic demonstration of lateral pharyngeal wall movement in real time – a preliminary report. *Cleft Palate Journal* 12: 299–303.

Stone, M., Morish, K., Sonies, B. C., & Shawker, T. H. 1987. Tongue curvature: A model of shape during vowel production. *Folia Phoniatrica* 39: 302–315.

Wilson, I. L. 2006. *Articulatory settings of French and English monolingual and bilingual speakers*. PhD dissertation, University of British Columbia.

Yule, G., & Macdonald, D. 1994. The effects of pronunciation teaching. In J. Morley, (ed.), *Pronunciation pedagogy and theory: New views, new directions*, Alexandria, VA: TESOL, 111–118.

Zagzebski, J. A. 1975. Ultrasonic measurement of lateral pharyngeal wall motion at two levels in the vocal tract. *Journal of Speech and Hearing Research* 18: 308–318.

CHAPTER 12

Technologies for prosody in context

Past and future of L2 research and practice

Dorothy M. Chun, Debra M. Hardison
and Martha C. Pennington
University of California at Santa Barbara / Michigan State University /
Georgia Southern University

Introduction

Phonology, like all of linguistics, is going through a period in which issues of autonomy and of the need to incorporate contextual features into description and explanation are increasingly being raised. One can observe in research a "rapidly shifting landscape and high level of activity [characterizing] current work in phonology in the attempt to account for the complex patterns of occurring forms within the context of language learning and language change" (Pennington & Clark 2002: 448). Pennington (2002) attributes the increased attention to the social context of phonology, even among formal linguists, in part to "increased quantities of research – and the resultant advances and discoveries – in both natural science and social science, which have filtered into linguistics" (p. 441), leading "phonology to move in the same general direction as syntax in the current era, i.e. towards more empirically and functionally grounded accounts of language" (p. 442).

This chapter examines the recent history of prosody in the study of second language (L2) phonology and traces the emergence of a contextual perspective in relation to the prosodic, or suprasegmental, level of language. The discussion focuses on the research and practice enabled by technological advances – particularly, in computer environments – and suggests a future direction that increasingly integrates the prosodic dimension within a larger context of communication and human behavior more generally. Jenkins (2004) recently stated "Of the recent findings of pronunciation research, the most influential in terms of pedagogic developments fall into two main groupings: those concerned with issues of context and those that relate to technological advances" (p. 109–110).

The discussion looks both backwards and forwards in the area of L2 prosody in context. Looking backwards, in the history of L2 phonology, prosody has been viewed as a level of language organizing and contextualizing the micro-level units of phonetic segments or phonemes (segmental phonology). In the 1980s, applied linguists recognized a need for more of an emphasis on the prosodic level in research and teaching to correct for a previous generation of studies focused on segmental phonology. Since that time, prosody has come to be recognized as an important part of the analysis of language in context, in traditions such as sociolinguistics, conversation analysis, pragmatics, and systemic-functional linguistics. Some encouraging signs of this same emphasis can be found in the L2 literature. Looking forwards to the potential for research and practice in L2 prosody, we propose that links will be made increasingly not only 'downwards' to segmentals but 'upwards' to the macro-level context of communication and other behavior.

History and potential of L2 discourse prosody

Over the last three decades, the study of intonation and prosody has gained increased attention in linguistics proper – in not only syntax but also pragmatics, discourse analysis, and conversation analysis – and in applied linguistics as well. Up to the 1980s, two main schools of thought on the teaching of prosody dominated. One of these, the American structuralist approach, associated pitch contours made up of sequences of pitch phonemes with sentence types such as declarative or interrogative and with emotions such as "impatience" or "surprise" (Leather 1983: 200). The British approach, rather than individual pitch phonemes, posited a finite inventory of "tunes" for a given language (e.g., O'Connor & Arnold 1961). Generative phonology, as a flourishing tradition in theoretical linguistics, generally did not offer direct applications to pronunciation teaching. It has, however, contributed much to our understanding of the theory and phonetic detail of intonation (cf. Pierrehumbert 1980). In L2 phonology proper, Major (2001) proposed that overall accent ("global accent") be broken down into three components: segmental, syllabic and suprasegmental. This inclusion of suprasegmentals has implications for the teaching of L2 phonology.

In the 1980s, the role of intonation in discourse and conversation emerged as an important research topic in L2 pedagogy. The field of ESL/EFL took the lead in reviving interest in pronunciation and in promoting the teaching of stress, rhythm, and intonation.[1] Pennington and Richards (1986) and Pennington (1989b) called for a "top-down" approach to the teaching of pronunciation, focusing on the

1. The field of ESL/EFL continues to promote the teaching of prosody, as evidenced by regular presentations in the Speech and Pronunciation Special Interest Sessions at the annual

rhythm, intonation, and trans-segmental properties of phrases and larger contexts as opposed to the "bottom-up" analysis or construction of larger units on a basis of individual sounds or words. In this and later work, the prosodic dimension is seen as a fundamental component of both listening comprehension and oral proficiency, and linked to intelligibility, as Derwing notes in Chapter 13 of this volume.

In the early 1990s, Morley (1991) included in her list of changing principles and priorities for pronunciation teaching the emerging focus on the link between perception and production on listening and speaking (p. 494). Leather and James (1991) reported a resurgence of interest in the acquisition of L2 speech over the previous decade and called for more attention to higher-level discourse patterning in research on the acquisition of L2 speech (see also Wennerstrom 1994), in parallel with work in linguistics proper (e.g. Pierrehumbert & Hirschberg 1990). As a result of the renewed interest in the acquisition of L2 phonology and the communicative functions of language, instructional materials, including computer software, are showing an increased emphasis on pronunciation in general and intonation in particular.

In the late 1990s, a dual-focus program of pronunciation teaching that incorporates both a *micro-level* focus on *speech production* (i.e., on discrete elements of articulation) and a *macro-level* focus on *speech performance* (i.e., on general features of communication) was advocated by many (cf. Clennel 1997; Pennington 1996a, 1998). At the micro-level, one might begin to teach prosody in terms of such features as syllable structure and the realizations of strong and weak stress, along with fluency phenomena such as elisions, assimilations, reductions, and contractions (cf. Brazil 1994; Celce-Murcia, Brinton & Goodwin 1996; Hahn & Dickerson 1999; Henrichsen, Green, Nishitani, & Bagley 1999). The crucial next step is to teach prosody at the macro-discourse level. Recent studies of discourse intonation suggest the need to base instruction on authentic conversations and to describe utterances and their accompanying prosodic features in naturally occurring contexts (cf. Chafe 2002; Couper-Kuhlen & Selting 1996; Wichmann 2000). L2 learners need to be able to understand intonational contrasts made by native speakers in authentic discourse while also making themselves understood in terms of their own intonation.

This orientation to teaching *discourse intonation*, i.e., with a focus on how intonation serves pragmatic or social functions in discourse, is not entirely new. In the late 1970s and early 1980s, this approach was developed at the University of Birmingham by David Brazil, along with John Sinclair and Malcolm Coulthard.

TESOL convention and in the special-topic issue "Re-conceptualizing Pronunciation in TESOL," Autumn 2005.

It became influential in English language teaching starting in the mid 1980s, both for teacher training and classroom pronunciation practice, but primarily in the U.K. The influence continues to grow, e.g., as noted by Cauldwell (2002)[2] and Jenkins (2004).

Research on teaching prosody in larger, authentic contexts is now flourishing on both sides of the Atlantic. In Europe, for example, Seidlhofer and Dalton-Puffer (1995) reviewed recent insights into the pragmatics of language use and language learning that point to the importance of larger prefabricated units. Cauldwell and Hewings (1996) examined two rules of intonation that they believed perhaps to be the most commonly found in ELT textbooks, namely, those concerning intonation in lists and intonation in questions. They found these rules to be inadequate as descriptions of what occurs in natural speech and offered alternative analyses of the patterns of intonation using Brazil's discourse intonation model. The goal of this and other current work in prosody is to provide learners with descriptions of intonation which will allow them to understand the communicative significance of the patterns of intonation identified in such rules, in other words, looking beyond syntactic types to the broader context of the surrounding communication or discourse.

Jenkins (2002) advocated the need for empirically established phonological norms and classroom pronunciation models for English as an International Language and also stressed intelligibility as a key goal. In addition to specific segmental items, including features of connected speech, especially assimilation, she listed the following suprasegmental items as important: appropriate use of contrastive stress, direction of pitch movements to signal attitude or grammatical meaning, the placement of word stress, and stress-timed rhythm (pp. 97–98). However, as Derwing notes in Chapter 13, more research needs to be done as this model is based on data from communication breakdowns between only a few learners.

In the U.S. and Canada, discourse intonation has been investigated by an increasing number of researchers. Park (2003), drawing on a framework of communicative competence proposed by Celce-Murcia et al. (1996), provided concrete examples of the different pronunciation features, primarily prosodic, that should be taught and learned if one is to demonstrate linguistic, discourse, actional, sociocultural and strategic competence (p. 5). Wennerstrom (2000) reported on a conversation analysis of naturally occurring dialogues using computerized speech equipment to analyze ESL speakers' pitch patterns. She showed that intonation is one of the important variables contributing to fluent speech in English.

2. Cauldwell is also the author of the website "Centre for Discourse Intonation Studies" (CDIS), found at the following URL: http://www.speechinaction.pwp.blueyonder.co.uk/CDIS_Home.htm.

Wennerstrom (2001) demonstrated the centrality of prosody in the interpretation of spoken texts. The role of prosody was considered in such discourse genres as casual conversation, oral narratives, courtroom testimony, lectures, and second language discourse. The studies established a framework for transcribing and analyzing prosody in discourse and provided a wealth of data illustrating a wide range of intonational phenomena in the analysis of conversations.

Since intonation can function in discourse in a variety of ways, the goal according to Chun (2002) is to describe and represent prosody as it serves the following functions (pp. 77, 206):

- to mark new information or the shared mutual knowledge of a speaker and listener
- to mark such boundaries as between sentences, paragraphs, topics, conversational turns
- to control interactive structure or organize conversational exchange
- to continue an established topic or to signal a new topic

In step with contextualizing the teaching of prosody, Levis (2001) proposed a functional approach to teaching ESL learners to predict focus in a sentence. The approach appeals to functional and meaning regularities in answering questions, correcting misinformation, and controlling repeated questions, and is easily applicable to normal conversation. The ultimate goal is to bridge the gap between classroom instruction and unplanned, authentic, contextualized communication.

Derwing, Munro and Wiebe (1998) provided evidence that learners who had received instruction on features such as speaking rate, intonation, rhythm, projection, word stress, and sentence stress showed significant improvement in comprehensibility and accentedness when they produced extemporaneous narratives. Learners who had received only segmental instruction improved their comprehensibility and accentedness when reading simple sentences but not when producing extemporaneous speech. Similarly, Derwing and Rossiter (2003) demonstrated that following 12 weeks of global pronunciation instruction (primarily prosodic features), non-native speakers' pronunciation improved significantly in terms of comprehensibility and fluency. They did "not advocate eliminating segment-based instruction altogether, but, if the goal of pronunciation teaching is to help students become more understandable, then this study suggests that it should include a stronger emphasis on prosody" (p. 14). In addition, they recognized the social nature of interaction and that reactions to accented speech are affected by many factors other than comprehensibility. They stated that the ultimate goal for teachers and researchers must be to reconcile the many factors that influence successful communication, which belong in a much larger discourse context (for an extended discussion of this issue and L2 pronunciation pedagogy in general, see Chapter 13 by Derwing in this volume).

Pickering (2001, 2002) examined data from naturally occurring university classroom interactions, focusing specifically on the contribution of prosodic cues to exchanges between Teaching Assistants (TAs) and students. Analysis of the prosodic structures used by native speakers (NS) and non-native speakers (NNS) revealed that NNS TAs were unfamiliar with the prosodic norms of typical classroom exchanges and their "miscues" at the level of pitch and pause structure led to cross-cultural communication failures. Her focus on prosody in the broader context of classroom interaction represents an important direction for future research in L2 phonology.

Given that both researchers and practitioners have recognized the importance of prosody in L2 phonology, the following section presents recent work on discourse intonation with an emphasis on both research and teaching as facilitated by technology, i.e., how computers and computer software have been utilized to implement some of the goals of prosody instruction described above.

Integrating technology into research and instruction: Present and future

Methodological options

As early as the 1960s, researchers were using visualizations of pitch contours to raise learners' awareness and enhance their understanding of intonation. Technological advances have made possible a direct comparison of learners' utterances with those of native speakers. Software tools have become increasingly accessible and affordable, enabling systematic research in the teaching of prosody using these tools. With these technologies, it is possible for instance to digitize authentic samples of speech and to incorporate them into a multimedia environment for purposes of teaching and research. This capability is not restricted to the audio form of these authentic contexts but includes, through video recording, the visual components of behavior, such as gestures, as these relate to the meaning conveyed by prosody. The potential for using video for analysis and teaching of L2 prosody in context within a multimedia environment is discussed, and recent original work by the authors and others is presented at the end of this section.

There are four main technological tools that have been used in L2 prosody instruction, each with its own benefits and limitations. First, **visualizations of pitch contours** are useful for sentence-level or discourse-level chunks of language, but there are screen limitations on how much is visible at one time. Display quality is subject to sustained phonation (e.g., vowels and nasals), but despite breaks in the contour due to voiceless consonants, most pitch displays are interpretable by non-specialists, including language learners. This type of technology has the advantage of not being language-specific. Second, **multimodal tools** (see description of *Anvil* below) allow integration of a video clip with the associated pitch contour, allowing optional annotations as well. This process is comparatively time-consuming and may be more suitable for smaller chunks of data. Tools such as *Anvil* are also applicable to any language and address the need to coordinate multiple components of a speech event for analysis and feedback. Third, **spectrographic displays** for segmental information are easily created by most current software programs; however, they are not as easily interpretable as pitch contours by non-specialists, particularly L2 learners. Fourth, **vowel analysis programs** such as *Sona-Match* by KayPENTAX (formerly Kay Elemetrics) can identify vowels produced by the user in relation to a "vowel space" chart, but they can only process isolated vowels in continuous (steady-state) production. This focus on individual phonemes is in contrast to the current move toward analysis of larger, contextualized speech.

In terms of methodological approaches to the use of technology for L2 prosody teaching, two main trends can be cited. The first is the use of isolated scripted sentences in developing prosodic awareness. The main advantage of this approach is that it provides the teacher or researcher with content control and allows for the focusing of learner attention. However, it does not transfer as well to discourse-level speech as training with discourse-sized units of natural speech samples (see Hardison 2004, 2005, discussed below). The second approach is the use of imitation as a training technique, also typical of the "accent-reduction" work carried out by some speech pathologists. An imitation-only approach does not appear to promote generalization to novel utterances and transfer to discourse-level speech in spontaneous conversations in the natural language environment. Rather, we would suggest that there appears to be a benefit to matching the training with the ultimate task.

The research base

We first briefly review selected past research on developing and applying instructional software for teaching prosody. A detailed account of the available hardware and software packages and of how computers have been utilized since the 1960s to provide learners with visualizations of intonation patterns can be found in Chun (2002: Chapter 5, 2007). Building on advances in the theory and measurement of intonation and aided by the growing accessibility of computational acoustic speech analysis, applications to teaching second and foreign languages have been developed and trialed, most notably in the last two decades.[3] In her dissertation,

3. Carey (2004) describes six products that provide learners with visual feedback: (1) Technologically Enhanced Accent Modification (TEAM), 1999 Version 2.0 (Erlbaum) $495, retrieved from the World Wide Web, April 23, 2004 http://www.ed.gov/about/offices/list/ope/fipse/lessons4/cleveland.html; (2) Accent Lab ($39.95); (3) Protrain; (4) Dr. Speech, which comes

Verhofstadt (2002) provided a critical analysis of computer-assisted pronunciation materials that included the teaching of prosody with electronically provided visual feedback. Her survey demonstrated that most research evaluating the contribution of both visual and auditory input in the acquisition of L2 prosody focused on the sentence level.

Several studies have shown that a visual display of a pitch contour is more effective than auditory-only input for improving learners' production of prosody in English (e.g., de Bot 1983; Weltens & de Bot 1984) (for research on the use of visual displays, in the form of ultrasounds, to improve segmental accuracy see Chapter 11 by Gick et al. in this volume). The display of prosodic features is reasonably easy to interpret by non-specialists – an important criterion in the selection of a computer-based pedagogical tool (Chun 1998). Moreover, the use of such a display provides valuable information about intonation as well as about tone. Leather (1990), for example, investigated the influence of (i) perception training on production and (ii) production training on perception of Chinese tone, using computer-based training that included comparison of the visual display of a learner's production with that of a native speaker.

Gorsuch (2001) reported on pronunciation training of EFL students and found that after production-focused instruction, students' perception of suprasegmentals seemed to improve, while their production did not, underscoring the notion that developments in second language speech perception and production do not necessarily parallel each other.[4] Hew and Ohki (2004) examined the effectiveness of imagery and electronic visual feedback in facilitating students' acquisition of the pronunciation of specific Japanese word pairs. Students who received text + audio + animated graphic annotations (AGA) significantly outperformed students in the text + audio only group in terms of pronunciation of pairs of words which have the same written form but differ in their pitch patterns. Students in the text + audio + immediate visual feedback (IVF) also significantly outperformed students in the text + audio only group. But there were no significant differences between the AGA and IVF groups.

Recent studies on the perception and acquisition of English intonation by native Cantonese speakers (Pennington & Ellis 2000; Pennington, Ellis, Lee, Lau, & Lock 2002) provide a basis for making pedagogical recommendations to be implemented on computer. The first of these (Pennington & Ellis 2000) is a study of

---

in two product versions, Real Analysis ($795) and Speech Training ($695); (5) Video Voice; and (6) Kay Sona-Match. Another product from Auralog, *TeLL me More*, is available for several languages http://www.auralog.com/.

4. However, she also cautions that the apparent gains or lack of gains may be dependent on the tests and testing procedures themselves.

A. Pragmatic Interpretation

1. *Focus*: Intonation cues unmarked (neutral) focus by means of prominence on the last constituent vs. marked (contrastive/emphatic) focus by means of an extra degree of prominence on an item or unit of the sentence.

Is HE driving the bus?	[special attention on *HE* for emphasis or contrast]
Is he driving the bus?	[no special emphasis or contrast]

2. *Tag*: Intonation cues a different pragmatic interpretation by means of a "closed" (falling) vs. "open" (rising) contour on an utterance-final tag. This contrast signals sentence modality as relatively certain (statement) and thereby soliciting agreement, or uncertain (question) and thereby soliciting an opinion.

He's a good boy, isn't he. (falling)	[I think he's a good boy and that you will confirm this.]
He's a good boy, isn't he? (rising)	[I think he's a good boy, but I am not sure.]

B. Syntactic Structure

3. *Boundary*: Intonation cues boundary, i.e., continuity/discontinuity of unit in (final) boundary position.

The fight is over Fred.	[The fight is about Fred.]
The fight is over, Fred.	[I am telling you Fred that the fight is finished.]

4. *Phrase Structure*: Intonation cues internal syntactic analysis, i.e., phrase structure, of unit.

She's a lighthouse keeper.	[Her job is to look after a lighthouse.]
She's a light housekeeper.	[She is a housekeeper who does light housework.]

Figure 1. Sentence Types for Intonation Processing Study (adapted from Pennington & Ellis 2000)

sentence recall with intonation as the key variable differentiating semantic interpretation. Stimuli were developed in contrasting pairs of four types as shown in Figure 1.

In two of the item types (A.1 and A.2), the intonational contrast signaled a difference in pragmatic interpretation, while in two other item types (B.1 and B.2), the intonational contrast signaled a difference in syntactic structure. Participants listened to two different sets of sentences and had to judge whether each sentence in the second set was the same as an original sentence in the first set or not. Sentences in the second set were either:

a. Identical: previously heard sentences, i.e., with same lexis and same intonation;
b. Intonationally contrastive: sentences with same lexis but contrasting intonation;
c. Lexically contrastive: sentences with different lexis and one of the two contrasting types of intonation.

In the first phase of the research, the recognition task was performed without any prior instruction. In the second phase, participants had their awareness of the contrasts raised in a session in which they heard sentences of each type represented in the pairs illustrated in Figure 1 and had to select from the contrasting interpretations given for each pair type as cued by intonation. Following the period of training the perception of the sentence contrasts, the recognition task was performed as before.

The phase 1 results showed that these L2 English speakers easily recognized sentences which were lexically contrastive but had poor recognition performance on sentences which were the same lexically but contrasted intonationally from those originally heard. They were somewhat better at recognizing completely identical (i.e., in both lexis and intonation) sentences. This pattern of response suggests that the Cantonese L1 speakers' English sentence processing was more focused on lexis than intonation. The only significant improvement from phase 1 to phase 2, following the period of awareness-raising focused on the intonational contrasts, was in correct recognition of sentences of type A.1. The findings reinforced the view as expressed elsewhere (e.g., Pennington 1996a, 1998) that adult learners need explicit focusing of attention on both the form and meaning of intonation to make improvements in phonology as in other areas of language.

The second study (Pennington, Ellis, Lee, Lau, & Lock 2002) investigated the learning of intonation on computer, comparing seven different pedagogical orientations. Four types of intonation were trained in the study – *wh-* question, echo question, *either-or* question, and statement – as in the two sets of sentences shown in Figure 2.

Participants were first exposed to a set of example sentences of each of the four types in a spoken version heard through earphones and in a written version seen on the screen. They were then given a pretest in which they had to repeat 6 different sets of the 4 sentence types presented in random order in spoken and written

	Set A	Set B
wh-question [falling]	*Why do you like him?*	*Where are you going?*
echo question [rising]	*You like him because he's intelligent?*	*You're going home?*
either-or question [rising then falling]	*You like him because he's intelligent,* *or because he's rich?*	*You're going home,* *or you're staying here?*
statement [falling]	*You like him because he's intelligent.*	*You're going home.*

Figure 2. Sentence Types for Intonation Training Study (Pennington, Ellis, Lee, Lau, & Lock 2002)

versions as before. The pretest was followed by a period of instruction using 6 new sets of the 4 sentence types presented in random order in spoken and written versions. Following the training period, each individual was given a posttest that was identical to the pretest. The tests and instructional treatments were delivered by means of the public domain program, *Psyscope*, originally developed as a means for delivery of psychology experiments on computer but adapted for use in this study.

As shown in Figure 3, three treatments were perception-focused, while three implemented a combined perception-production focus. A final treatment used the Kay Elemetrics Visi-Pitch to deliver self-controlled real-time visual feedback of intonation. This was a special treatment that required a small period of familiarization with an assistant following the pretest. Pretests and posttest recordings were analyzed by Visi-Pitch, and effectiveness of treatment was assessed by significant changes pretest to posttest for each treatment and sentence type, using various measurements of the contours such as pitch range and ratios of pitch minima or maxima to pitch range or average pitch (for details, see Pennington, Ellis, Lee, Lau, & Lock 2002).

The most significant effects were found for *either-or* questions, followed by echo questions, with *wh*-questions and statements showing few significant changes from pretest to posttest. The treatment showing the greatest effect of training was that involving a simple focusing of attention on intonation by listening and not repeating each sentence as it was heard and seen (treatment 1). That this treatment was effective in improving participants' intonation in this brief period supported the findings of other studies (e.g., De Bot & Mailfert 1982; Leather 1990) that productive practice is not necessary for improving production of tone or intonation in L2 and that improvement in L2 phonology can occur in an instructional context that focuses attention on form alone. The two other treatments which showed some significant pre/post improvements in intonation were treatment 5 combining repetition and a visual representation of the intonation contour and treatment 6 combining repetition and an extracted auditory pitch contour. The effectiveness of these two treatments suggested the value of raising awareness of pitch cues in the context of whole utterances outside of any attention to meaning.

Both studies underscored the need to explicitly focus language learners' attention on intonation and give directions for providing such a focus in computer-based instruction. The results of the second study suggested the effectiveness for improved production of intonation in L2 of concentrated attending in a highly focused instructional context such as can be delivered in individual computer workstations. Those results further suggested ways that such instruction may be delivered without the additional resource of on-line acoustic analysis and real-time visual feedback.

Input Training
1 – Focusing of attention: Participants were instructed only to listen and pay attention but not to repeat the intonation of the sentences as they were spoken once each and appeared at the same time in written form.

2 – Perception training: Participants were to listen to the sentences and mark intonation as falling or rising by ticking a rising or falling arrow next to the written version of each sentence as it was spoken.

3 – Written description + Perception training: The same instructional format as 2 preceded by a presentation on the screen of the following verbal descriptions of each sentence type, accompanied by a spoken version with the indicated intonation:

Wh-Question
A wh-question is one beginning with one of the wh-words *when*, *where*, *why*, *who*, *which*, *what* or *how*. For any wh-question, the speaker's voice falls at the end of the question. EX 1: *Why do you like him?*

Echo Question
Any sentence can be made into a question by having the voice rise at the end of it. Speakers sometimes repeat information they have heard, but with a rising voice, in order to ask whether or not it is correct. This is sometimes called an echo question. EX 2: *You like him because he's intelligent?*

Either/Or Question
In an either/or question, two alternatives are presented, of which the listener is expected to choose one. For this type of question, the speaker's voice rises at the end of the first part (the part before *or*) and then falls at the end of the second part (the part after *or*). EX 3: *You like him because he's intelligent, or because he's rich?*

Statement
In a statement, the speaker's voice falls at the end. EX 4: *You like him because he's intelligent.*

Input-Output Training
4 – Repetition practice: Participants were to repeat each sentence three times after they heard it once and saw its written form on the screen.

5 – Repetition practice + visual reinforcement: This treatment added visual displays of simplified intonation contours of the sentences to the repetition practice of 4. These were derived from Visi-Pitch tracings of the sentences by drawing over them with a felt-tip marker to eliminate small perturbations of pitch, and these "smoothed out" contours were input to the computer by means of a scanner. They appeared on the screen above the written version of each sentence when it was spoken.

6 – Repetition practice + aural reinforcement: This treatment added an extracted pitch contour for each sentence to the repetition practice of 4. Contours were extracted from the spoken words of each sentence by means of the *Signalize* speech analysis system and then input as audio files to *Psyscope*. The extracted contours (instead of the full utterances) were played as the written form of the each sentence appeared on the screen.

7 – Repetition practice +Visi-Pitch split-screen on-line visual feedback: This treatment allowed the user to self-monitor his/her intonation in comparison to a real-time visual model of the intonation contours provided by speech analysis equipment (the Kay Elemetrics Visi-Pitch), with the possibility of multiple repetitions. This treatment required additional time following the pretest for instructions from the assistant and practice with one of the example sentences on how to use the additional Visi-Pitch equipment, which was set up in the soundproof room next to the PC used for the pretest and posttest. With Visi-Pitch in split-screen mode, as each sentence was played, its intonation contour appeared at the top of the screen. Within the remaining time of the 10 sec. interval between sentences, the learner could try to match the intonation contour by repeating the sentence one or more times, with Visi-Pitch analysis provided each time to display his/her own intonation at the bottom of the screen for comparison with the target contour at the top of the screen.

Figure 3. Types of Instruction in Intonation Training Study (Pennington, Ellis, Lee, Lau, & Lock 2002)

In another recent study, Hardison (2004) found that visual feedback was effective in the acquisition of French prosody by foreign language learners (L1 English), and demonstrated generalization to novel sentences and transfer to improved segmental production accuracy – hallmarks of successful training. In addition, questionnaire responses indicated a very positive evaluation of computer-assisted training. Respondents also noted increased confidence in their oral production of French and heightened awareness of the elements that make up speech.

The preceding sections describe major research and pedagogical findings that bring us a step closer to understanding the acquisition of an L2 sound system including its prosody. One of the most important results discussed is that prosody-focused training results in generalization to novel utterances and to other areas of language production, specifically, to segmental accuracy. In addition, practice with computer-based tools facilitates accurate speech by learners, which, in turn, should contribute to the development of automaticity in oral production. However, it should be noted that while the tools provide a type of feedback, this is not human interaction, which has been shown to promote interlanguage development. From a cognitive perspective, prosody and lexical information appear to be linked in long-term memory, and perhaps stored together.[5] Finally, qualitative data collected through questionnaires have demonstrated affective benefits of prosody training. Learners in the Hardison (2004) study responded that they felt they had learned about different aspects of the language and wished they had the same practice opportunities during their regular French course.

Expanding the context of research, pedagogy, and technology for prosody

Past use of the computer to help learners access intonation via visualizations of pitch changes occurring in speech was restricted, initially by the relative inaccessibility of hardware and software, technical limitations on the representations of intonation, and a lack of pedagogical input related to those visualizations. A further restriction was the usual focus on sentence-level intonation of contrasting sentence (syntactic) types, such as declarative statements, *yes-no* questions, *wh*-questions, and exclamations. For the future, Chun (1998, 2002) suggested that developers should move beyond these basic contrasts to apply pitch visualizations of learners' utterances to teach the *communicative* and *sociocultural* functions of discourse-level communication. Computers and computer software should be utilized to:

---

5. This would be compatible with an episodic model of learning.

1. Provide learners with visualizations of their intonation patterns and with immediate feedback to help them improve their speech perception and production;
2. Provide learners with models of authentic speech;
3. Facilitate, record, and analyze speech including interactions of two or more speakers;
4. Offer tools for research, e.g., data collection tools to record students' performance and their steps toward self correction.

We now turn to the use of technology for teaching discourse-level, contextualized prosody and will review the most recent research and instructional software to determine the progress that has been made with respect to the above recommendations.

As has been realized for some time (see, e.g., Chun 2007; Pennington 1989a), multimedia technologies offer exciting possibilities for teaching intonation and other aspects of prosody in context, though as Pennington (1989b, 1996b, 1998, 1999, 2000) pointed out, there is a still a need for further experimentation and research to determine the effectiveness of various approaches. With regard to recommendations #3 and #4 above, one of the greatest potential advantages of computer-based analysis and display of intonation is the dual function of the machine as instructional and research tool. Such computer-based tools can function as a teaching environment while simultaneously keeping detailed records of student performance. Current advances in technology support the use of computers in instructional, evaluative, and research functions. This means, as Pennington (1989a) stressed in her early work, that computer software has the potential to provide an autonomous system integrating training, testing, and research in perception and production of prosody. For pronunciation pedagogy, the greatest potential lies in the incorporation of interactive functions and discourse contexts.[6]

In the past, research studies in L2 speech typically used scripted sentences, often decontextualized, for testing and training. While this affords experimental

6. Three commercial software packages have attempted to incorporate principles of discourse/contextualized intonation into their programs: *In Tune with English* (reported in Kaltenboeck 2001); *Connected Speech*, available from Protea Textware, Australia, http://www.proteatextware.com.au and reviewed by Darhower 2002, http://calico.org/CALICO_Review/review/conspeech.htm and Egbert 2004, http://llt.msu.edu/vol8num1/review2/default.html; *Streaming Speech: Listening and Pronunciation for Advanced Learners of English*, produced by Richard Cauldwell and reviewed by Petrie 2003, http://calico.org/CALICO_Review/review/streaming.htm, Setter (2003), and Lian 2004, http://llt.msu.edu/vol8num2/review2/default.html and Chun (2003) in *TESOL Quarterly* 39: 559–562. However, to date there are no published empirical studies to our knowledge on the effectiveness of these programs.

control over materials, it is not clear whether this type of learning can transfer to improvement in the production of natural discourse-level prosody, which is the objective in communicative language teaching. Hardison (2005) recently investigated this issue with L1 speakers of Mandarin Chinese who were advanced L2 speakers of English and graduate students at an American university. Their pre-training difficulties with English prosody are captured in the category described by Chun (2002) as discourse functions of intonation that contribute to cohesion in speech including the marking of thought groups with appropriate pausing and pitch movement, and the use of stress and intonation to mark information focus. These are important elements in the functional use of prosody that contribute to higher production ratings for L2 speakers of Mandarin Chinese (Wennerstrom 1998).

The specific goal of the Hardison (2005) study was to investigate the effects on the production of discourse-level English prosody of different types of contextualized training using segments from the participants' own oral presentations on familiar topics. Two computer-based tools were used to compare two weeks of training with and without the visual context of the speech event, and with discourse-level input versus isolated sentences. The tools were:

(1) web-based *Anvil* (Kipp 2001),[7] which provides a screen display integrating the audio and video components of a speech event with the associated pitch contour created in *Praat*,[8] a public domain phonetic tool; and

(2) *Real-Time Pitch* (RTP) program in conjunction with the *Computerized Speech Lab* (KayPENTAX), which produces a pitch contour in real-time and allows on-screen comparison of a learner's utterance with that of a native speaker, including overlay of one contour on the other.

Two groups received training input using *Anvil* and practiced production focused on prosody with RTP including feedback from a native speaker. Two groups used only RTP to view their pitch contours and to practice, and received the same type of feedback. Within these pairs, one group received discourse-level input and the other individual sentences taken from their recorded presentations. Each group served as its own control in a time-series design incorporating a series of five presentations (three prior to training to establish the individual's normal use of

7. See http://www.dfki.de/~kipp/anvil. There is a link to a demo screen shot. Directions are given for those who wish to obtain the address for downloading the files. It is free for research purposes.

8. Created by Boersma (2001) and Weenink. Available free of cost at http://www.fon.hum.uva.nl/praat.

prosody and two following training to assess improvement and retention). Mean performance across groups was comparable prior to training.

Native speakers provided ratings of global prosody for each participant's series of five oral presentations. Comparison of pre- and post-training data revealed that although all groups improved, discourse-level input produced better transfer to novel natural discourse and the presence of video (i.e., using *Anvil*) was more helpful with discourse-level input than with individual sentences. The discourse-level training materials more closely resembled the type of connected speech on which the assessment of improvement was based (i.e., the participants' oral presentations) – the type of speech one is more often required to produce. This linguistic context is crucial to the assignment of stress and intonation in English and likely contributed to successful transfer to natural speech. It is also likely that materials produced by the participants themselves were more meaningful for them especially when the speech event was recreated in training both auditorily and visually. Ratings of speech samples one week after training revealed sustained improvement. Questionnaire results also supported the use of computer-based tools and authentic speech samples in L2 speech learning.

For methodological reasons, the above study focused on prosody as one measurable aspect of a speech event; however, other applications are possible. A tool such as *Anvil* can be configured to provide feedback in the form of comments on micro-level elements of L2 speech such as segmental performance as well as macro-level features of communication including beat gestures by a speaker that accompany the production of stressed syllables, iconic gestures that correspond to meaning, eye contact, facial expression, and other accompaniments to speech. This type of feedback can be beneficial to learners of any language and level of proficiency. Given the amount of information contained in these various channels, instructors might consider directing learners' attention to features one at a time. In addition, viewing videos of speech events produced by native speakers with time-aligned pitch contours could complement instruction in L2 speech. Given the beneficial role of the presence of visual cues from a talker's face in improving L2 perceptual accuracy (Hardison 2003), one might focus attention at one stage on the talker's lip movements. This type of auditory-visual input could also facilitate shadowing or mirroring exercises by L2 speakers (Hardison & Sonchaeng 2005). Shadowing is an exercise that involves repetition or echoing of a talker's utterance. In addition to this vocal repetition, mirroring involves imitating a speaker's posture, gestures and other movements (see Goodwin 2004).

To review, the context of a speech event has numerous components. Prosody is linked to the phonological, syntactic, semantic and pragmatic aspects of linguistic context and to the gestural and expressive components of the macro-level of communication (see Bolinger 1983). Technology provides the opportunity to integrate these components in L2 speech analysis and teaching.

Outstanding problems and future directions for research

Audio, video, and computer technology have made possible a vast increase in the amount and types of spoken language data that can be stored and analyzed. Archives of spoken language in digital form include the Santa Barbara Corpus of Spoken English (for information, visit http://www.linguistics.ucsb.edu/research/sbcorpus.html), the Speech Accent Archive (housed at George Mason University http://classweb.gmu.edu/accent) and The Michigan Corpus of Academic Spoken English (for information visit http://legacyweb.lsa.umich.edu/eli/micase/index.htm). Almost twenty years ago, Pennington (1989a) suggested that "Perhaps the most exciting possibilities combining language training, assessment, and research involve two-person interactions which are both facilitated and analyzed by a computer" (p. 119). At the same time as it makes language more accessible to researchers, the amount and complexity of information in such naturally occurring interactions is so great that it creates problems of analysis.

Solutions to these problems can be found in the technologies themselves. For example, based on speech input, a computer can extract and display prosodic features as a separate channel of information. This capability has resulted in new research and new understandings of prosody in English and other languages (Chun 2002, 2007). Using computer software, it has been possible, for example, to independently manipulate the audio signal in coordination with co-produced visible articulatory movements (especially of the lips) in order to research the contribution that the visible aspects of articulation make to speakers' perception of speech sounds. Through such manipulation, the specific effects of mismatches in these two channels have been studied, and it has also been determined that people perceive spoken language with greater accuracy when even extremely brief co-occurring mouth movements can be seen (Hardison 1999). The computer also allows manipulation of individual frames of a video for research focused on minute details of the audio-visual complex of speech. Once isolated, each channel can be studied individually and as part of the complex in which it occurs (Chun, Hardison & Pennington 2004).

Evolving technology provides the basis for continuing to isolate the components of spoken language in context in order to study their properties and functioning individually and in coordination with each other. The authors are engaged in exploratory research to isolate and visually highlight components of facial expression and hand gestures and their interaction with prosodic features, lexis, and syntax. A prototype system for highlighting the components and movements of hands, lips, eyes and eyebrows is in development, as is a transcription system for coding each of these channels in relation to prosodic features, lexis and syntax (Chun et al. 2004; Pennington, Chun, & Hardison 2002).

Expansion of research domain to include gestures and movement

Spoken language is part of a larger constellation of activity that includes movement of head and body, facial expression, and gestures of arms and hands (Clark 1996). Each of these components is further divisible; for example, facial expression incorporates mouth postures, eye gaze, eyebrow movements, etc. (McNeill 1992). Each component tends to be studied in isolation from the others. Yet in spoken language, syntactic patterning is linked to prosodic patterning, and prosodic patterning is linked to the larger constellation of activity within which speech occurs. For example, a rise in pitch often co-occurs with raised chin and eyebrows (Pennington 1996b: Ch. 4), and emphasis on a specific word is typically shown not only by articulatory stress as manifested in increased acoustic energy but also by different types of gestural "stress" such as a co-occurring brief increase in eye opening, bob of the head, or finger pointing gesture.

In graduate seminars at the University of California, Santa Barbara conducted annually by Lerner and Thompson (2004) titled "Language and the Body," the objective was to bring together the methods and findings of functional linguistics and those of conversation analysis in a dialogue centering on the visible behavior of the body in the organization of talk-in-interaction. A primary goal was the formulation of new understandings of the ways in which spoken language, including "grammar" and "meaning," is intertwined with gesture, gaze, and body movements in face-to-face interactions. This was to be accomplished on the basis of naturally occurring interaction, focusing on the social organization of gesture, gaze, and body behavior and their relationship to the use of language. While some researchers have characterized such body behaviors as "nonverbal communication," Lerner and Thompson focused on the interaction and integration of body behaviors and systematic aspects of spoken language. Such a focus holds considerable promise for the understanding and training of L2 prosody in context.

A pilot study carried out by the authors (Pennington, Chun, & Hardison 2002) based on digitized video clips of lecturers in a wide range of fields from the University of Warwick has demonstrated a strong tendency for multiple signals of these sorts to co-occur in coordination with prosodic features in spoken language. Support for this pilot can be found in a study by Loehr (2004), which investigated the relationship between gesture and intonation based on data collected during natural conversations. Gestures were coded according to the guidelines published by McNeill (1992) and colleagues, and the videos were processed with the *Anvil* software (Kipp 2001). Loehr found: (1) no evidence for Bolinger's (1983) hypothesis that pitch and body parts rise and fall together to reflect increased or decreased tension; (2) consistent alignment of the apexes of gestural strokes and pitch accents; (3) occasional but striking instances where gestural and intonational meanings converged; (4) a rich relationship and interplay among hands, head, and intonation.

In one of the few L2 studies involving gesture, Sueyoshi and Hardison (2005) investigated the contribution of gestures to L2 listening comprehension. A total of 42 ESL learners at different levels of proficiency were randomly assigned to three stimulus conditions: AV-gesture (auditory and visual presentation of hand gestures and facial cues), AV-face (auditory and visual of facial cues only), and A-only (auditory-only input). Comprehension was significantly better for the lower proficiency learners in the AV-gesture condition, whereas the higher proficiency learners performed best in the AV-face condition. The results demonstrated that facial speech cues such as lip movements may not contribute as much information for lower level learners in a larger discourse context of speech. The learners were positive about the value of visual cues to aid comprehension, and these positive attitudes coupled with the experimental findings of the study suggested the value of raising learner awareness to nonverbal aspects of communication.

Remaining issues and challenges

Despite the research progress of the last decade, problems exist in comparing findings across studies because of differences in subject populations, tools, instructions, amount and content of feedback, testing and training conditions and procedures, and other types of limitations in knowledge or technology. Longitudinal studies are lacking as are studies that examine the long-range effects and resilience of training. Limitations continue to exist for pedagogical applications because of equipment cost and the time needed to train instructors, although public domain tools such as *Praat, Anvil* and *CECIL (Computerised Extraction of Components of Intonation in Language)* have contributed to bridging the gap between technology and practical applications. In addition to training instructors, there is the problem of how much training learners require and what type is most effective. What is the potential, in particular, of computerized tools for self-study? It is likely that learners will need some tutorial assistance at least initially to suggest an appropriate sequencing of attention and activities to respond to their needs, but the hope is that prosody could be practiced in an electronic environment to a large extent individually. From a pedagogical point of view, matching available computerized tools with varied learner needs and objectives will continue to be a challenge for teachers, as will materials selection and provision of multiple native-speaker models.[9] There is the further issue of how to incorporate the use of technology

9. We note that following individual practice, learners will need to re-connect with their human tutors or instructors to receive feedback beyond what computer programs can provide.

focused on prosody into classrooms with a communicative approach. Of particular pedagogical importance is the need to maintain a dual emphasis on the micro- and macro-levels of communication with an appropriate sequence to focus learner attention. Incorporating all of these needs into computer-based instructional tools remains challenging.

Future directions for technology and for research

In terms of the technologies themselves, the future lies in tools applicable to any language that capture and integrate multiple components of the speech event for storage, data analysis, replay with comments for feedback and reference, all with the ability to handle large amounts of data. Future directions for research must consider the relationship between perception and production. Studies have shown that transfer of perceptual training (generally, at a segmental level) can improve production, but there have been fewer studies on the influence of production training on perception. A critical issue is whether training with computer-based tools transfers to face-to-face interaction. It has been shown to contribute to improved prosodic use in discourse-level speech, but its application to natural conversations has not yet been examined. Evidence of such transfer is difficult to obtain but would provide a link between focused training and the type of interaction that promotes negotiation and interlanguage development. Lastly, future research should analyze the integration of the multiple components of speech events by both native and nonnative speakers of a variety of languages.

Conclusion

Prosody in context can be conceptualized in two different ways. From the first perspective, that of linguistics, prosody is the context for lower level units. From the second perspective, that of communication, prosody is a unit within a higher level organization of behavior. This chapter has suggested an evolutionary path for research and practice on prosody within applied linguistics and second language acquisition from the first to the second of these perspectives. It has highlighted work making use of computer-based technology for research and teaching-learning applications according to both of these perspectives and outlined a direction for future work using digital video and multimedia tools. The discussion and examples show ways in which work in applied linguistics and L2 acquisition is helping to chart a course that is moving the language disciplines away from an isolationist, autonomous linguistics and towards a situated, contextual linguistics.

Suggested readings

Brazil, D. 1997. *The communicative value of intonation in English*. Cambridge: Cambridge University Press.

Chun, D. M. 2002. *Discourse intonation in L2: From theory and research to practice.* Amsterdam: John Benjamins.

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

Pennington, M. C. 1996. *Phonology in English language teaching: An international approach.* London: Longman.

Watt, D. L. E. 1994. *The phonology and semology of intonation in English: An instrumental and systemic perspective*. Bloomington, IN: Indiana University Linguistics Club Publications.

References

Bolinger, D. 1983. Intonation and gesture. *American Speech* 58(2): 156–174.

Boersma, P. 2001. PRAAT, a system for doing phonetics by computer. *Glot International* 5(9/10): 341–345.

Brazil, D. 1994. *Pronunciation for advanced learners of English.* Cambridge: Cambridge University Press.

Carey, M. 2004. Visual feedback for pronunciation of vowels: Kay Sona-Match. *CALICO Journal* 21(3): 571–601.

Cauldwell, R. 2002. Streaming speech: Listening and pronunciation for advanced learners of English. In D. Teeler (ed.), *Talking computers*, Whitstable, UK: IATEFL, 18–22.

Cauldwell, R., & Hewings, M. 1996. Intonation rules in ELT textbooks. *ELT Journal* 50(4): 327–334.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. 1996. *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge: Cambridge University Press.

Chafe, W. 2002. Prosody and emotion in a sample of real speech. In P. Fries, M. Cummings, D. Lockwood, & W. Sprueill (eds.), *Relations and functions within and around language*, London: Continuum, 277–315.

Chun, D. M. 1998. Signal analysis software for teaching discourse intonation. *Language Learning & Technology* 2: 61–77. Retrieved July 27, 2007 from http://llt.msu.edu/vol2num1/article4

Chun, D. M. 2002. *Discourse intonation in L2: From theory and research to practice.* Amsterdam: John Benjamins.

Chun, D. M. 2007. Technological advances in researching and teaching phonology. In M. C. Pennington (ed.), *Phonology in context*, London: Palgrave, 274–299.

Chun, D. M., Hardison, D. M., & Pennington, M. C. 2004. Technologies for prosody in context: Past and future of L2 research and practice. Paper presented at the AAAL Annual Conference, Portland, OR.

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

Clennel, C. 1997. Raising the pedagogic status of discourse intonation teaching. *ELT Journal* 51(2): 117–125.

Couper-Kuhlen, E., & Selting, M. (eds.). 1996. *Prosody in conversation*. Cambridge: Cambridge University Press.

de Bot, K. 1983. Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech* 26: 331–350.

de Bot, K., & Mailfert, K. 1982. The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly* 16: 71–77.

Derwing, T. M. 2008. Curriculum issues in teaching pronunciation to second language learners. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Derwing, T. M., Munro, M. J., & Wiebe, G. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48(3): 393–410.

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

Gick, B., Bernhardt, B., Bacsfalvi, P., & Wilson, I. 2008. Ultrasound imaging applications in second language acquisition. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Goodwin, J. 2004. The power of context in teaching pronunciation. In J. Frodesen & C. Holten (eds.), *The power of context in language teaching and learning*, Boston, MA: Thomson/Heinle, 225–236.

Gorsuch, G. J. 2001. Testing textbook theories and tests: The case of suprasegmentals in a pronunciation textbook. *System* 29: 119–136.

Hahn, L., & Dickerson, W. 1999. *Speechcraft: Discourse pronunciation for advanced learners.* Ann Arbor: The University of Michigan Press.

Hardison, D. M. 1999. Bimodal speech perception by native and nonnative speakers of English: Factors influencing the McGurk effect. *Language Learning* 49: 213–283.

Hardison, D. M. 2003. Acquisition of second-language speech: Effects of visual cues, context and talker variability. *Applied Psycholinguistics* 24: 495–522.

Hardison, D. M. 2004. Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology* 8: 34–52. Retrieved July 27, 2007 from http://llt.msu.edu/vol8num1/hardison

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

Hardison, D. M., & Sonchaeng, C. 2005. Theatre voice training and technology in teaching oral skills: Integrating the components of a speech event. *System* 33: 593–608.

Henrichsen, L., Green, B., Nishitani, A., & Bagley, C. L. 1999. *Pronunciation matters.* Ann Arbor: The University of Michigan Press.

Hew, S.-H., & Ohki, M. 2004. Effect of animated graphic annotations and immediate visual feedback in aiding Japanese pronunciation learning: A comparative study. *CALICO Journal* 21(2): 397–420.

Jenkins, J. 2002. A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics* 23(1): 83–103.

Jenkins, J. 2004. Research in teaching pronunciation and intonation. *Annual Review of Applied Linguistics* 24: 109–125.

Kaltenboeck, G. 2001. A multimedia approach to suprasegmentals: Using a CD-ROM for English intonation teaching. In *Phonetics Teaching and Learning Conference Proceedings,* 19–22. Retrieved April 23, 2004 from http://www.phon.ucl.ac.uk/home/johnm/ptlc2001/ptlc2001_abstracts.htm

Kipp, M. 2001. *Anvil* – A generic annotation tool for multimodal dialogue. In *Proceedings of the 7th European Conference on Speech Communication and Technology*, Aalborg, Denmark: Eurospeech, 1367–1370.

Leather, J. 1983. Second-language pronunciation learning and teaching. *Language Teaching* 16: 198–219.

Leather, J. 1990. Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In J. Leather & A. James (eds.), *New Sounds 90*, Amsterdam: University of Amsterdam, 72–97.

Leather, J., & James, A. 1991. The acquisition of second language speech. *Studies in Second Language Acquisition* 13: 305–341.

Lerner, G., & Thompson, S.A. 2004. Language and the body. Graduate seminar, University of California, Santa Barbara.

Levis, J. M. 1999. Intonation in theory and practice, revisited. *TESOL Quarterly* 33(1): 37–63.

Levis, J. M. 2001. Teaching focus for conversational use. *ELT Journal* 55(1): 47–54.

Loehr, D. P. 2004. *Gesture and intonation*. Unpublished doctoral dissertation, Georgetown University, Washington, D.C.

Major, R. C. 2001. *Foreign accent*. Mahwah, NJ: Lawrence Erlbaum.

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

Morley, J. 1991. The pronunciation component in Teaching English to Speakers of Other Languages. *TESOL Quarterly* 25(3): 481–520.

O'Connor, J.D., & Arnold, G. F. 1961. *Intonation of colloquial English* (2nd ed., 1973). London: Longman.

Park, J. K. 2003. Pronunciation and communicative competence. *As we speak* 6(1): 5.

Pennington, M. C. 1989a. Applications of computers in the development of speaking and listening proficiency. In M. C. Pennington (ed.), *Teaching languages with computers: The state of the art*, La Jolla: Athelstan, 97–121.

Pennington, M. C. 1989b. Teaching pronunciation from the top down. *RELC Journal* 20: 20–38.

Pennington, M. C. 1996a. *Phonology in English language teaching: An international approach*. London: Longman.

Pennington, M. C. (ed.) 1996b. *The Power of CALL*. Houston: Athelstan.

Pennington, M. C. 1998. The teachability of phonology in adulthood: A re-examination. *International Review of Applied Linguistics in Language Teaching* 36(4): 323–341.

Pennington, M. C. 1999. Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning* 12(5): 427–440.

Pennington, M. C. 2000. Focusing learners' attention on prosody. Revised version of a plenary address delivered on 20 May 2000 at the University of Durham Seminar on Exploring Research-based Innovations for Teaching L2 Pronunciation.

Pennington, M. C. 2002. Real language phonology: Optimality Theory meets Sociolinguistics. *Journal of Sociolinguistics* 6(3): 418–448.

Pennington, M. C., Chun, D. M., & Hardison, D. M. 2002. *VISI-PRO Video-based pronunciation training system*. Revenue Investment Fund Project, University of Luton. Principal Investigator, M. C. Pennington; Consultants, D. M. Chun and D. M. Hardison.

Pennington, M. C. & Clark, B. Z. 2002. Review of three books on Optimality Theory: Kager (1999), McMahon (2000), and Tesar & Smolensky (1999). *Language in Society* 31(3): 443–449.

Pennington, M. C., & Ellis, N. C. 2000. Cantonese speakers' memory for English sentences with prosodic cues. *Modern Language Journal* 84(3): 372–389.

Pennington, M. C., Ellis, N. C., Lee, Y. P., Lau, L., & Lock, G. 2002. Instructing intonation in a second language: Lessons from a study with Hong Kong Cantonese undergraduate English majors. Unpublished ms.

Pennington, M. C., & Richards, J. C. 1986. Pronunciation revisited. *TESOL Quarterly* 20(2): 207–225.

Pickering, L. 2001. The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly* 35: 233–255.

Pickering, L. 2002. Patterns of intonation in cross-cultural communication exchange structure in NS TA and ITA classroom discourse. In *Proceedings of the Seventh Annual Conference on Language, Interaction and Culture* at the University of California, Santa Barbara, 1–17.

Pierrehumbert, J. 1980. *The phonology and phonetics of English intonation*. Unpublished doctoral dissertation, MIT, Cambridge, MA.

Pierrehumbert, J., & Hirschberg, J. 1990. The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollock (eds.), *Intentions in communication*, Cambridge, MA: MIT Press, 271–312.

Seidlhofer, B., & Dalton-Puffer, C. 1995. Appropriate units in pronunciation teaching: Some programmatic pointers. *International Journal of Applied Linguistics* 5(1): 135–147.

Setter, J. 2003. Review of Richard Cauldwell, *Streaming speech: Listening and pronunciation for advanced learners of English*. Birmingham: Speech in Action, 2002. Windows CD-ROM, *Journal of the International Phonetic Association, 33*(2).

Sueyoshi, A., & Hardison, D. M. 2005. The role of gestures and facial cues in second-language listening comprehension. *Language Learning* 55: 661–699.

Verhofstadt, K. 2002. *A critical analysis of commercial computer-assisted pronunciation materials*. Unpublished doctoral dissertation, University of Ghent. Retrieved July 29, 2002 from http://members.tripod.com/katrienverhofstadt/

Weltens, B., & de Bot, K. 1984. Visual feedback of intonation II: Feedback delay and quality of feedback. *Language and Speech* 27: 79–88.

Wennerstrom, A. 1994. Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics* 15(4): 399–420.

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

Wennerstrom, A. 2000. The role of intonation in second language fluency. In H. Riggenbach (ed.), *Perspectives on fluency*, Ann Arbor, MI: University of Michigan Press, 102–127.

Wennerstrom, A. 2001. *The music of everyday speech*. Oxford: Oxford University Press.

Wichmann, A. 2000. *Intonation in text and discourse*. Harlow: Longman.

CHAPTER 13

Curriculum issues in teaching pronunciation to second language learners

Tracey M. Derwing
University of Alberta

Introduction

This chapter critically examines the current state of pronunciation instruction in the second language classroom. As such, it touches upon many of the theoretical issues addressed in other chapters of this volume and relates them to the practical challenges of trying to help learners improve their pronunciation in a second language (L2). The chapter first surveys the key issues that instructors must consider when designing pronunciation courses and argues that the primary aim of instruction is improved intelligibility and comprehensibility of L2 learners. The following section of the chapter is an examination of relevant research findings that have implications for the nature and content of the L2 pronunciation curriculum, including factors that influence pronunciation and intelligibility. Issues related to instructional practice and teacher preparation are then discussed in detail. The chapter concludes with the identification of some of the most pressing needs for future research in L2 pronunciation teaching and curriculum design.

Why teach pronunciation?

In any discussion of the value of teaching pronunciation to second language learners, three primary considerations are (a) the context in which a learner communicates, (b) the learner's perceived need or desire for pronunciation instruction, and (c) the speaker's intelligibility. As Jenkins (2002) has indicated, most students of English as an International Language (EIL) will find that much of their communication is with other nonnative speakers of English, most often people who use English as a lingua franca to conduct business or for other instrumental reasons. Their needs may be somewhat different from those of learners of English as a Foreign Language (EFL) who may be interested in working with native English speakers in their own countries, or in English-speaking countries abroad. Other EFL students may not be particularly concerned with issues of pronunciation at all. Some, for example, may simply want to gain a reading knowledge of English in order to access materials that are not available in their first language. Another substantial constituency is that of English as a Second Language (ESL) learners: individuals who have moved permanently to a largely English-speaking country such as Australia, New Zealand, the United States, or Canada, many of whom wish to integrate into the local society, both socially and through employment. ESL students will likely interact with other nonnative speakers of English (NNSs) as well as native speakers (NSs) and, because they will encounter English in most parts of their lives, good communication skills are invaluable. For English teachers, a consideration of the milieu in which their students find themselves is critical in designing a curriculum that adequately addresses pronunciation needs.

In addition to context, teachers should be aware of their students' own perceived pronunciation goals. Timmis (2002), in a survey of 400 EFL, EIL and ESL students from 14 countries, found that 67% would prefer to speak English like a native speaker. Fully 95 of the 100 adult ESL students interviewed by Derwing (2003) reported that they would like to pronounce English like a native speaker, many of them because they felt they would be respected more if they did not speak with an L2 accent. Other learners are interested in maintaining some aspects of their accent as an identity marker (Gatbonton, Trofimovich, & Magid 2005; see also Hansen Edwards, Chapter 9 this volume). As will be discussed below, only a very small percentage of L2 speakers are able to achieve native-like status, but clearly, the goals of learners vary, a fact that teachers must bear in mind when planning a curriculum that involves pronunciation.

Taking into consideration the learners' goals, the primary aim of the pronunciation instructor should be improved intelligibility within the context in which the learners find themselves as opposed to general accent reduction (for a comprehensive overview of work on intelligibility, see Munro, Chapter 7, this volume). Those elements of an L2 learner's speech that interfere with a listener's comprehension should be the focal point for any instruction. Anderson-Hsieh, Johnson, and Koehler (1992) and Munro and Derwing (1995) presented evidence that intelligibility is particularly affected by suprasegmental features. These studies confirm the teaching experiences of pronunciation experts who have called for a greater emphasis on suprasegmentals (e.g., Firth 1992; Gilbert 1984; Pennington 1989). Following a classroom experiment in which one group of ESL students received segmental instruction and a second received lessons that focused on suprasegmentals, Derwing, Munro and Wiebe (1998) found that only the latter group's extemporaneous productions had improved in comprehensibility. In other words, pronunciation teaching that was based primarily on suprasegmental aspects of speech affected the ESL students' productions to the extent that listeners found them to be easier to understand as a result of the instruction. This study is an indication that pronunciation instruction can effectively enhance intelligibility. As Chun, Hardison and Pennington note in Chapter 12, prosody-focused instruction may also improve segmental features. However, there is no documented evidence that any type of pronunciation instruction can fully eliminate an accent.

Factors that affect success

Several factors affect pronunciation instruction success, including motivation, first language (L1), aptitude, age, social identity, exposure and choice of instructional approach. In an examination of 61 learners, Purcell and Suter (1980) found that concern with pronunciation accuracy was significantly correlated with pronunciation. Moyer (1999) conducted a study with advanced second language learners in which she found that phonological attainment was correlated with professional motivation. It stands to reason that motivation, which is connected to progress in other aspects of language learning, would also have an effect on pronunciation.

The phonological distance between L1 and L2 also appears to have a bearing on pronunciation. Bongaerts, van Summerin, Planken, and Schils (1997) had NSs judge the productions of Dutch speakers who had learned English post-puberty, and many of them were assessed as native-like English speakers. In a later study, Bongaerts, Mennen and van der Slik (2000) collected speech samples from a group of L2 speakers of Dutch who learned their L2 in adulthood and who came from a wide variety of language backgrounds. Although the proficiency level of all of the participants was extremely high, the only ones who were judged to have native-like accents were people whose L1 was closely related to Dutch (German and English). None of the participants from typologically unrelated language backgrounds, e.g., Berber, were assessed as having native-like pronunciation.

Aptitude is another variable that plays a role in determining the degree of accentedness retained by a NNS. Ioup, Boustagi, El Tigi, and Moselle (1994) identified some extraordinary native English speakers whose productions of Arabic were deemed native-like, despite the fact that they learned their L2 as adults. They attribute this unusual finding to an exceptional linguistic aptitude. This justification is intuitively pleasing to teachers who see differences in performance in their classrooms that are not easily explained by other factors. Why is it that two learners from the same L1 background, who started learning their L2 at the same time, under similar conditions, and who appear to share the same levels of motivation can be markedly different in their ability to produce their new language with relatively accurate pronunciation? In the absence of other explanations, aptitude seems likely.

Age of learning the L2 has long been viewed as a crucial factor in ultimate attainment, particularly with regard to pronunciation (for a detailed discussion of this issue, see Chapter 2 by Ioup, this volume). For several decades it was believed that there is a critical period for second language learning coinciding with the onset of puberty. However, in the most comprehensive study of age of acquisition, Flege, Munro and MacKay (1995) determined that even young children who started learning their second language before the age of six are sometimes identifiable as NNSs. This study nevertheless suggests that age matters, and that the older the learner is, the more likely he or she is to retain aspects of the L1 in L2 productions. What is not clear at this point is whether age has a bearing on L2 phonological acquisition in senior adult learners. In a study by Burda, Scherz, Hageman, and Edwards (2003), older listeners apparently had more difficulty understanding foreign accents than did younger adults. Whether they would also have more difficulty mastering L2 phonological patterns than younger adults is an intriguing empirical question.

Degree of exposure to the L2 appears to have an effect on pronunciation. Munro (1993) found that Arabic speakers' production of English vowels improved with increased exposure to English. Flege, Bohn, and Jang (1997) compared accent ratings of L2 speakers who had been in the United States for an average of seven years with those of speakers who had been there for an average of a little over half a year. The ratings favoured the experienced speakers. Another study that addresses exposure suggests that extent of use of the L1 may affect pronunciation in the L2. Flege, Frieda and Nozawa (1997) examined the relative use of Italian and English among age-matched Italian immigrants to Canada, and found that those who reported greater use of Italian were judged to have stronger accents when speaking English (see also the discussion by Hansen Edwards, Chapter 9 this volume, on social factors influencing L2 accent). Piller (2002), in a study of cross-cultural marriages between English and German speakers, found that several participants claimed to be able to "pass" as native speakers of their L2 in first encounters with strangers. Although the findings are based on self-report alone, they suggest that massive amounts of exposure may result in a close to native-like accent.

Many of the factors discussed thus far are beyond the control of the pronunciation teacher, although presumably motivation and amount and quality of exposure could be influenced to a certain degree. However, one further variable that has a demonstrable effect on L2 productions is the efficacy of the pronunciation instruction itself. Until recently there were very few studies of pronunciation instruction outcomes. Bongaerts et al. (1997) indicated that of the five speakers who were judged to sound like NSs in their study, all reported having had formal pronunciation instruction. Derwing, Munro, and Wiebe (1997, 1998) conducted two before/after studies in which students received suprasegmental instruction; both studies provided evidence of increased intelligibility as a result of pronunciation teaching. In Derwing et al. (1998), another group of students received only segmental instruction. Their post-test speech samples were not judged to have improved in comprehensibility, despite a clear indication (details in Derwing & Rossiter 2003) that they produced segmentals more accurately at the post-test; thus the focus of instruction makes a difference to outcomes. Similarly, Elliott (1997) conducted pre- and post-tests with English-speaking students of Spanish who had received segmental instruction. He found no significant improvements in spontaneously produced speech. Although Couper (2003) did not obtain measures of intelligibility for the productions of students who attended a pronunciation course that included both segmental and suprasegmental work, he did identify errors in pre- and post-instruction tests (a reading task and a free production task). Students showed improvement, that is, fewer errors, on both tasks after instruction. However, without obtaining direct measures it is impossible to know which of the changes the students made in their productions would have an effect on intelligibility. Ultimately, the choices a pronunciation teacher makes should be based on factors that have been shown to influence intelligibility and comprehensibility. There is no reason to suppose that accent reduction, in and of itself, will necessarily result in improved listener understanding, because some aspects of an accent appear to have little effect on intelligibility (Munro & Derwing 1995). The preceding studies indicate that the choice of focus can make a significant difference to the overall efficacy of instruction, in that suprasegmentals appear to have more impact on overall intelligibility and comprehensibility than segmentals.

Pronunciation curriculum development

Needs analysis

In any curriculum development project the first task is to undertake a needs analysis. In the case of pronunciation, individual assessments of students' needs are highly advisable whether or not the learners share an L1 background. Many texts offer a diagnostic tool to help the teacher identify speech errors. Gilbert (2005), for instance, provides both a listening and a speaking test to assist teachers in developing a pronunciation profile for each student. (These tests are helpful for needs analysis, but caution is necessary in using them to assess progress – see Gorsuch 2001). Firth (1992) also suggests a student diagnostic profile that serves to guide the instructor in determining where the student's problems lie. These assessment tools deal with both segmental and suprasegmental information. Firth's includes a *General Speaking Habits* section that encompasses aspects of an individual's speech behavior that are often ignored in pronunciation materials, but which clearly affect comprehensibility. If a student doesn't speak with sufficient volume, for example, he or she will not be intelligible, regardless of how well-pronounced the speech sample is. The diagnostic information that these tools and others provide serves to help teachers monitor several aspects of students' productions, though not all of them have a listening component. However, the students' perceptual skills may be as important as their ability to produce, and it is thus essential that the teacher test both perception and production.

Several explanations have been proposed for some of the difficulties experienced by ESL students. According to current models of L2 speech learning, the most common cause of pronunciation problems is perceptual: a student may not hear a given contrast the same way that a native speaker does (see the discussion on perception by Strange and Shafer in Chapter 6 of this volume). Indeed, if a particular sound in the L2 is relatively close to a sound in the L1, the student will sometimes have more difficulty perceiving a difference than if a new L2 sound is completely distinct from anything in the L1 inventory (Best 1995; Flege 1995). The resulting errors can be of two types: students may transfer an L1 form directly into the L2 inventory (e.g., some speakers of Spanish dialects will replace English /y/ as in 'yes' with /dʒ/, or two L2 categories may be perceived as a single L1 sound (for example, Spanish learners of English may hear both /ɪ/ and /i/ as /i/).

Not all mispronunciations are based on erroneous perception, however. At the word level, for instance, students may simply not know how a given lexical item should be pronounced; they may have no trouble saying the word accurately once they hear a correct model, but because of a faulty representation (perhaps influenced by an odd spelling), they produce the item inappropriately. Finally, students may have an accurate perceptual representation of an L2 form but may struggle with producing it satisfactorily, for articulatory reasons.

The importance of assessing each person individually cannot be over-emphasized. There are numerous materials that characterize the errors that L2 speakers of particular languages make (Nilsen & Nilsen 1971; Swan & Smith 1987), but not all speakers of a given language will have these errors, and some individuals will exhibit problems that are not shared by others from the same L1. To illustrate, Cantonese speakers are generally portrayed as having difficulty distinguishing /l/ from /r/, yet some individuals will actually have more trouble with the contrast between /l/ and /n/.

Jenkins (2002) has suggested that for EIL speakers, there is a predictable "core" set of linguistic features that should be taught, most of which are segmentals. While this may be so, the available evidence is very limited, based on a small sample of communication breakdowns across very few learners. Until considerably more work is done in this area, both observational and experimental, it is safer to rely on individual assessments with reference to research on intelligibility than to assume that the proposed core is adequate for all EIL learners.

Factors that influence pronunciation and intelligibility

Of the many features that one should attend to in a needs analysis, those that have been shown to affect intelligibility are the most important (see Chapter 7 by Munro, this volume). Many facets of pronunciation have been identified in the literature as having an influence on intelligibility, including general speaking habits, voice quality, several aspects of intonation, primary (nuclear) stress and segmentals, as well as predictability of syntax, lexical choice and discourse markers.

As mentioned above, Firth (1992) identified several general speaking habits, such as volume, eye gaze, clarity (e.g., hand in front of mouth, poor posture) that can influence intelligibility. Obviously if a listener has a great deal of difficulty hearing the productions of an L2 learner, intelligibility will be severely compromised. Firth also suggested that speech rate affects intelligibility. In a series of experiments investigating rate, Munro and Derwing (2001) determined that the common caution to second language learners to 'slow down' is not appropriate across the board; indeed most L2 speakers naturally produce language more slowly than NSs. There appears to be a U-curve in performance, such that quite slow speakers would benefit from speeding up and very fast talkers should probably lower their speech rate slightly. When teachers who have had little or no pronunciation training evaluate an L2 speaker who is difficult to understand, they may be prone to blaming rate: it may serve as a scapegoat for listeners who actually have difficulties with other phonological problems that listeners have trouble describing.

Voice quality, the long-term characteristics of a speaker's vocal output, is also related to L2 accent, and as Munro, Derwing and Burgess (2003) have shown, it is extremely salient to NS listeners. In a series of backwards speech experiments these authors found evidence that listeners were able to distinguish between native and nonnative speakers on the basis of voice quality alone, not only at the level of the sentence, but even when they heard a single word. Esling and Wong (1983) have suggested that when a voice quality that is characteristic of a speaker's L1 is radically different from that of speakers of the L2, intelligibility is at risk. For instance, factors such as retroflexion and creaky voice (vocal fry) may interfere with comprehensibility. Esling and Wong recommend that teachers bring students' attention to voice quality by having them observe and make notes while watching NSs on TV, by imitating English-speakers' accents in their own L1, and by making comparisons of the voice quality in L1 phrases produced by several students in mixed L1 classes. Jones and Evans (1995) also advocate teaching voice quality in the pronunciation classroom and recommend activities that help L2 students recognize the connections between voice quality and the expression of emotion. Kerr (2000) conducted a case study with a Vietnamese speaker in which she worked on improving voice quality, in part by training the individual to relax his tongue, to open his jaw to a greater extent, and to use lip rounding and lip spreading more effectively. She had NSs transcribe some of the student's utterances both before and after 12 instructional sessions. Although the results were not conclusive, Kerr noted that the "judges were better able to understand longer sections of utterances" post instruction (p. 9).

Wennerstrom (1994) compared several aspects of Spanish, Thai, and Japanese speakers' intonation in English with NS productions, and found that the learners did not use pitch contrasts in the same way as native speakers. In a later study (1998), she tested the hypothesis that the stronger their grasp of the English intonation system, the better NNSs performed on a global test of English proficiency. Her analysis of the intonation of Mandarin speakers of English revealed that the paratone (an increase in pitch that signals topic shift) predicted test scores. She suggested that intonation carries meaning and creates cohesion in discourse, and, like Chun (1988) and Clennell (1996), called for a discourse-level approach to intonation teaching and the use of authentic texts. Levis and Pickering (2004), in an examination of sentence-level versus discourse-level intonation, recommend that technology be employed to teach intonation at the discourse level, along with clear explanations of the meaning of intonational choices (see also Chun et al., Chapter 12 of this volume, for a detailed discussion of research on L2 prosody and intonation).

In a careful comparison of the lectures of native speaker teaching assistants at an American university with those of international teaching assistants (ITAs), Pickering (2001), using Brazil's (1997) model of intonation, determined that the performance of the two groups was significantly different, such that the ITAs underused rising tone choice. She suggested that their speech would be interpreted negatively by students as being a sign of unattractive personality traits. Even though they were lecturing, the NSs used a dialogic model, assuming "cooperative achievement" (p. 253) whether their students responded orally or not, whereas the ITAs may have distanced themselves from the students because of their tone choice. Pickering recommends that L2 learners be made aware of the pragmatic nature of rising, falling, and steady tones.

Hahn (2004) undertook an innovative study in which she had a highly proficient native speaker of Korean read three identical mini-lectures in English, in which assignment of primary stress had been manipulated. Three groups of undergraduate students were then asked to listen to the lectures and both respond to comprehension questions and rate the quality of the lecture. She discovered that the students who heard the lecture with the appropriate stress assignment understood significantly more than either of the other two groups. These students also judged the speaker more favourably than did the other groups.

Segmentals have long been the mainstay of many pronunciation programs, and certain segmental difficulties are strikingly salient in L2 accented speech. The

English interdental fricatives are often identified by L2 students as their most troublesome sounds (Derwing 2003), but it has been compellingly argued by Catford (1987) and Brown (1991) that these are not very important sounds because they carry a low functional load. That is to say, the inability to produce certain segments, such as /p/ and /b/ for instance, is far more likely to cause a breakdown in communication than other less frequently occurring segments. Brown has argued that the cumulative frequency of English phonemes and the number of minimal pairs be considered in determining which segments are most important. On the basis of such information he provides a rank ordering of vowels and consonants to guide teachers in deciding which segments their students may need to work on. Munro and Derwing (2006) have carried out a preliminary study in which they tested the arguments for functional load, by examining the intelligibility of low and high functional load errors. They found some evidence to suggest that Catford's (1987) and Brown's (1991) approach is correct.

A factor that appears to interact with L2 accent in influencing comprehensibility ratings is the predictability of both grammatical structures and vocabulary. Varonis and Gass (1982) determined that naïve listeners' perceptions of pronunciation were affected by the grammaticality of utterances. That is, ungrammatical, and thus somewhat unpredictable structures, led to harsher judgments of pronunciation, which the authors argued was indicative of increased processing costs on the part of the listener. In a complementary study they also found that familiarity promotes comprehensibility (Gass & Varonis 1984). Schmid and Yeni-Komshian (1999) also determined that predictability plays a role at the level of individual vocabulary items. They conducted a listening study in which mispronunciations of predictable and less predictable words were targeted. Listeners had an easier time recognizing the meaning of the mispronounced version of 'carpet' in the phrase *shag garpet* than *rag garpet*. Schmid and Yeni-Komshian suggest that "the increased variability [that is] characteristic of non-native speech" (p. 56) requires additional processing on the part of the listener. In a longitudinal study of pronunciation development, Derwing and Thomson (2004) worked with an individual who prided himself on his diverse vocabulary, but his predilection to use low frequency words such as 'halt' where a native speaker would say 'stop,' and 'draw' in a context where a native speaker would say 'pull,' interfered with his intelligibility. The lack of lexical predictability involved, combined with pronunciation problems, made this person quite difficult to understand. Thus classroom work on idiomatic expressions, predictable lexical chunks, and oral grammar are likely to benefit listeners' perceptions of L2 learners' productions.

Finally, a number of discourse factors will also influence listeners' perceptions of accented speech, including lexical discourse markers and degree of lexical specificity. Tyler (1992) asked a native speaker of English to deliver two lectures (one of which had originally been spoken by a native speaker of Mandarin, and the other by an English native speaker). Even with an L2 accent removed, the lecture that was developed by the Mandarin speaker was more difficult to understand. Tyler attributed this finding to a lack of cohesion caused by missing discourse markers.

Instruction

Mixed vs. same L1 classes

The question of whether pronunciation is better taught in mixed versus same L1 classes is moot in many situations as there is often no choice: either the students share the same L1 (this is the usual case in EFL settings) or classes are linguistically diverse (most ESL and many EIL contexts). There are advantages and disadvantages to both. In same L1 settings, the students are likely to share a number of problems at both the segmental and suprasegmental levels, whereas in mixed L1 classes, the variability at the level of segments is bound to be tremendous. In that case, it may be somewhat easier for the instructor of a homogeneous class to ensure that the activities will be beneficial to most of the students. The disadvantage is that it may be easier for the teacher to miss individual differences. Furthermore, much of the input to which the students are exposed in shared L1 classes (that is, the speech of their fellow classmates) may reinforce L1 patterns. In addition, in some, but not all instances, there appears to be a minor advantage to intelligibility if an L2 learner interacts with someone who has the same L1 (Major, Fitzmaurice, Bunta, & Balasubramanian 2002; Munro, Derwing & Morton 2006), thus students in homogeneous classes may have a somewhat skewed impression of how clear their speech actually is. In mixed L1 classes, on the other hand, students hear a variety of L2 productions that give them helpful listening practice, and they also have to work harder to make themselves comprehensible to other members of the class.

Occasionally, large ESL programs offer courses tailored to a particular L1 group. In the early 1980s, for example, when substantial numbers of Vietnamese, Laotian and Cambodian refugees came as immigrants to North America, shared L1 pronunciation courses were established in many cities because the nature of these newcomers' pronunciation difficulties appeared to be both quantitatively and qualitatively distinct from those of other ESL students. The advantages and disadvantages cited above hold true in ESL settings, but in some instances a mixed class is inappropriate if there is extreme linguistic distance between English and only some of the L1 languages involved. For example, many Vietnamese are completely unfamiliar with the notion of rhyme in English, whereas students from European language backgrounds are fully aware of rhyme in their own first languages and are able to relate to rhyme in English immediately. Having students from both these backgrounds in the same pronunciation class might prove frustrating for everyone concerned.

Regardless of the nature of the class, mixed or shared L1, ESL, EFL or EIL, it is important that other models of L2 speech be provided. In EFL and some EIL settings, media such as radio, television, movies, books on tape and the Internet may all provide a range of voices, dialects, and accents to augment classroom-based input. English language media as well as teacher-directed contact activities are also useful in ESL settings, where one might think that there are many opportunities for authentic interactions with native speakers. In fact, the only significant contact some students have outside of an ESL class is with compatriots who share the same L1.

Stand-alone classes vs. incorporation into the general language curriculum

As Levis and Grant (2003) suggest, there are generally two types of approaches to teaching pronunciation and speaking. First, there is the stand-alone class in which students are supposed to move from controlled practice of pronunciation to communicative activities. The authors point out that in many of these classes the communicative aspects end up being ignored in favour of a strong emphasis on controlled practice. On the other hand, Levis and Grant propose that in most classes where the focus is on speaking, pronunciation practice is unsystematic or even non-existent. The authors agree with Murphy (1991), who argued that pronunciation is a part of oral proficiency and one should not be isolated from the other. Levis and Grant put forward a set of guiding principles for incorporating pronunciation into ESL/EFL classrooms that would bring about a balanced approach. These principles are: (1) "to aim for a primary though not an exclusive focus on suprasegmentals"; (2) "to maintain a central focus on speaking in the class"; and (3) "that pronunciation instruction should fit the constraints of the speaking task" (p. 14).

Chela-Flores (2001) also advocates the inclusion of pronunciation in the general L2 curriculum, rather than in stand-alone classes. She notes that many programs that offer stand-alone classes require students to be at least at an intermediate or advanced level of proficiency because most pronunciation materials assume a relatively strong grasp of the language, particularly metalinguistic terminology. Chela-Flores argues that it makes more sense to include pronunciation as part of the curriculum right from the beginning, focusing on meaningful units of speech (that is, thought groups, phrases and short sentences), such that students' attention is brought to bear on rhythm and intonation in a non-technical way at the earliest stages of language learning. Students' awareness of these patterns can then be 'recycled' (reinforced throughout the program), unlike the situation in many stand-alone classes, where a distinct facet of pronunciation is covered in each lesson. Seidlhofer and Dalton-Puffer (1995) also support the notion of using larger units of pronunciation units or 'phonological chunks' as the basis of instruction. They suggest teaching phrases that have set intonation patterns, such as 'in no time' and 'for a change'. These chunks serve to enhance a learner's overall pronunciation in the same way that formulaic sequences help a speaker to sound fluent.

Textbooks and technology

Compared to other aspects of second language acquisition such as grammar, there are relatively few pronunciation reference books intended for teachers, and of those, very few make substantive reference to research (three notable exceptions are Celce-Murcia, Brinton & Goodwin 1996; Chun 2002; and Dalton & Seidlhofer 1994). Moreover, many general preparation books for second language teachers give short shrift to pronunciation teaching (e.g., Davies & Pearse 2000; Hedge 2000; Nunan 1999). Many pronunciation materials have been available for classroom use for several decades, but until lately, most have focused heavily on segmentals, even if the title suggests otherwise (e.g., *Pronouncing American English: Sound, Stress and Intonation* [Orion 1997]). The few older materials that did incorporate suprasegmentals, such as *Jazz Chants* (Graham 1978) and *Clear Speech* (Gilbert, originally published in 1984 and now in 3rd edition) are still popular among L2 teachers (Breitkreutz, Derwing & Rossiter 2001). More recently, several student books have been published that deal extensively with suprasegmental features (e.g., Grant 1993; Hahn & Dickerson 1999; Hewings & Goldstein 1999; Matthews 1994; Reed & Michaud 2005). There is a caveat, however, on the usefulness of many of the exercises and activities even in more recent textbooks. As Cauldwell and Hewings (1996) and Levis (1999) have pointed out, despite advances in our understanding of intonation, and despite numerous suggestions for approaches to teaching intonation that are communicative in nature, many commercial materials continue to teach a subset of intonation patterns in a way that is not likely to serve L2 students well. Levis recommends the following four principles for the teaching of intonation: (1) teach intonation in an explicit context; (2) make learnable and generalizable statements about meaning; (3) teach intonation in the context of a communicative focus; and (4) teach intonation with realistic language.

Advances in technology have cleared the way for pronunciation materials that students can use without direct supervision; however, teachers should be wary of many of the electronic pronunciation programs available. The possibilities for truly superb materials are great, but at this point, many of the technological publications have not exploited the potential benefits, partly because there appears to have been much less input from people who understand pronunciation than from designers who work on the bells and whistles or the "Wow" factor as Murray and Barnes (1998) called it. In Breitkreutz et al.'s (2001) survey, for instance, the top three CDs favoured in ESL programs all focused exclusively on segmentals. Although some technological resources deal with suprasegmentals (e.g., *Streaming Speech,* Cauldwell [2002] and *Connected Speech*, Westwood & Kaufman [2005]), more work is needed in this area. A number of training techniques using new technology that have been employed in research settings could be valuable models for commercial development (see Chun, Hardison & Pennington for a comprehensive review of the use of technology to teach prosody in context, Chapter 12 this volume, and Wang & Munro 2004).

Measuring improvement

Ongoing assessment of the effects of instruction is useful, not only because it provides the teacher with guideposts as to what is working and what is not, but it is also beneficial for the student, who may not believe that changes are actually taking place. The first signs of progress one would look for have to do with awareness. For example, many L2 students who have a relatively good grasp of vocabulary and grammar do not realize until it is pointed out to them that a change in primary stress makes a difference in meaning (e.g., Does *he* speak Russian? vs. Does he speak *Russian*?) It is a revelation for many learners to discover that the same words in the same order can have different meanings, and, in the case of the questions in the example here, they would elicit different responses. Assessment, then, should include perception tasks that deal with the pronunciation elements covered by the teacher.

Production can be evaluated in a variety of ways. Clearly a narrow transcription of students' speech samples will indicate whether changes have begun to occur, but intelligibility and comprehensibility measures are also a good indication of improvement (cf. Munro, Chapter 7, this volume). Indeed, until we have a better sense of which factors directly affect intelligibility and comprehensibility, feedback from listeners is probably the best window on the students' progress. There are several ways of obtaining listener assessments, including dictation tasks (in which listeners write down what they hear the L2 student saying), rating tasks (in which listeners make comprehensibility judgments as to how easy or difficult an utterance is to understand), and videos or audio recordings of student presentations, followed by comprehension questions. In each instance, the inclusion of listeners other than the teacher is essential; collaboration with an L2 teacher preparation program would be very helpful here. Student teachers would gain experience listening in an analytic manner to L2 speech, and the pronunciation teacher could access a ready pool of listeners to assess intelligibility and comprehensibility. Another group of listeners who would provide invaluable feedback regarding intelligibility are other L2 speakers who are representative of the students' NNS interlocutors. In ESL, EIL and EFL settings, students from other classes in the same program could be enlisted to assess intelligibility and comprehensibility.

Intelligibility measures can, however, be misleading, as Zielinski (2004) has pointed out. For example, if a listener writes down nine of the ten words in an utterance successfully but misses a word that is critical to understand the utterance, a score of 90% does not adequately characterize the degree of comprehension. Despite such difficulties, however, the responses of listeners are the best way to gauge whether there is actual improvement in L2 learners' productions.

Social issues

Thus far the curriculum elements recommended here deal exclusively with matters of language perception and production, but there is another issue that is of critical importance, whether in a stand-alone course or integrated into a communicative language classroom. The whole point of pronunciation teaching is to enhance the communication experiences that L2 speakers have with their interlocutors. Nonetheless, how the people they interact with receive them is influenced by a complex set of social variables. Ample evidence suggests that L2 speakers are sometimes overtly discriminated against because of their accents (Lippi-Green 1997; Munro 2003; see also Chapter 7 by Munro, this volume), but there are more subtle negative reactions that interlocutors may not be consciously aware of themselves. Rubin (1992), for example, conducted a study in which he gave two groups of native English-speaking psychology students an audio-taped mini-lecture, produced by a NS of English who shared the students' Ohio dialect. One group saw a photograph of a Caucasian woman as they listened to the tape recording, while the other class saw a picture of a Chinese woman. The two photos were very similar in all respects other than racial background. Interestingly, when the students responded to comprehension questions about the mini-lecture material, the people who had seen the picture of the Chinese woman scored significantly worse than those who had seen the Caucasian. All of the pronunciation teaching in the world could not have helped in this situation, where listeners were attending more to visual stimuli than to what they heard. The students' assumption that the Chinese woman would be less clear influenced their comprehension, and in fact, some of the psychology students even complained about her "foreign" accent.

In another study of NS listeners, Derwing, Rossiter and Munro (2002) taught native speakers of English to listen to Vietnamese-accented speech. Many of the participants, students in a social work program, admitted at the end of the course that they had often avoided talking to L2 speakers, for fear that they would not understand them. They were not racist, but rather they were concerned that they did not have the ability to follow accented speech. Some admitted to "tuning out" as soon as they heard someone with an L2 accent start to talk. For these and other reasons, it is essential that there be discussions with ESL learners regarding the social consequences of having an L2 accent. (This may be less significant for EIL or EFL students.) It is of vital importance that students understand that the responsibility for the success of any interaction should be shared, and that sometimes it is the interlocutor who is at fault if communication breakdown takes place. Open discussions of attitudinal research, readings from books such as Lippi-Green's (1997) *English with an Accent*, clips from films where accent is at issue (e.g., the Korean grocery store scene in *Falling Down*), and discussions of strategies for dealing with difficult interlocutors are appropriate additions to pronunciation curricula, yet this aspect of pronunciation is not even acknowledged in most student-oriented material.

Another topic worthy of discussion with L2 speakers is the proliferation of accent reduction courses which make extravagant claims, and which often cost a great deal of money. Of course, the motto *caveat emptor* should be heeded with regard to any service or merchandise, but L2 students may not be aware of the degree of hucksterism in this area. In the most egregious cases, people who have no background in a related field set themselves up as experts. Consider the class in Toronto that is advertised as a course to "Canadianize" foreign accents (Stuparyk 1996). The instructor tells the students (all new immigrants) to place a marshmallow gently between the top and bottom teeth, thus holding the mouth open. They are then told to recite the tongue twister, *Peter Piper picked a peck of pickled peppers. How many peppers did Peter Piper pick?* in order improve their production of the phoneme /p/. As any student who has taken an introductory course in linguistics knows, /p/ requires closure of the lips, followed by a burst of air, something that is impossible with a marshmallow stuck between one's teeth, and yet this trusting class of students faithfully followed the directions for this and other equally useless exercises, with the expectation that their "expert" teacher knew what he was doing.

At the time of writing, a quick perusal of the web under the heading of accent reduction turned up 633,000 hits with sites such as (1) Accent Reduction Made Easy – Learn in your Car; (2) Foreign Accent Reduction Therapy; (3) Accent Reduction Speech and Stuttering Improvement; and (4) Accent Reduction, Physical Therapy, Job and Career Advice. Note that numbers 2–4 treat an accent as a pathology that requires therapy. The second entry offers its services in a Head and Neck Rehabilitation Center, which also treats laryngeal cancer and assesses tumor staging among other things (doing a course in one's car as suggested in the first entry might actually be more comfortable, if ineffective). These websites are suggesting that L2 learners are abnormal, and thus require the services of a speech therapist for remediation of their pathology. Discussion of these issues with adult L2 students will serve to heighten their awareness of the politics of accent, and may encourage them to think critically about their options.

Teacher preparation

Since an L2 accent is *not* a pathology but a completely normal characteristic of L2 speech, only those features that cause some intelligibility problems or issues that a student wants to work on should be of concern. The individuals who are best suited to provide appropriate instruction are thus second language teachers familiar with second language acquisition theory and practice. Unfortunately, many language teachers have little or no formal preparation to teach pronunciation. Murphy (1997) carried out a comprehensive survey of MATESOL programs in the USA. Although all of the programs had at least one course with a component involving phonology in some manner, far fewer programs had a course solely devoted to phonology issues. He makes several suggestions for ways in which university preparation programs could enhance the readiness of ESL teachers to work on pronunciation with their students, including arranging for a hands-on experience in an L2 classroom; developing case narratives from the perspectives of both L2 speakers and pronunciation teachers; helping teachers to assess technologies with a critical eye in order to make the most of what is available; relating pronunciation teaching to current research and practice in TESL generally; preparing teachers to determine their students' needs in authentic speaking situations; and teaching in "styles that enrich graduate education" (p. 756).

In a survey of 67 ESL programs across Canada, Breitkreutz et al. (2001) found that only 30% of teachers had any formal pronunciation teacher training (some of which consisted of a single workshop), and that the opportunities for ongoing professional development were very limited. Seventy-nine percent indicated that conference presentations were a primary source of upgrading in this area, while only 12% could access a university course. MacDonald (2002) has indicated that many Australian ESL teachers are quite uncomfortable with the idea of teaching pronunciation to their students, and Burgess and Spencer (2000) have also called for improved training of English language teachers in pronunciation in Britain. One consequence of a lack of preparation is that some teachers will rely heavily on published materials or software, without regard to their students' individual problems. Breitkreutz et al. (2001) determined that many ESL programs make extensive use of computer labs for pronunciation practice. When they examined the software packages most often used by the language programs surveyed, they found that all of them concentrated almost exclusively on segmentals, thus shortchanging those students whose primary problems are at the suprasegmental level.

Another consequence of insufficient teacher training is the provision of inaccurate information to L2 students. Wang and Munro (2004) indicate that many teachers tell their students that the main difference between a vowel pair such as /ɪ/ and /i/ is one of length. Just like dictionary categorizations of vowels into 'short' and 'long', teachers sometimes inform their students that the primary distinction between two words such as 'did' and 'deed' is purely durational in nature. In fact, vowel quality is a more important cue. Wang and Munro trained L2 listeners to pay attention to vowel quality rather than the secondary cue of length and found that "pedagogical misdirection" could be put right relatively easily.

Some pronunciation teachers are very well trained, while others, through intuition and strong observational skills, have developed their own successful strategies for teaching pronunciation; for many, however, there is insufficient support. In order to address the problems associated with a lack of teacher training (avoidance of pronunciation instruction, too heavy a reliance on set materials, a lack of understanding of what will help a student, or pedagogical misdirection), it is incumbent upon teacher preparation programs to include at least one course on pronunciation teaching, ideally one that would cover recent research findings, pronunciation curriculum development, and a hands-on component in which student teachers could implement some elements of what they are being taught.

Future directions

There are countless possibilities for future research in pronunciation teaching, but I will highlight some avenues that I believe should be explored sooner rather than later. First, there is a need for longitudinal studies that measure not only the effects of pronunciation instruction, but also the degree to which performance is permanently improved. Hahn's (2002) doctoral dissertation is a starting point, in that she measured L2 learners' productions of nine patterns of stress at three times – before instruction, immediately after the pronunciation course, and then at a later time – to determine whether the gains that had been made from Time 1 to Time 2 were retained at Time 3. She found first that the instruction resulted in more accurate productions at Time 2, and second that most of her participants' productions at Time 3 were better than the Time 1 samples. However, all but one person showed some backsliding across the board and, in the case of one third of the stress patterns, the students' productions returned to the accuracy levels at Time 1. Further innovative work of this type is needed to examine retention of other features. Also, in Hahn's study, Time 3 was not fixed; in future work the measures should be collected at constant intervals to permit comparisons among participants. Furthermore, it would be worthwhile to know whether periodic interventions would ensure better retention.

A second type of longitudinal study that would inform pronunciation instruction is large scale measures of developmental sequences. If clear patterns of L2 phonological development emerge in learners from many different L1 backgrounds, teachers would benefit from knowing what the stages of development are.

Third, factors affecting intelligibility should be explored to a greater extent. Studies that follow the model of Hahn's (2004) investigation of primary stress would be most useful in identifying the degree to which different linguistic variables affect intelligibility. Pickering's (2001) findings for tone choice, for example, could be extended by modifying tone in mini-lectures for a group of listeners, followed by comprehension questions and general impressions. Although there are several worthy candidates for specific investigation, it would also be valuable to study the interaction of a variety of modifications to establish whether certain combinations of L2 accent features are more problematic than others.

Fourth, although Catford (1987) and Brown's (1992) arguments for the importance of functional load are compelling, they have been examined in only a very preliminary study (Munro & Derwing 2006). Experiments that further test the hypothesis that low functional load errors interfere less with intelligibility than do high functional load errors should be carried out.

Fifth, given the interest in computer-assisted language learning, and the rapid development of new learner materials, a comprehensive examination of the effectiveness of prosodically-based software should be undertaken.

In addition to the possibilities listed above, it would be well worth pursuing more research that focuses on the mutual intelligibility of NNSs from a variety of L1 backgrounds, at different levels of proficiency. Finally, on a related matter, it would be most constructive to undertake a systematic series of tests of Jenkins' (2002) proposed core to determine whether the comprehensibility of the features she suggests and other pronunciation features identified in the literature actually hold across a large number of listeners.

Acknowledgements

I would like to thank my long-time friend and research partner, Murray Munro, for his ongoing contributions to the ideas in this chapter. Thanks also to Marian Rossiter and Ron Thomson for their comments. I am grateful to the Social Sciences and Humanities Research Council of Canada for financial support.

Suggested readings

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. 1996. *Teaching pronunciation: A reference for teachers of English to speakers of other languages.* Cambridge: Cambridge University Press.

Dalton, C., & Seidlhofer, B. 1994. *Pronunciation.* Oxford: Oxford University Press.

Derwing, T. M., & Munro, M. J. 2005. Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly* 39: 379–397.

Derwing, T. M., Munro, M. J., & Wiebe, G. E. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48: 393–410.

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

Levis, J. M., & Grant, L. 2003. Integrating pronunciation into ESL/EFL classrooms. *TESOL Journal* 12 (2): 13–19.

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

Rubin, D. L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English speaking teaching assistants. *Research in Higher Education* 33: 511–531.

References

Anderson-Hsieh, J., Johnson, R., & Koehler, K. 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning* 42: 529–555.

Best, C. T. 1995. A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception in linguistic experience: Theoretical and methodological issues*, Timonium, MD: York Press, 171–206.

Bongaerts, T., Mennen, S., & van der Slik, F. 2000. Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica* 54: 298–308.

Bongaerts, T., van Summerin, C., Planken, B., & Schils, E. 1997. Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition* 19: 447–465.

Brazil, D. 1997. *The communicative value of intonation in English.* Cambridge: Cambridge University Press.

Breitkreutz, J. A., Derwing, T. M., & Rossiter, M. J. 2001. Pronunciation teaching practices in Canada. *TESL Canada Journal* 19: 51–61.

Brown, A. 1991. Functional load and the teaching of pronunciation. In A. Brown (ed.), *Teaching English pronunciation: A book of readings.* London: Routledge.

Burda, A. N., Scherz, J. A., Hageman, C. F., & Edwards, H. T. 2003. Age and understanding of speakers with Spanish or Taiwanese accents. *Perceptual and Motor Skills* 97: 11–20.

Burgess, J., & Spencer, S. 2000. Phonology and pronunciation in integrated language teaching and teacher education. *System* 28: 191–215.

Catford, J. C. 1987. Phonetics and the teaching of pronunciation: A systemic description of the teaching of English phonology. In J. Morley (ed.), *Current perspectives on pronunciation: Practices anchored in theory*, Washington DC: TESOL, 83–100.

Cauldwell, R. 2002. *Streaming speech: Listening and pronunciation for advanced learners of English.* Birmingham, UK: speechinaction.

Cauldwell, R., & Hewings, M. 1996. Intonation rules in ELT textbooks. *ELT Journal* 50: 327–334.

Celce-Murcia, M., Brinton, D., & Goodwin, J. 1996. *Teaching pronunciation: A reference for teachers of English to speakers of other languages.* Oxford: Oxford University Press.

Chela-Flores, B. 2001. Pronunciation and language learning: An integrative approach. *International Review of Applied Linguistics in Language Teaching* 39: 85–101.

Chun, D. M. 1988. The neglected role of intonation in communicative competence and proficiency. *Modern Language Journal* 72: 295–303.

Chun, D. M. 2002. *Discourse intonation in L2: From theory and research to practice.* Amsterdam: John Benjamins.

Chun, D. M., Hardison, D. M., & Pennington, M. C. 2008. Technologies for prosody in context: Past and future of L2 research and practice. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Clennell, C. 1996. Promoting the role of English prosody in a discourse based approach to oral interaction. *Prospect* 11: 17–28.

Couper, G. 2003. The value of an explicit pronunciation syllabus in ESOL teaching. *Prospect* 18, (3): 53–70.

Dalton, C., & Seilhofer, B. 1994. *Pronunciation*. Oxford: Oxford University Press.

Davies, P., & Pearse, E. 2000. *Success in English language teaching.* Oxford: Oxford University Press.

Derwing, T. M. 2003. What do ESL students say about their accents? *Canadian Modern Language Review* 59: 547–566.

Derwing, T. M., & Munro, M. J. 2005. Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly* 39: 379–397.

Derwing, T. M., Munro, M. J., & Wiebe, G. E. 1997. Pronunciation instruction for "fossilized" learners: Can it help? *Applied Language Learning* 8: 217–235.

Derwing, T. M., Munro, M. J., & Wiebe, G. E. 1998. Evidence in favour of a broad framework for pronunciation instruction. *Language Learning* 48: 393–410.

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

Derwing, T. M., Rossiter, M. J., & Munro, M. J. 2002. Teaching native speakers to listen to foreign-accented speech. *Journal of Multilingual and Multicultural Development* 23: 245–259.

Derwing, T. M., & Thomson, R. I. 2004, May. Effects of L1 fluency on L2 fluency development. Paper presented at the meeting of the American Association of Applied Linguistics, Portland, OR.

Elliott, A. R. 1997. On the teaching and acquisition of pronunciation within a communicative approach. *Hispania* 80: 95–108.

Esling, J., & Wong, R. F. 1983. Voice quality settings and the teaching of pronunciation. *TESOL Quarterly* 17: 89–95.

Firth, S. 1992. Pronunciation syllabus design: A question of focus. In P. Avery & S. Ehrlich (eds.), *Teaching American English pronunciation*, Oxford: Oxford University Press, 173–183.

Flege, J. E. 1995. Second language speech learning: Theory, findings and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues*, Timonium, MD: York Press, 229–273.

Flege, J. E., Bohn, O., & Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25: 437–470.

Flege, J. E., Frieda, E. M., & Nozawa, T. 1997. Amount of native language (L1) use affects the pronunciation of an L2. *Journal of Phonetics* 25: 169–186.

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

Gass, S., & Varonis, E. M. 1984. The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning* 34: 65–89.

Gatbonton, E., Trofimovich, P., & Magid, M. 2005. Learners' ethnic group affiliation and L2 pronunciation accuracy: A sociolinguistic investigation. *TESOL Quarterly* 39: 489–511.

Gilbert, J. 1984. *Clear speech: Pronunciation and listening comprehension in American English.* New York: Cambridge University Press.

Gilbert, J. 2005. *Clear speech: Pronunciation and listening comprehension in American English, (3rd edition).* New York: Cambridge University Press.

Gorsuch, G. J. 2001. Testing textbook theories and tests: The case of suprasegmentals in a pronunciation textbook. *System* 29: 119–136.

Graham, C. 1978. *Jazz chants: Rhythms of American English for students of English as a second language.* New York: Oxford University Press.

Grant, L. 1993. *Well said: Advanced English pronunciation.* Boston, MA: Heinle & Heinle.

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

Hahn, L. D., & Dickerson, W. B. 1999. *Speechcraft: Discourse pronunciation for advanced learners.* Ann Arbor, MI: University of Michigan Press.

Hahn, M. K. 2002. *The persistence of learned primary phrase stress patterns among learners of English.* Unpublished PhD Dissertation, University of Illinois at Urbana-Champaign.

Hansen Edwards, J. G. 2008. Social factors and variation in production. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Hedge, T. 2000. *Teaching and learning in the language classroom.* Oxford: Oxford University Press.

Hewings, M. & Goldstein, S. 1999. *Pronunciation plus: Practice through interaction.* Cambridge: Cambridge University Press.

Ioup, G. 2008. Exploring the role of age in the acquisition of a second language phonology. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Ioup, G., Boustagi, E., El Tigi, M., & Moselle, M. 1994. Re-examining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition* 16: 73–98.

Jenkins, J. 2002. A sociolinguistically-based, empirically-researched pronunciation syllabus for English as an International Language. *Applied Linguistics* 23: 83–103.

Jones, R. H., & Evans, S. 1995. Teaching pronunciation through voice quality. *ELT Journal* 49: 244–251.

Kerr, J. 2000. Articulatory setting and voice production: Issues in accent modification. *Prospect* 15 (2): 4–15.

Levis, J. M. 1999. Intonation in theory and practice revisited. *TESOL Quarterly* 33: 37–54.

Levis, J. M., & Grant, L. 2003. Integrating pronunciation into ESL/EFL classrooms. *TESOL Journal* 12 (2): 13–19.

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

Lippi-Green, R. 1997. *English with an accent: Language, ideology and discrimination in the United States.* New York: Routledge.

MacDonald, S. 2002. Pronunciation: Views and practices of reluctant teachers. *Prospect* 17 (3): 3–18.

Major, R. C., Fitzmaurice, S. E., Bunta, F., & Balasubramanian, C. 2002. The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly* 36: 173–190.

Matthews, C. 1994. *Speaking solutions: Interaction, presentation, listening and pronunciation skills.* Englewood Cliffs, NJ: Prentice Hall Regents.

Moyer, A. 1999. Ultimate attainment in L2 phonology: The critical factors of age, motivation and instruction. *Studies in Second Language Acquisition* 21: 81–108.

Munro, M. J. 1993. Productions of English vowels by native speakers of Arabic: Acoustical measurements and accentedness ratings. *Language and Speech* 36: 39–66.

Munro, M. J. 2003. A primer on accent discrimination in the Canadian context. *TESL Canada Journal* 20 (2): 38–51.

Munro, M. J. 2008. Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

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

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

Munro, M. J., & Derwing, T. M. 2006. The functional load principle in ESL pronunciation instruction: An exploratory study. *System* 34: 520–531.

Munro, M. J., Derwing, T. M., & Burgess, C. S. 2003. The detection of foreign accent in backwards speech. In M.-J. Solé, D. Recasens & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, ES: Universitat Autonoma de Barcelona, 535–538.

Munro, M. J., Derwing, T. M., & Morton, S. 2006. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition* 28: 111–131.

Murphy, J. M. 1991. Oral communication in TESOL: Integrating speaking, listening and pronunciation. *TESOL Quarterly* 25: 51–75.

Murphy, J. M. 1997. Phonology courses offered by MATESOL programs in the US. *TESOL Quarterly* 31: 741–761.

Murray, L., & Barnes, A. 1998. Beyond the "wow" factor – evaluating multimedia language learning software from a pedagogical viewpoint. *System* 26: 249–259.

Nilsen, D. L. F., & Nilsen, A. P. 1971. *Pronunciation contrasts in English.* New York: Regents.

Nunan, D. 1999. *Second language teaching and learning.* Boston: Heinle & Heinle.

Orion, G. F. 1997. *Pronouncing American English: Sounds, stress and intonation.* Pacific Grove, CA: Heinle & Heinle.

Pennington, M. C. 1989. Teaching pronunciation from the top down. *RELC Journal* 20: 20–38.

Pickering, L. 2001. The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly* 35: 233–255.

Piller, I. 2002. Passing for a native speaker: Identity and success in second language learning. *Journal of Sociolinguistics* 6: 179–206.

Purcell, E. T., & Suter, R. W. 1980. Predictors of pronunciation accuracy: A reexamination. *Language Learning* 30: 271–287.

Reed, M., & Michaud, C. 2005. *Sound concepts: An integrated pronunciation course.* McGraw Hill.

Rubin, D. L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English speaking teaching assistants. *Research in Higher Education* 33: 511–531.

Schmid, P. M., & Yeni-Komshian, G. H. 1999. The effects of speaker accent and target predictability on perception of mispronunciations. *Journal of Speech, Language and Hearing Research* 42: 56–64.

Schumacher, J. (Director). 1993. *Falling down.* Hollywood: Andrew Kopelson Productions in association with Warner Brothers.

Seidlhofer, B., & Dalton-Puffer, C. 1995. Appropriate units in pronunciation teaching: Some programmatic pointers. *International Journal of Applied Linguistics* 5: 13–146.

Strange, W., & Shafer, V. L. 2008. Speech perception in second language learners. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, Philadelphia: John Benjamins.

Stuparyk, M. 1996. New Canadians eager to master English. *The Edmonton Journal*, July 14, G6.

Swan, M., & Smith, B. (eds.) 1987. *Learner English: A teacher's guide to interference and other problems.* Cambridge: Cambridge University Press.

Timmis, I. 2002. Native-speaker norms and International English: A classroom view. *ELT Journal* 56: 240–249.

Tyler, A. 1992. Discourse structure and the perception of incoherence in international teaching assistants' spoken discourse. *TESOL Quarterly* 26: 713–729.

Varonis, E. M., & Gass, S. 1982. The comprehensibility of non-native speech. *Studies in Second Language Acquisition* 4: 114–136.

Wang, X., & Munro, M. J. 2004. Computer-based training for learning English vowel contrasts. *System* 32: 539–552.

Wennerstrom, A. 1994. Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics* 15: 399–420.

Wennerstrom, A. 1998. Intonation as cohesion in academic discourse. Studies in *Second Language Acquisition* 20: 1–25.

Westwood, V. W., & Kaufman, H. 2005. *Connected speech (Rev. ed.)* Hurstbridge, Australia: Protea Textware.

Zielinski, B. 2004. May. Measurement of speech intelligibility: What are we actually measuring? Paper presented at the meeting of the American Association of Applied Linguistics, Portland, OR.

Author index

Subject index

In the series *Studies in Bilingualism* (SiBil) the following titles have been published thus far or are scheduled for publication:

37 **NIÑO-MURCIA, Mercedes and Jason ROTHMAN (eds.):** Bilingualism and Identity. Spanish at the crossroads with other languages. ix, 357 pp. + index. *Expected March 2008*

36 **HANSEN EDWARDS, Jette G. and Mary L. ZAMPINI (eds.):** Phonology and Second Language Acquisition. 2008. vi, 380 pp.

35 **ROCCA, Sonia:** Child Second Language Acquisition. A bi-directional study of English and Italian tense-aspect morphology. 2007. xvi, 240 pp.

34 **KOVEN, Michèle:** Selves in Two Languages. Bilinguals' verbal enactments of identity in French and Portuguese. 2007. xi, 327 pp.

33 **KÖPKE, Barbara, Monika S. SCHMID, Merel KEIJZER and Susan DOSTERT (eds.):** Language Attrition. Theoretical perspectives. 2007. viii, 258 pp.

32 **KONDO-BROWN, Kimi (ed.):** Heritage Language Development. Focus on East Asian Immigrants. 2006. x, 282 pp.

31 **BAPTISTA, Barbara O. and Michael Alan WATKINS (eds.):** English with a Latin Beat. Studies in Portuguese/Spanish – English Interphonology. 2006. vi, 214 pp.

30 **PIENEMANN, Manfred (ed.):** Cross-Linguistic Aspects of Processability Theory. 2005. xiv, 303 pp.

29 **AYOUN, Dalila and M. Rafael SALABERRY (eds.):** Tense and Aspect in Romance Languages. Theoretical and applied perspectives. 2005. x, 318 pp.

28 **SCHMID, Monika S., Barbara KÖPKE, Merel KEIJZER and Lina WEILEMAR (eds.):** First Language Attrition. Interdisciplinary perspectives on methodological issues. 2004. x, 378 pp.

27 **CALLAHAN, Laura:** Spanish/English Codeswitching in a Written Corpus. 2004. viii, 183 pp.

26 **DIMROTH, Christine and Marianne STARREN (eds.):** Information Structure and the Dynamics of Language Acquisition. 2003. vi, 361 pp.

25 **PILLER, Ingrid:** Bilingual Couples Talk. The discursive construction of hybridity. 2002. xii, 315 pp.

24 **SCHMID, Monika S.:** First Language Attrition, Use and Maintenance. The case of German Jews in anglophone countries. 2002. xiv, 259 pp. (incl. CD-rom).

23 **VERHOEVEN, Ludo and Sven STRÖMQVIST (eds.):** Narrative Development in a Multilingual Context. 2001. viii, 431 pp.

22 **SALABERRY, M. Rafael:** The Development of Past Tense Morphology in L2 Spanish. 2001. xii, 211 pp.

21 **DÖPKE, Susanne (ed.):** Cross-Linguistic Structures in Simultaneous Bilingualism. 2001. x, 258 pp.

20 **POULISSE, Nanda:** Slips of the Tongue. Speech errors in first and second language production. 1999. xvi, 257 pp.

19 **AMARA, Muhammad Hasan:** Politics and Sociolinguistic Reflexes. Palestinian border villages. 1999. xx, 261 pp.

18 **PARADIS, Michel:** A Neurolinguistic Theory of Bilingualism. 2004. viii, 299 pp.

17 **ELLIS, Rod:** Learning a Second Language through Interaction. 1999. x, 285 pp.

16 **HUEBNER, Thom and Kathryn A. DAVIS (eds.):** Sociopolitical Perspectives on Language Policy and Planning in the USA. With the assistance of Joseph Lo Bianco. 1999. xvi, 365 pp.

15 **PIENEMANN, Manfred:** Language Processing and Second Language Development. Processability theory. 1998. xviii, 367 pp.

14 **YOUNG, Richard and Agnes Weiyun HE (eds.):** Talking and Testing. Discourse approaches to the assessment of oral proficiency. 1998. x, 395 pp.

13 **HOLLOWAY, Charles E.:** Dialect Death. The case of Brule Spanish. 1997. x, 220 pp.

12 **HALMARI, Helena:** Government and Codeswitching. Explaining American Finnish. 1997. xvi, 276 pp.

11 **BECKER, Angelika and Mary CARROLL:** The Acquisition of Spatial Relations in a Second Language. In cooperation with Jorge Giacobbe, Clive Perdue and Rémi Porquiez. 1997. xii, 212 pp.

10 **BAYLEY, Robert and Dennis R. PRESTON (eds.):** Second Language Acquisition and Linguistic Variation. 1996. xix, 317 pp.

9 **FREED, Barbara F. (ed.):** Second Language Acquisition in a Study Abroad Context. 1995. xiv, 345 pp.

8 **DAVIS, Kathryn A.:** Language Planning in Multilingual Contexts. Policies, communities, and schools in Luxembourg. 1994. xix, 220 pp.

7 **DIETRICH, Rainer, Wolfgang KLEIN and Colette NOYAU:** The Acquisition of Temporality in a Second Language. In cooperation with Josée Coenen, Beatriz Dorriots, Korrie van Helvert, Henriette Hendriks, Et-Tayeb Houdaïfa, Clive Perdue, Sören Sjöström, Marie-Thérèse Vasseur and Kaarlo Voionmaa. 1995. xii, 288 pp.

6 **SCHREUDER, Robert and Bert WELTENS (eds.):** The Bilingual Lexicon. 1993. viii, 307 pp.

5 **KLEIN, Wolfgang and Clive PERDUE:** Utterance Structure. Developing grammars again. In cooperation with Mary Carroll, Josée Coenen, José Deulofeu, Thom Huebner and Anne Trévise. 1992. xvi, 354 pp.

4 **PAULSTON, Christina Bratt:** Linguistic Minorities in Multilingual Settings. Implications for language policies. 1994. xi, 136 pp.

3 **DÖPKE, Susanne:** One Parent – One Language. An interactional approach. 1992. xviii, 213 pp.

2 **BOT, Kees de, Ralph B. GINSBERG and Claire KRAMSCH (eds.):** Foreign Language Research in Cross-Cultural Perspective. 1991. xii, 275 pp.

1 **FASE, Willem, Koen JASPAERT and Sjaak KROON (eds.):** Maintenance and Loss of Minority Languages. 1992. xii, 403 pp.