

The Pennsylvania State University
The Graduate School
Department of Spanish, Italian, and Portuguese

**THE ACQUISITION OF PROBABILISTIC PATTERNS IN SPANISH
PHONOLOGY BY ADULT SECOND LANGUAGE LEARNERS:
THE CASE OF DIPHTHONGIZATION**

A Thesis in

Spanish

by

Matthew Thomas Carlson

© 2007 Matthew Thomas Carlson

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2007

The thesis of Matthew Thomas Carlson was reviewed and approved* by the following:

Henry Gerfen
Associate Professor in Spanish Linguistics
Thesis Advisor
Chair of Committee
Interim Associate Head of the Department of Spanish, Italian, and
Portuguese

Paola E. Giuli Dussias
Assistant Professor of Spanish and Second Language Acquisition

Judith Kroll
Liberal Arts Research Professor of Psychology and Linguistics

Daniel Weiss
Assistant Professor of Psychology

*Signatures are on file in the Graduate School

ABSTRACT

The increasingly global nature of our society bears witness more and more to the importance of learning additional languages. Human language is, however, an extremely complex phenomenon and decades of research have raised at least as many questions as they have answered. In particular, the learning of additional languages has fueled a wealth of research, especially regarding the cognitive underpinnings of adult second language acquisition. This dissertation seeks to shed light on the question of adult language acquisition by exploring learners' knowledge of fine-grained statistical patterns in Spanish morphophonology. Crucially, it adds to this an exploration of the role of certain kinds of memory and attentional control in modulating adults' sensitivity to these subtle patterns in their second language grammar.

Two areas of recent research are especially relevant to the approach taken here. First, recent findings point to the ability of both adults and children to segment words and extrapolate grammatical rules from statistical patterns in (generally artificial) language input. Second, a large body of psycholinguistic literature shows that adults are sensitive to frequency at a variety of levels of structure, including the positional probability of segments and syllables in words. These findings suggest that at least some grammatical information is resonant in individuals' experience of language, and that both adults and children have access to statistical learning mechanisms that allow them to capitalize on frequency information in their first language. Findings such as these have led to the development of emergentist and usage-based theories of language that address phenomena from child language acquisition to historical linguistics, and which develop a

view of grammar as instantiated in statistical patterns in actual language, dynamically evolving in response to subtle changes in the frequency of particular structures.

Importantly, such a view of language structure, learning, and representation suggests both a continuity in the mechanisms of language acquisition throughout the lifespan, and also specific reasons for the obvious differences in the course and outcome of language acquisition in children and adults. This opens the door to investigating the contribution of these mechanisms to adult second language learning, and to exploring the differences between child and adult second language acquisition based on the cognitive structures that underlie and modulate frequency-based learning processes.

In line with these findings, the present research investigates the sensitivity of adult second language learners to a subtle probabilistic subpattern of Spanish diphthongization in derived words, and compares their performance with that of native Spanish speakers. The Spanish diphthong/mid-vowel alternation may be explained through a highly consistent rule in Spanish verbs, but in suffixed forms such as those examined here, the probability that a given form will conform to the consistent rule varies according to the suffix used. The present experiments measure learners' processing of neologisms created using real Spanish stems with alternating diphthongs and real derivational suffixes as a way of assessing their sensitivity to the probability of a diphthongized stem in specific morphological contexts.

The consequences of this probabilistic pattern in diphthongization are measured in participants' perception and production on a lexical decision task and a conditional naming task in which participants say *palabra* 'word' if they hear a real Spanish word, and repeat the nonword items, including the neologisms. Learners are also compared

based on Spanish proficiency and on three measures of cognitive capacity accuracy in repeating words from an unfamiliar foreign language (Korean), inhibitory control, as measured on the Simon task, and working memory, based on reading span scores.

Native speakers performing the same tasks are shown to be sensitive to this subtle probabilistic variability in their processing of neologisms, although the results vary in strength between the lexical decision and conditional naming tasks. Strikingly, learners also appear to be sensitive to this variability, despite the complex morpholexical and paradigmatic knowledge required to detect these patterns. However, learner behavior differs from that of native speakers in interesting ways. The lexical biases of some suffixes appear to be more easily acquired than others, and learner performance differs more dramatically between reaction time and error rate measures, and between lexical decision and conditional naming. This suggests that the morphophonological pattern in question may not affect all levels of lexical processing, and that its effects are also contingent to some extent on task demands. Finally, there is also evidence that phonological memory, and to a lesser extent working memory also modulate learners' sensitivity to probabilistic variation in Spanish diphthongization, but these effects also vary from one dependent measure to another.

The results have important implications for theories of adult second language acquisition, suggesting that adults are sensitive to similar kinds of statistical information in language experience to that exploited in emergentist theories of first language acquisition. This indicates that adult language acquisition may rely on many of the same cognitive processes as does child language acquisition, although adults' more developed cognitive resources also lead to important differences.

TABLE OF CONTENTS

LIST OF FIGURES	ix
LIST OF TABLES	xii
LIST OF ABBREVIATIONS.....	xiii
ACKNOWLEDGEMENTS.....	xiv
Chapter 1 The role of statistical information in language structure and processing... 1	
1.1 Overview.....	1
1.1.1 The cognitive bases of language acquisition	3
1.1.2 Implications of emergentism for second language acquisition	5
1.1.3 Testing for emergent grammar in adult second language acquisition..	8
1.2 Probability in language structure: Computational and descriptive evidence..	12
1.3 The impact of lexical statistics on human language processing and use	19
1.3.1 Probabilistic effects in language usage and language change	20
1.3.2 Frequency effects in phonology and morphology	24
1.3.3 Probabilistic phonotactics in language processing	36
1.4 Human learning based on statistical patterns.....	44
1.4.1 Evidence for emergent language knowledge in a second language	51
1.5 Cognitive mechanisms in first and second language acquisition	56
1.5.1 Usage-based theory and Cognitive Grammar.....	57
1.5.2 Working memory, phonological memory, and attentional control	62
1.6 Summary.....	65
Chapter 2 The Diphthong/Mid-Vowel Alternation in Spanish Derived Words	68
2.1 Chapter overview.....	68
2.2 The Spanish diphthong/mid-vowel alternation.....	70
2.2.1 Analyses of diphthongization in generative phonology	72
2.2.2 Descriptive and experimental accounts of gradience in Spanish diphthongization.....	76
2.2.3 Spanish diphthongization in adult L2 learners: forced choice data.....	84
2.3 The present experiments: Research questions and predictions.....	86
Chapter 3 Methodology	91
3.1 Overview.....	91
3.2 Tests of phonological processing.....	93
3.2.1 Perception experiment.....	95
3.2.2 Production experiment.....	97
3.2.3 Item construction and selection, LDT and Conditional Naming.....	98

3.2.3.1 Test neologisms.....	100
3.2.3.2 Distractors	105
3.2.3.3 Recording of auditory stimuli	106
3.2.3.4 Procedures for LDT and conditional naming tasks.....	108
3.2.4 Proficiency measures.....	109
3.2.5 Cognitive resources	111
3.2.5.1 Foreign word repetition.....	112
3.2.5.2 Simon task.....	114
3.2.5.3 Reading span task.....	116
3.3 Participants	117
3.4 Procedure	118
3.5 Data coding and analysis	119
Chapter 4 Results: Native Speakers.....	123
4.1 Chapter overview.....	123
4.2 Results: Lexical decision task.....	126
4.2.1 False alarm rates: LDT	126
4.2.2 Reaction time analyses: LDT	128
4.3 Results: Conditional naming task	131
4.3.1 False alarm rates: Conditional naming.....	131
4.3.1.1 Reversals of diphthong status: Conditional naming.....	133
4.3.2 Reaction time analyses: Conditional naming	134
4.4 Discussion.....	136
4.4.1 Effects of wordlikeness on lexical processing: Facilitation or inhibition?.....	138
4.4.2 Gradient patterns in corpus data, forced choice, false alarms, and reaction times	150
4.5 Conclusions.....	158
Chapter 5 Results: Adult Learners.....	161
5.1 Chapter overview.....	161
5.2 Results for adult learners: Lexical decision task	165
5.2.1 Diphthongizing stems in neologisms with derivational suffixes.....	165
5.2.2 Individual differences between learners.....	168
5.2.2.1 Considerations for data presentation.....	168
5.2.2.2 Effects of Spanish proficiency and experience.....	171
5.2.2.3 Individual differences in cognitive resources.....	174
5.3 Results for adult learners: Conditional naming task.....	177
5.3.1 Diphthongizing stems in neologisms with derivational suffixes.....	177
5.3.2 Individual differences between learners.....	181
5.3.2.1 Effects of Spanish proficiency and experience	182
5.3.2.2 Individual differences in cognitive resources.....	188
5.4 Discussion.....	191

5.4.1 The acquisition of gradience in diphthongization	191
5.4.2 The role of proficiency and language learning experiences	203
5.4.3 The role of cognitive resources	206
5.5 Conclusions.....	211
Chapter 6 Conclusions: Emergence, SLA, and Multilingualism.....	214
6.1 Overview.....	214
6.2 The present findings	216
6.2.1 Diphthongs, monophthongs, and the processing of neologisms	216
6.2.2 Experiential and cognitive factors	225
6.3 Emergence, age, and language acquisition	227
6.4 Emergent grammar in adult SLA.....	234
6.4.1 Estimating the grammatical information available for acquisition: Corpus research	234
6.4.2 Learner interaction with L2 experience.....	237
6.5 Emergent grammar in multilingualism.....	241
6.6 Conclusion: Adult L2 learners as a testing ground for emergentist theory	243
References.....	246
Appendix A Test items for the lexical decision and conditional naming tasks	
Table A-1	260
Appendix B Picture Naming Items	270
Appendix C Korean word repetition task items.....	272
Appendix D Language History Questionnaire Items.....	274
Appendix E Reaction Time Figures in Milliseconds.....	277

LIST OF FIGURES

Figure 2-1: Mean correctness ratings for neologisms with unstressed diphthongs or monophthongs, by suffix (Carlson, 2006).....	85
Figure 4-1: Native speaker selection of neologisms with and without diphthongs from Eddington (1998) for the suffixes in the present study.....	125
Figure 4-2: False alarm rates for neologisms in LDT by native speakers	127
Figure 4-3: Log RTs for neologisms with multisyllabic stems in LDT by native speakers, analysis by items	129
Figure 4-4: False alarm rates for neologisms in conditional naming by native speakers, analysis by subjects.....	132
Figure 4-5: Log RT data for neologisms in conditional naming by native speakers, analysis by items.....	135
Figure 5-1: logRTs for neologisms in LDT by learners	166
Figure 5-2: Monophthong bias in Eddington (1998) for native speakers on a forced choice task	169
Figure 5-3: Monophthong bias for native speakers on the LDT, all neologisms	170
Figure 5-4: False alarm rates by diphthong status in the LDT by learners, by Spanish picture naming accuracy	172
Figure 5-5: Monophthong bias (logRT) by suffix on the LDT for learners with and without study abroad experience	173
Figure 5-6: Monophthong bias (logRT) by suffix on the LDT for learners with greater accentedness in Korean word repetition.....	174
Figure 5-7: Monophthong bias (logRT) by suffix on the LDT for learners with less accentedness in Korean word repetition.....	175
Figure 5-8: Monophthong bias (logRT) by suffix on the LDT for low and high memory span groups.....	176
Figure 5-9: Mean logRT and false alarm rates for neologisms in conditional naming by adult learners, by diphthong status	179
Figure 5-10: False alarm rates for neologisms in conditional naming by adult learners.....	180

Figure 5-11 : logRTs by diphthong status for neologisms in conditional naming by adult learners, by Spanish picture naming accuracy and efficiency	183
Figure 5-12 : Diphthong bias (logRT) by suffix for learners on the conditional naming task with low and high accuracy in Spanish picture naming.....	185
Figure 5-13 : Diphthong bias (logRT) by suffix for learners with low and high efficiency in Spanish picture naming	185
Figure 5-14 : Monophthong bias (false alarms) for neologisms in conditional naming by adult learners with more or less efficient Spanish picture naming scores.	187
Figure 5-15 : Diphthong preference (logRT) in conditional naming by adult learners as a function of accuracy in Korean word repetition	189
Figure 5-16 : Diphthong bias (logRT) for neologisms in conditional naming for adult learners varying in attentional control ability	190
Figure E-1 : RTs in milliseconds for neologisms with multisyllabic stems in LDT by native speakers, analysis by items (Figure 4-3).....	277
Figure E-2 : RTs in milliseconds for neologisms in conditional naming by native speakers, analysis by items (Figure 4-5).....	278
Figure E-3 : RTs in milliseconds for neologisms in LDT by learners (Figure 5-1)	278
Figure E-4 : Monophthong bias in (RT in milliseconds) for native speakers on the LDT, all neologisms (Figure 5-3).....	279
Figure E-5 : Monophthong bias (RT in milliseconds) by suffix on the LDT for learners with and without study abroad experience (Figure 5-5)	279
Figure E-6 : Monophthong bias (RT in milliseconds) by suffix on the LDT for learners with greater accentedness in Korean word repetition (Figure 5-6).....	280
Figure E-7 : Monophthong bias (RT in milliseconds) by suffix on the LDT for learners with less accentedness in Korean word repetition (Figure 5-7).....	280
Figure E-8 : Monophthong bias (RT in milliseconds) by suffix on the LDT for low and high memory span groups (Figure 5-8)	281
Figure E-9 : RTs in milliseconds by diphthong status for neologisms in conditional naming by adult learners, by Spanish picture naming accuracy and efficiency (Figure 5-11).....	281

Figure E-10 : Diphthong bias (RT in milliseconds) by suffix for learners on the conditional naming task with low and high accuracy in Spanish picture naming (Figure 5-12).....	282
Figure E-11 : Diphthong bias (RT in milliseconds) by suffix for learners with low and high efficiency in Spanish picture naming (Figure 5-13).....	282
Figure E-12 : Diphthong preference (RT in milliseconds) in conditional naming by adult learners as a function of accuracy in Korean word repetition (Figure 5-15).....	283
Figure E-13 : Diphthong bias (RT in milliseconds) for neologisms in conditional naming for adult learners varying in attentional control ability (Figure 5-16).....	283

LIST OF TABLES

Table 2-1: Examples of diphthongizing verbs in Spanish	70
Table 2-2: Derived words with diphthongizing stems	71
Table 2-3: Percentage of attested derivations with monophthongs and percentage of monophthong forms of neologisms selected by native speakers(Eddington, 1996, 1998). Suffixes are ranked according to Eddington (1998).....	80
Table 3-1: Experimental Tasks	91
Table 3-2: Stimulus counts for LDT and Conditional Naming tasks	99
Table 4-1: Reversals of diphthong status in conditional naming by native speakers ..	134
Table 5-1: Reversals of diphthong status in conditional naming by adult learners	181

LIST OF ABBREVIATIONS

ADIOS	Automatic DIstillation Of Structure
AML	Analogical Modeling of Language
ANOVA	Analysis of Variance
ANCOVA	Analysis of Covariance
ERP	Event-Related Potential
L1	First language
L2	Second language
LDT	Lexical Decision Task
LHQ	Language History Questionnaire
MMN	Mismatch Negativity
RT(s)	Reaction time(s)
SLA	Second Language Acquisition
VOT	Voice Onset Time

ACKNOWLEDGEMENTS

I owe a tremendous debt of gratitude to the many individuals who have guided, taught, trained, and encouraged me over the past years. I first wish to express thanks to my dissertation advisor, Chip Gerfen, and the members of my doctoral committee, Judith Kroll, Giuli Dussias, and Daniel Weiss. They were from the start actively involved, constantly available, and provided invaluable questions, suggestions, and good sound advice. Alongside my committee, I am also indebted to Drs. Scott Payne and James Lantolf, who invested a great deal in me and in the work leading to this dissertation, and to the members of the Center for Language Science at Penn State, who listened and offered much help and encouragement at many stages in the evolution of this project. My fellow graduate students in the Center are deserving of special mention and thanks, especially Susan Bobb, Noriko Hoshino, Jared Linck, and Mike Shelton. I have been so fortunate to work in such a community of intellectuals, with a great openness to talking about new ideas and supporting each other's research, and to sharing a tremendous depth of experience and knowledge. In addition to enabling me to do so much more than I could have done alone, they have also taught me how to be a true colleague, a lesson I will not forget.

In addition to much guidance and preparation, the experiments reported here demanded considerable resources. I am grateful to Dr. Maya Misra for providing sound booths for recording stimuli, and to Drs. Teresa Bajo, and Pedro Macizo for providing lab space for running the experiments and for their gracious hospitality during data collection in Spain. Special thanks are also due to Dr. Nicolás Gutiérrez-Palma for his assistance in

norming the experimental stimuli. I would like to thank Eduardo Gregori, Jesús Rueda López, Susana García Prudencio, and Timothy Marston for their help in preparing and recording the Spanish stimuli, Joel Koeth, Haewon Byun, Kyungja Ahn, So-Eun Cho, and Soojung Choi for their invaluable help in selecting, recording, and coding the Korean experiment and data, Ruth Hummel for help with the statistics, and the many fellow students in the Spanish department who helped me test and pilot the experiments.

Research Assistants Sherlyne Dalupang and Gabriela Salerno are deserving of particularly hearty thanks, not only for their many hours of work running experiments and coding data, but also for their interest in the entire endeavor and their readiness to learn. I am glad to have had the chance to work with these growing scholars, and wish them all the best in the future.

Above all I thank my wife, Joanna. She stood with me, encouraged me, and kept me going through the inevitable challenges, setbacks, and at times just plain fatigue. Her love and support made my work lighter by far.

This research was funded by National Science Foundation Dissertation Improvement Grant BCS-0544969: *The acquisition of probabilistic patterns in Spanish phonology by adult foreign language learners*, under the direction of Dr. Chip Gerfen, dissertation advisor

Chapter 1

The role of statistical information in language structure and processing

1.1 Overview

The past two decades have seen a dramatic increase in research on how individuals acquire and represent statistical information in their first language (L1) grammars. This work has yielded considerable progress in determining the extent to which grammatical information is available from statistical distributions across individuals' experience of their L1 and in exploring the cognitive mechanisms by which speakers may be able to exploit this information to learn, represent, and use languages. Despite this robust and growing research tradition revealing speakers' remarkable sensitivity to the statistical structure of their L1 grammars, surprisingly little is known about how or whether speakers of a second language (L2) acquire similarly fine-grained knowledge of their L2 grammar. This dissertation begins to explore the implications of a statistical approach to grammar for L2 research by examining learners' knowledge of probabilistic structures in Spanish phonology at different levels of proficiency and by evaluating the contribution of specific cognitive resources to that knowledge.

Within the L1 domain, researchers have increasingly explored probabilistic phenomena in grammar and language processing at a variety of levels, from phonology to lexical processing to syntax and semantics (e.g. Bod, Hay, & Jannedy, 2003; Bybee & Hopper, 2001b). Findings from this research have led to the development of a number of

theoretical frameworks that seek to further the goal of accounting for all and only those structures that are possible in human language (Pierrehumbert, 2001a) by grounding both exceptionality and regularity in language structure in the statistical structure of language in use and in the general cognitive mechanisms that contribute to the human language ability. These approaches share much in common, and are variously termed “emergentist”, frequency-based, stochastic, cognitive grammar, and usage-based (Barlow & Kemmer, 2000; Bod et al., 2003; Bybee, 2001b; Bybee & Hopper, 2001b; Hopper, 1998; Langacker, 1987, 1991, 2000; MacWhinney, 2001, 2007, 1999). While this collection of approaches is still a somewhat loose conglomeration (Bybee, 2006), I will use the term “emergentist” to refer in general to approaches that attribute language acquisition and representation to general cognitive processes operating on statistical distributions across features in language use and experience.

With some variation, the goal of research within this paradigm has been to determine what grammatical information is available from individual’s language experience (most often based on statistical distributions and patterns of structure, Newport & Aslin, 2000) and to account for language acquisition and representation through general cognitive mechanisms (Langacker, 2000), usually relying on a unified model of the lexicon and grammar (Bates & Goodman, 1999; MacWhinney, 2005b, 2007). By attempting to determine what grammatical information may be acquired directly from frequency information in language input and how it may be acquired, this approach frames long-standing questions about language acquisition and knowledge in a new light and allows for a much more nuanced exploration of how human cognition may contribute to language structure.

1.1.1 The cognitive bases of language acquisition

It has long been observed that all children, except in cases of profound disability, acquire language, a finite set of resources capable of a seemingly limitless range of expression, with apparent effortlessness and universal success. Furthermore, this occurs despite the limited and varied exposure to language that children receive and despite the lack of explicit negative evidence that might keep the system from overgenerating. These facts have often been taken to require an innate, domain-specific learning system, a Language Acquisition Device, which supplies information about the range of possible structures in human language and guides the acquisition process (Bley-Vroman, 1989; Chomsky, 1965, 1980, 1986; Pinker, 1994). However, focusing in this way on the effortlessness, universality, and inexorability of child language acquisition obscures significant observations about language structure and human cognition that might elucidate language acquisition in greater detail.

Recent research on language from an emergentist perspective opens two significant avenues for advancement in this respect. First, by focusing on how statistical patterns emerge and become robust through the accumulation of repeated language experience over time, it allows for a principled way to test the assumption of the poverty of stimulus, suggesting that children's language input is not as impoverished as was previously thought. Moreover, the statistical robustness of existing patterns and the underrepresentation of ungrammatical or extremely low-probability patterns in language use have been shown to constrain overgeneration in grammar, fulfilling the task of negative evidence (Pierrehumbert, 2003b). Tomasello takes this as evidence that

“children *can* get from here to there” (2003, p. 3), meaning that detailed and comprehensive knowledge of grammar can be acquired from the input children receive, without recourse to innately specified grammatical universals.

The second way forward opened by an emergentist approach follows naturally from the first. If the grammatical structure of a language may be induced from emergent statistical distributions in accumulated language experience, there must be a set of cognitive mechanisms on which this process depends. Tomasello (2003, pp. 3-4) lists two sets of such mechanisms. In the first instance, a set of intention-reading abilities grounds language functionally in concrete social activity. Tomasello argues that this allows for different experiences to be compared, enabling the second kind of ability. This second set comprises a set of pattern-finding abilities, including categorization, the formation of sensory-motor schemas from recurrent patterns, the tracking of statistical distributions in perception and behavior, and the creation of analogical links between elements of more complex entities. These cognitive abilities are not specific to language (see also Langacker, 2000), but in an emergentist view they combine synergistically, yielding the human ability to acquire language.

This proceeds naturally from an evolutionary perspective in which wholly new and complex structures do not appear abruptly in organisms. Rather, existing structures are reorganized and added to gradually, with new capabilities emerging through this process (Bates & MacWhinney, 1989; Ullman, 2004). The unique constellation of abilities that contribute to human language is in a very real sense innate, and the particular characteristics of this system are likely to place constraints of some kind on the range and diversity of structures that are possible in human language. By investigating

this constellation of cognitive abilities in light of a detailed understanding of how grammar emerges through the accumulation of language exposure, emergentist theories make a significant contribution to our understanding of how language is acquired and used through interaction between human cognition and experience.

1.1.2 Implications of emergentism for second language acquisition

Such a view of language structure, acquisition, and representation has major implications, and holds significant promise, for work in the L2 domain, especially with respect to adult L2 learning (Achard, 1997; Achard & Niemeier, 2004b; Ellis, 2002a; MacWhinney, 2005b, 2007). Recourse to general cognitive mechanisms to explain L2 learning after puberty is not a new idea. According to Bley-Vroman's Fundamental Difference Hypothesis (Bley-Vroman, 1989), general problem-solving abilities are exactly what adults rely on in L2 learning because the innate, domain-specific learning mechanisms that enable children to learn language are no longer available. According to this hypothesis, children, on the other hand, have not yet developed general problem-solving abilities, and they rely instead on the availability of the Language Acquisition Device.

Krashen's Monitor Model (Krashen, 1981, 1982) proposed a slightly different version of this contrast. He distinguished between language acquisition and language learning, with part of the difference being that the former refers to unconscious and the latter to conscious processes. While Krashen argues that the processes underlying his concept of acquisition remain through the lifespan, in adults the Affective Filter may

render all or some input inaccessible to unconscious *acquisition*, leaving adults again to conscious learning and problem-solving mechanisms for the learning of additional languages. Acquisition processes are not necessarily coterminous with a strict notion of Chomsky's Language Acquisition Device (Gregg, 1984), but Krashen emphasizes the continuity between acquisition, in his sense of the word, in adults and children (see also Schwartz, 1986). The discontinuity in the processes underlying language acquisition/learning in individuals at different ages stems primarily from age-related changes in access to those processes.

The views of both Krashen and Bley-Vroman respond to a certain extent to the apparent effortless of child language acquisition, and the apparent effortfulness and varying success of adult language acquisition (here in the general sense of the word). As noted above, this has a tendency to obscure important insights into both the cognitive mechanisms that allow humans to learn language, and into the constraints that these may place on the nature of language itself. The crucial insight of emergentist approaches to language is that children too rely on general cognitive mechanisms, and not on cognitive structures specific and dedicated to language. These general mechanisms are not necessarily conscious problem-solving abilities such as those held by Bley-Vroman and Krashen to underlie adult second language acquisition, but rather refer to the ways in which humans record and structure memories of experience and use that knowledge productively in a variety of domains including language.

The complexities of language arise from interactions between general cognitive mechanisms and experience – in this case involving statistical distributions of features in language use (Ellis, 2002a; Langacker, 2000) – in a manner similar to the way

computational models can acquire certain aspects of grammar through the unsupervised operation of simple algorithms on language samples (Edelman, Solan, Horn, & Ruppin, 2003; Klein, 2005; Solan, Horn, Ruppin, & Edelman, 2003; Solan, Ruppin, Horn, & Edelman, 2002), but this is not limited to automatic or subconscious processes (see Ellis, 2006). Rather than a binary distinction between those processes relied on by children and those used by adults, small changes in some aspect of the total constellation of cognitive and environmental factors underlying language learning, representation, and use may alter the course and outcome of language acquisition. This recalls the thinking of Krashen (see above), who held that the processes underlying acquisition (as he defined acquisition) remained in adulthood but were rendered inaccessible by another aspect of adult cognition. The emergentist view opens the door for a more nuanced exploration of the interaction of multiple processes in a dynamic cognitive environment and their effect on language development.

An emergentist approach provides ways of comparing the cognitive resources and development of individuals at all stages in the lifespan and, critically, between young children and adults. It should not be taken to imply that language development or learning functions in exactly the same way in adults and children. Rather, the theories and methods discussed here allow for both differences and similarities to be evaluated with a great deal of precision regarding both the ways individuals of different ages participate in and are exposed to language in use and the cognitive abilities they bring to the task. It frames the questions of adult and child language acquisition in a similar way by asking what information is available to the learner through experience, and how do the learner's

cognitive abilities and resources act on and interact with that experience in the construction (Tomasello, 2003) of language knowledge?

1.1.3 Testing for emergent grammar in adult second language acquisition

This dissertation brings these insights to bear on adult L2 acquisition by evaluating learners' sensitivity to probabilistic patterns in Spanish phonology, and by probing the cognitive resources that may modulate this sensitivity. In doing so, it also opens the door to exploration of the implications of nascent bilingualism for a usage-based or emergentist understanding of language knowledge and processing (see Sebastián-Gallés & Kroll, 2003, on the usefulness of studying bilingualism in language processing research). It sees the L2 domain as a crucial testing ground for emergentist and stochastic theories and methodologies (Achard & Niemeier, 2004b; Ellis, 2002a, 2006; MacWhinney, 2005b). This dissertation combines a nuanced understanding of a particular probabilistic pattern in Spanish grammar, the diphthong/mid-vowel alternation, with a tradition of experimental psycholinguistic methodology sensitive to the subtle consequences of the unique structure of particular lexical items, and it adds a series of tests aimed at gauging the effect of various aspects of memory and cognitive control.

Spanish diphthongization has been selected for examination because research using metalinguistic judgment tasks has shown that native speakers have fine-grained knowledge of its statistical subtleties, and because that knowledge requires detailed sensitivity to different levels of phonology, morphology, and paradigm relationships. This dissertation first attempts to extend findings of native speakers' probabilistic

knowledge of Spanish diphthongization to online language processing tasks, and then seeks to use these tasks to detect the consequences of probabilistic variation in Spanish diphthongization in adult learners.

If adult learners are sensitive to the statistical distributions of this phenomenon in ways similar to native speakers, then their processing should reflect a dynamic progression approaching a native-like pattern of variability. On the other hand, if their behavior is more categorical in nature, this would indicate that adults in some way impose categorical structure on the variability encountered in actual language use. A third possibility is that their behavior may be variable, but that this variability may not approximate native-like patterns. In this case, it is possible that the learners may still rely on statistical learning abilities, but they may be tracking the distributions of a different set of features from those used by native speakers. Crucially, learner performance with Spanish diphthongization is evaluated against differences between learners' cognitive resources, experiences, and proficiency. The specific research questions to be addressed are as follows:

1. Are university-age foreign language learners sensitive to probabilistic patterns in Spanish diphthongization?
2. Does their pattern of behavior approach native speaker norms as proficiency increases?
3. Do individual differences in cognitive resources, namely working memory, attentional or inhibitory control, and phonological memory affect the development of probabilistic knowledge?

The goals of this dissertation draw heavily on several different areas of research. Therefore, the remainder of this chapter will serve to outline the prior work on emergent grammar, statistical learning, lexical processing, and cognitive resources on which the current project builds, and to integrate this work into the present approach. The strength of this perspective, again, is that it directly investigates the interactions between the environment and the specific cognitive resources and processes that give rise to the acquisition, representation, and use of language in all its structural complexity. As both the environment and cognitive resources evolve over the lifespan, the characteristics of this interaction and its results should be expected to change, but emergentism incorporates the dynamic nature of this evolution into a unified model of language development throughout the lifespan. It is this argument that forms the basis for this investigation of second language learning in university-age foreign language learners.

The following sections address three questions that follow from the emergentist perspective by reviewing a broad base of research into the statistical basis of language structure and processing. First, is the grammatical structure of a language available for acquisition based solely on statistical patterns in the use of that language? If so, is human language processing and intuition sensitive to this statistical information? Finally, what are the mechanisms by which humans are able to exploit statistical information in the acquisition and cognitive representation of grammar? Related to this last question, to what extent might working memory, phonological memory, and attentional or inhibitory control modulate language acquisition based on the statistics of usage?

The confluence of these strands of research provides a strong basis for the prediction that statistical patterns in the grammar of an L2 have consequences for

learners' online processing of novel lexical items that will gradually emerge over the course of L2 learning, modulated by individual differences in cognitive resources. This idea is developed as follows in the remainder of this chapter.

The following section reviews research showing how grammatical information at differing levels of complexity and abstraction is instantiated statistically in natural languages, and how that information may be extracted computationally. Furthermore, statistical patterns in language use have been shown to play a significant role in language change. Subsequently, we turn to a large and growing body of research demonstrating that the probability of structures at varying levels also impacts humans' use and processing of language, supporting the notion that the cognitive representation of language structure is grounded in the statistical distribution of features across language use. I then review work using artificial languages showing that children and adults are capable of extracting structural information from input in which the only cues available are statistical distributions. In addition, current research is discussed suggesting that adults may utilize the same kind of statistical information to extract at least some levels of grammatical structure from L2 exposure. The chapter closes with a discussion of the cognitive processes thought to undergird the emergence of grammar and a brief review of three types of memory and attentional control that may particularly affect adult L2 acquisition.

Following this discussion, in chapter 2 I present the particular diphthongization pattern examined in the present work, highlighting theoretical and empirical work showing how subtle yet statistically predictable patterns of variability in Spanish diphthongization have systematic consequences for native speaker behavior. In chapter 3,

I lay out the methodology used in the present study, and in chapters 4-6 I present and discuss the results.

1.2 Probability in language structure: Computational and descriptive evidence

The notion that grammatical rules may be extracted from patterns of language in use is fundamental to emergentist thinking about language, and the frequency of specific structures has been a central tool in showing how this might be done. Work in computational linguistics has explored how statistical information may be exploited for the development of machine translation, artificial intelligence, part-of-speech taggers, corpus tools, and pedagogical tools, among other technologies (Jurafsky & Martin, 2000). Of particular relevance here is the question of whether statistical distributions in language use can be used to induce grammatical structure without *a priori* knowledge about language structure.

A key insight into this question is that humans acquire language in an essentially unsupervised manner (Edelman et al., 2003). That is, the corpus of language experience of a child does not come tagged and labeled with phonological, morphological, syntactic, or other information.¹ Klein (2005) demonstrated that, with the use of two probabilistic models, one inducing bracketed tree structures and the other calculating word-to-word dependencies, accurate analyses could be obtained from small samples of text with no

¹ Children's language experience does come marked with information about the intentions of speakers. Tomasello assigns the ability to read the intentions of others, which emerges in children at about the time language begins to appear, a central role in unlocking the ability to acquire productive knowledge of language structure from frequency distributions in usage.

labeled examples in several languages. This is a useful result, given the labor-intensive nature of tagging a corpus, and shows that syntactic structure can indeed be extracted probabilistically from language experience. Of course, Klein's model was not entirely unsupervised. His corpus was divided into words and tagged for word category. Nonetheless, it is striking that the model arrived at an accurate analysis of syntactic structure given only knowledge of words and their categories.

The Automatic DIstillation Of Structure (ADIOS) model (Edelman et al., 2003; Solan et al., 2003; Solan, Horn, Ruppin, & Edelman, 2004) goes one step further. This model relies on a pattern acquisition algorithm that detects significant patterns in an untagged corpus and runs recursively until no further significant patterns are found. After training the model on 300,000 sentences from the CHILDES corpus of English (MacWhinney & Snow, 1985), they subjected the model to a multiple-choice grammaticality judgment test used in secondary-level ESL classes in Sweden. The model answered half the questions with 60% accuracy (in the remaining questions the model scored more than one answer as correct), even though it had only been trained on child-directed speech (Solan et al., 2004).

Computational models such as these show that it is possible, even with an untagged corpus (Solan et al., 2004, suggest that the ADIOS model could bootstrap itself given a corpus of words broken down into their constituent characters), to induce detailed and accurate syntactic structure from the statistical properties of the corpus. They are not, of course, meant to model how the human mind might extract these statistical patterns, but they do show that it is possible.

Another way in which grammatical information might be extracted from a corpus of language in use has been explored by researchers working in an analogical framework. Analogical models calculate the form of novel words according to the pattern of the most similar item in the lexicon.

Eddington used Analogical Modeling of Language (Skousen, 1989, 1992, 2002) to assign stress to Spanish words. The AML assumes that all known words are stored whole in memory (Eddington, 2000).² In order to determine the stress pattern of a novel word, the AML contained an algorithm which searched the lexicon for the most similar word based on the phonemic content and structure of the final three syllables along with tense and person (for verb forms), and it assigned stress to the novel item according to the pattern found in that word. If there were many similar forms with the same pattern, the probability that any one of them would be selected increased, but a single highly similar item would be more likely to be selected than less similar items. Based on the phonemic content and syllable structure of the final three syllables of each word, the model made correct stress assignments to 94% of the 4,970 words in Alameda and Cuetos (1995).

Eddington obtained similar data using a related algorithm, the Tillburg Memory-based Learner (Daelemans, Zavrel, van der Sloot, & van den Bosch, 1999), and, in a separate study, showed that the AML was also capable of choosing correct diminutive forms 92% of the time based on a corpus of other existing diminutives, with about half of the errors representing other attested forms (Eddington, 2002). Significantly, this implies

² This is a simplifying assumption made for the sake of the computational model. While there is ample evidence that humans store many morphologically complex words whole, or at least those higher in frequency, there is evidence that less frequent words are generally decomposed in storage.

productivity; existing forms may be memorized, but they may also form the basis for the construction of novel forms.

The issue of productivity is a natural area in which to explore a probabilistic approach to modeling grammar. Within the realm of morphology, Baayen (2003) has provided a way of modeling productivity based on the number of types (different entities) that appear in a certain set of tokens (individual instances of entities). Taking advantage of the notion that frequent words with a given affix are more likely to be stored whole in memory, such that their internal structure has less effect on processing (see also Bybee, 2001b), the productivity of that affix will rise if it contributes to a large number of infrequent forms. In other words, morphological productivity can be understood in terms of the contribution of an affix to the growth of the vocabulary as a whole.

Baayen offers several examples of this reasoning, illustrating the gradient nature of productivity. One significant example is the use of his model, MATCHCHECK, to explain the productivity of the German plural affix /-s/. This structure has been held up as evidence that regular rules depend on a different kind of processing, namely combinatorial symbolic rules, than do irregular structures (Clahsen, Rothweiler, & Marcus, 1993; Marcus et al., 1995; Pinker, 1998), because although German plural /-s/ is infrequent, it applies as a default, without respect to the phonological or frequency characteristics of stems. Baayen's model, on the other hand, shows that, given the existing lexicon of German plurals, the productivity of /-s/ can be calculated probabilistically.

One very interesting result of this model is that MATCHCHECK predicted that there would be no frequency effect for /-s/ plurals in German lexical access. This is in line with

lexical decision results for human language processing (Clahsen, Eisenbeiss, & Sonnenstuhl-Henning, 1997; Sonnenstuhl-Henning & Huth, 2002). The result is interesting because the lack of a frequency effect had been interpreted as evidence against probabilistic processing of default morphological structures such as this. Baayen's model shows that this seemingly contradictory pattern of behavior can result from the particular kinds of words and affixes that participate in productive morphological patterns. This result argues for a significant role for probability in modeling structure, while leaving open the possibility that a number of other factors join with probability in contributing to grammatical structure (e.g. speakers' reluctance on social grounds to use the Dutch feminine suffix /-ster/, which is otherwise predicted to be productive, via Baayen's model).

The studies reviewed in this section so far show that grammatical structure at certain levels is resonant in, and can be extrapolated from, lower levels of structure based on patterns of frequency. For language acquisition, this raises the question of where emergent structure begins. The ADIOS model of Solan and his colleagues discussed above provides evidence that complex syntax can be extrapolated from a corpus broken down into a string of constituent characters. If we assume that this can be extended to a string of phones, and not only graphemes, then a basis for phonology in something other than abstract linguistic entities will provide a grounding from which structure at all levels may ultimately be derived. This would not exclude the influence of other factors on grammar, such as the social reluctance to use Dutch /-ster/ alluded to above, or constraints on memory or processing, but it would support the notion that statistical patterns in language use underlie grammatical structure at all levels.

The interaction of vocal and oral anatomy, human audition, acoustics, and cognitive factors such as perception, memory, and motor control suggest such a base-level starting point for the emergence of phonemes and phonological grammar.

Pierrehumbert (2000; 2001a; 2001b; 2003b) presents a model in which all sounds are first perceived as distinct and are recorded in memory with fine-grained detail in a multidimensional phonetic space. The dimensions of this space are grounded in the physical realities of acoustics and anatomy, which serve as a starting point for organizing these detailed memories of sound.

As exemplars (specific instances of language experience, both in perception and production) accumulate in memory, certain areas of the phonetic space begin to emerge as statistical peaks, normal distributions reflecting the frequency with which exemplars are experienced at different points in acoustic or articulatory space. As frequency peaks emerge, they granularize the phonetic space in progressively greater detail, yielding emergent phonetic categories, which at first may correspond to something more akin to articulatory and acoustic gestures (Browman & Goldstein, 1992; Goldstein & Fowler, 2003; Hume & Johnson, 2001) than to single, steady-state phonemes. As new experiences continue to be logged, the probability that new exemplars will correspond to the peaks and not the valleys continues to increase, strengthening the integrity and resolution of the categories.

This notion of emergent phonetic categories receives strong empirical support from studies by Maye and colleagues (Maye, Weiss, & Aslin, In Press; Maye, Werker, & Gerken, 2002). They synthesized phones along a continuum of voice onset time and formant transition characteristics and presented them to infants in two experimental

conditions. In one condition, the frequencies of exemplars from different steps along the continuum formed a bimodal distribution, such that the infants heard a large number of tokens from each of two regions on the continuum, but fewer tokens from the area between the two peaks. In the other condition, the frequency distribution of exemplars was unimodal. Infants trained on the bimodal distribution were highly sensitive to contrasts across the boundary between the two distributions, but those trained on the unimodal distribution became desensitized to differences between steps on the continuum. This finding provides strong support for the conclusion that humans divide up perceptual space based on the statistical distributions among exemplars in their own language experience (C. Best, 1995; Flege, 2003; Kuhl, 2000).

Pierrehumbert's model thus provides a statistical approach to the systematicity of phonological grammar that is grounded in phonetics. The probabilistic nature of the model enables it to deal with the variation encountered in language between speakers, registers, acoustic conditions (e.g. a noisy restaurant, on the telephone), and a host of other variables. Furthermore, because the statistical patterns that emerge are highly robust, i.e. language in use contains far more exemplars of structures than it does distinct structures, this model also constrains the number of possible structures.

These studies are a small sample of the literature supporting the conclusion that detailed information about language structure, from phonology and morphology (Bybee, 2001b; Pierrehumbert, 2003b) through syntax (Croft, 2001; Goldberg, 1995; Manning, 2003; Schmitt, ; Wray, 2002) and sentence processing (Cuetos, Mitchell, & Corley, 1996; Gibson & Pearlmuter, 1994), may be extracted from statistical distributions across a corpus of actual language use. This conclusion forms the basis for Tomasello's claim that

grammar can be acquired from input without recourse to innately specified structural universals (2003). If all the necessary grammatical information is resonant in the accumulated language experience of a learner (whether child or adult), the next step is to ask whether there is evidence that human knowledge and use of language does indeed emerge based on the statistical distribution of features and elements in language use. That is, is a statistically based theory of language psychologically plausible (Langacker, 2000)?

1.3 The impact of lexical statistics on human language processing and use

To answer this question we now turn to research demonstrating that human language use is sensitive to the same kinds of probabilistic information that have been exploited in Natural Language Processing and other computational modeling research at a number of levels (Bod et al., 2003). First, research on language change provides evidence that the grammatical specifics of a particular language are malleable over time and respond to shifts in the frequency of words and structures resulting in the gradual emergence of new phonological, morphological, and syntactic entities. A second body of research shows that even very consistent rules and patterns are probabilistic in their behavior, and that speakers respond to this probability in their language use, processing, and metalinguistic judgments. Taken together, these areas of research provide evidence that human language knowledge is highly attuned to frequency in language experience. Linking this with the conclusion reviewed above from the computational literature that the grammatical structure of a language is resonant in language use through frequency,

this finding indicates that humans have access to grammatical structure at all levels through frequency distributions of linguistic elements in their experience of language.

1.3.1 Probabilistic effects in language usage and language change

Studies in the first category have sought to show that frequency plays a role in processes of language change. Zuraw (2003), in fact, argues that without a probabilistic model, language change could not take place. Generally, evidence of the role of frequency in language change is seen in the state of grammar and usage at a given point in time. One example of this is the tendency of English speakers to make “case errors” such as using *Sally and I* in an object Noun Phrase. While these are often hypercorrections based on prestige factors, Boyland (2001) points out that the most frequent *X and I* phrase (i.e. *you and I*) is also the most likely to appear as a hypercorrection. She argues that this kind of effect cannot be explained by prestige factors and suggests that the frequency with which hypercorrect forms are encountered positively affects the likelihood that similar forms will be produced subsequently. As this frequency increases, the use of non-standard forms such as these, or conversely, the appearance of object case forms (e.g. *him and me*) in subject Noun Phrases, become less and less likely to be perceived as errors, effectively changing the grammar.

Similar frequency effects can be noted in morphology, as for instance in the replacement of the past tense of English *go* with *went*, which was previously the past tense form of *wend*. According to Bybee (1985), this was possible because high frequency inflections such as these are less closely related to their base words and have

strong enough lexical representations to stand on their own. Thus, when a base form such as *wend* falls out of use, its past tense may remain and be reinterpreted as an inflectional form of a different verb, in this case *go*. Low frequency inflections that are more dependent on their base forms, on the other hand, are more likely to undergo the same changes as the base form.

This frequency effect in language change is also reflected in the observation that words with irregular morphology (e.g. *buy-bought*) tend to be higher frequency items. Phillips (2001) reviews evidence from a number of sound changes between Old and Modern English, including /t,d/ deletion, flapping, vowel lengthening, schwa deletion, /ö:/>/e:/ (Middle English), and stress shift. She makes the observation that changes which require morphological analysis of the lexical item affect low frequency items first. This supports the notion that low frequency items depend more on their membership in classes, such that when a change begins to affect the class, the low frequency item is likely to change sooner than a high frequency item with weaker connections to the class.

This represents an effect of *token frequency*, that is, the frequency of particular items in the lexicon. The more frequent an item, the stronger its representation or the more *entrenched* it becomes (Bybee, 2001b; Langacker, 2000), and consequently, the more independent its behavior in actual language use. That token frequency may affect units larger than words may be seen, for example, in French liaison, where, similarly, higher-frequency phrases tend to preserve liaison even as it disappears as a productive process in lower-frequency phrases (Bybee, 2001a, 2001b). This leads to a critical conclusion about frequency in human language. If the frequency of individual lexical

items can affect the likelihood that their use will change over time, then frequency must somehow be represented in humans' knowledge of grammar.

The additional observation that items with high token frequency are more likely to undergo leniting sound change provides further support for this conclusion, suggesting that each new experience of an item both strengthens its representation and increases the efficiency with which it may be produced and identified (Pierrehumbert, 2001a). This may be seen in the consonant deletion that gave rise to French liaison in the first place, and also in the phonetic reduction of frequent words and phrases such as *I don't know*, which may take the form [ãjɾəɔ̃] (Bybee, 2001b). In the latter case, the frequency with which the phrase is produced leads to reduction in the neuromuscular effort employed as well as in the phonetic detail required to identify it. Token frequency of this kind may also play a role in L2 learning, making some forms easier to remember than others, and rendering other forms more susceptible to regularization (in the case of irregular items) or to errors based on similar, but unrelated forms.

Type frequency, or the number of items that a given pattern applies to, affects the strength and scope of particular grammatical patterns, a factor that also may figure in the learning of consistent patterns or rules in an L2. An overwhelmingly dominant pattern, such as English past tense *-ed*, is almost always applied to borrowings or coined words, but other patterns may also be productive. For example, Spanish verb forms ending in *zco* (Bybee & Pardo, 1981), and large classes of English irregular verbs such as *string-strung* may induce errors such as *bring-brung* or be applied to nonce words like *spling-splung* (Bybee & Moder, 1983). It is this kind of frequency that Baayen (2003) exploited in his

model of morphological productivity (see above). As the type frequency of a structure increases, it is applied to more and more low-frequency items. These low-frequency items are more likely to be processed according to their constituents (e.g. morphemes) rather than as whole units in themselves, thus strengthening the productivity of the structural rule or process.

These observations suggest that the dividing line between the regular and the irregular is not well-defined, but rather gradient. Some theorists have argued that there are truly regular, “default” rules that have a different nature from other more gradient patterns (e.g. Clahsen et al., 1997; Pinker, 1998, 1999), but many of the rules held up as true defaults have been challenged as having their origins in the statistics of language use as well (e.g. Baayen, 2003; Tomasello, 2003, see above). Either way, it is apparent that statistical distributions of features and items in actual language use can be used to induce at least the majority of a language’s grammar, and further, that knowledge of those statistical distributions is necessary to explain a number of structural characteristics and trajectories of change in natural languages. This provides support for the role of statistical distributions in humans’ use of language over time, because, after all, language change takes place in and between human beings using language in actual speech communities. Extending these conclusions, we now examine a large and growing body of psycholinguistic research demonstrating the consequences of frequency, both for real-time language processing in individuals and for their metalinguistic judgments about language.

1.3.2 Frequency effects in phonology and morphology

Frequency effects in language processing and behavior have been reported in the literature for some time now. One of the most robust findings in the psycholinguistic research, for example, is the finding that high-frequency words are accessed faster than low-frequency words (see Jurafsky, 2003 for a review of probabilistic research in psycholinguistics). Moving beyond words to grammar, the role of frequency has been investigated in large-scale structures, such as relative-clause attachment (Cuetos et al., 1996; Gibson & Pearlmutter, 1994), as well as small-scale structures such as phoneme discrimination and phonotactics. The research reviewed in this section comes from a broad range of methodologies and takes a wide variety of approaches to grammatical structure. The research in this dissertation is situated at the confluence of work on phonological and morphological grammar, lexical access, and the impact of phonotactics, neighborhood density, and morphological families on language processing. The studies to be reviewed here all speak to different facets of this interdisciplinary work.

Given the relevance of phonotactic and morphophonological structure to this dissertation, I begin this section by reviewing a series of studies indicating probabilistic knowledge of varying levels of phonological and morphological grammar including allophonic distributions across varying phonological contexts and morphophonological alternations. Following this, I discuss a series of studies on probabilistic effects of phonotactics on language comprehension and production. The methodologies used in the research in this section range from metalinguistic judgments of the well-formedness of nonce words to reaction time studies, showing that the frequency of grammatical

structures is operative in both offline and online language processing. The findings reviewed here argue for a model of phonological grammar as probabilistic generalizations that are emergent over statistical patterns in a finely-detailed lexicon (Pierrehumbert, 1999).

One way in which frequency effects have been explored is in the notion of a *gang effect*, in which a pattern with high type frequency may come to exert influence over other words. This resembles the measure of productivity discussed above (Baayen, 2003), in which patterns that apply to a large number of items are more productive. In one early study (Bybee & Pardo, 1981), Spanish native speakers were presented with nonce verbs in written form (see also Section 2.2.2). The nonce verbs were constructed to contain alternating diphthongs, similar to a large class of diphthongizing verbs, e.g. *p[e]nsar* ‘to think’ vs. *p[je]nso* ‘I think’, in which the alternating syllable contains a diphthong when that syllable is stressed but a monophthong when it is not stressed. Despite the highly consistent nature of this pattern in alternating Spanish verbs, and despite being presented with two forms of each nonce verb, such that participants could clearly see that the nonce stems alternated, participants produced forms consistent with the alternation only 77% of the time. Eddington (1998) conducted a replication of this experiment utilizing a forced choice task instead of a free-response paradigm and obtained similar results.

Both Bybee and Pardo and Eddington argued that this result indicates that the diphthongization rule is to a certain extent lexically-bound. That is, it applies to certain lexical items, but does not as easily extend to new ones. Eddington also pointed to a group of verbs with non-alternating diphthongs, e.g. *adiestrar* ‘to train’, *adiestro* ‘I train’. He argued that the existence of unstressed diphthongs in forms of these verbs may have

exerted a gang effect, leading participants in these studies to produce and select forms of the nonce verbs with unstressed diphthongs even though the alternation would have prohibited them from appearing.

Bybee and Pardo conducted a subsequent experiment (1981), in which some participants were given the infinitive and third-person (3s) singular present form of the nonce verbs, and others were given the 3s preterit instead of the infinitive, in order to illustrate that the nonce verb followed the diphthongization “rule”. When asked to produce another preterit form, the subjects who had seen the 3s preterit were more likely to produce a monophthong in the critical syllable, in accordance with the “rule”, than the other group. The authors take this to indicate a relationship between surface forms in a paradigm, such that the presentation of another preterit form influenced participants’ responses in addition to the influence of the larger class of Spanish alternating verbs. The results of this study as well as Eddington (1998) paved the way for later work on stochastic phonology by showing that the application of even consistent patterns to novel forms depends in part on their similarity (in this case both phonological and semantic) to other forms. The size of the gang defined by this similarity (i.e. its type frequency) is one way in which frequency may influence grammar when speakers must compute the form of new words.

Another study probed the structure of “gangs” further. Bybee and Moder (1983) examined the class of English strong verbs exemplified by *string-strung* and defined a prototype for the class based on a series of phonological parameters. They then created a list of nonce verbs that varied in their similarity to this prototype (e.g. *spling*) and asked native English speakers to give the past tense forms (e.g. *splung*, *splinged*). The results

indicated that speakers were sensitive to the degree of similarity between a nonce verb and the prototype, suggesting that morphological classes are defined gradiently, with more and less central members. The closer a form is to the center of the distribution, the more likely it is to be a member of the class. Furthermore, this knowledge is productive, such that nonce forms are more or less likely to be subsumed in a class based on their centrality in that class, and speech errors may occur even in extant verbs with only marginal similarity to a prototype, as in *eat-ut* (Bybee & Moder, 1983, p. 263). This bears a strong resemblance to the frequency distributions of phonemes in Pierrehumbert (2001a; 2003b), by which sounds in experience are more or less likely to be classified as a certain phoneme based on their position in the distribution of prior exemplars in memory.

All of these studies showed a certain amount of productivity of their respective patterns, but they also demonstrate the gradience of morphological productivity and its susceptibility to the influence of similar forms in the lexicon. They also support the idea that morphologically complex words may be stored whole in the lexicon (Bybee, 2001b; Langacker, 2000), especially if they are higher in frequency. As noted above, this may also help to explain morphological productivity, because higher frequency words are less likely to be processed according to their component morphemes and therefore do not contribute as much to the productivity of their classes as do lower frequency items that are more likely to be processed as morphologically complex (Baayen, 2003). On the other hand, errors such as *ut* for the past tense of *eat* (above) show that even frequent forms may be recomputed during use, in this case using incorrect morphology. This is in

line with usage-based models of phonology (Bybee, 2001b), and receives support from a lexical decision study by Meunier and Segui (1999).

Meunier and Segui measured the priming effect of frequent and infrequent derivations on their stems, as compared to the priming effect of the stem on itself, and they also measured the priming effect of stems on frequent and infrequent derivations in French. They found that infrequent derivations prime their stems as much as the stems do themselves, but frequent derivations showed a smaller priming effect. This suggests that the frequent derivations are not decomposed into their component morphemes during lexical access, which results in a weaker priming effect. In their second experiment, however, they found that stems prime frequent derivations more than infrequent ones, an effect which seems to indicate that the processing of high-frequency derivations may be facilitated by priming their component parts. This asymmetry supports a model of lexical storage in which high-frequency derivations are stored both whole and decomposed, and in which these representations are accessed differently under different conditions.

This illustrates one effect of frequency, namely that high token frequency can lead to greater entrenchment of a form and thus to its functioning as a unit in memory and processing regardless of its decomposability into parts (Bybee, 2001b; Goldstone, 2000; Langacker, 2000). The representation at various levels (minimally, decomposed and whole) of morphologically complex words in the lexicon has obvious consequences for models in which grammatical patterns are extrapolated over the lexicon. First, the studies by Bybee and Pardo, Bybee and Moder, and Eddington reviewed above suggest that gang effects in the lexicon affect the application of productive morphological and phonological patterns. Changes in the contents of the lexicon due to different levels of representation

would affect the number of words that participate in a gang, and would also serve to reduce the productivity of patterns by reducing the impact of high-frequency forms (Baayen, 2003). These changes also have consequences for the Analogical Modeling studies on Spanish stress and diminutives reviewed above by altering the forms available for analogical comparison.

A slightly different tack on the probabilistic productivity of morphophonological patterns is taken in a series of studies by Albright and colleagues (Albright, 2002; Albright, Andrade, & Hayes, 2001; Albright & Hayes, 2003). These studies all involve computational models that induce a set of rules pertaining to a certain structure and then assess their reliability based on their performance in an actual sample lexicon. The resulting set of rules (in the case of Albright et al., 2001, an especially large set of over 3000 rules) included some very consistent rules, and other less consistent subpatterns, but all of the rules applied probabilistically, based on the phonological shape of the stem.

The results of the models were then tested against speakers' well-formedness ratings. Albright (2002) and Albright and Hayes (2003) examined the past tense in Italian and English, respectively, and in both cases found that participants' ratings of possible past tenses of nonce words depended on the phonological shape of the stem, even for the most regular pattern. Albright et al. (2001) applied a similar stochastic algorithm to the Spanish diphthongization pattern examined in Bybee and Pardo (1981), with similar results (see also Section 2.2.2). This result is significant because it demonstrates that the particular structure of individual lexical items influences the probability that a given rule or pattern will apply to a similar but novel item.

The influence of the lexicon on the probability of lower level phonological patterns receives support from research on quantity sensitivity, or the tendency of heavy syllables (i.e. those with codas or diphthongs) to attract stress, in Spanish (Alvord, 2003; Eddington, 2004a; Face, 2000, 2004; Waltermire, 2004). In particular, a study by B ark anyi (2002) involving trisyllables with differing patterns of light and heavy syllables (all light, heavy final, heavy penult, all heavy) indicated that quantity sensitivity is not a productive pattern in modern Spanish. When asked to mark the syllable they thought should receive stress in written nonce words (presented in sentential context), participants did not follow the canonical pattern of penultimate stress in words with a light final syllable and final stress for those ending in a heavy syllable. B ark anyi did, however, find some evidence that similarity to existing words in the lexicon influenced speaker preferences. She suggests that trisyllabic nonce words with three heavy syllables were disproportionately given antepenultimate stress because they resemble borrowings (e.g. *Washington*). This is consistent with a lexically-based model of grammar.

Other studies, however, have suggested that quantity sensitivity is at least partially productive. Syllable weight was a significant factor in Face (2000), in which native speakers listened to nonce words synthesized so as to contain no acoustic cues to the location of stress. While a replication of this study in which syllable duration was held constant (Face, 2004) showed an effect only for heavy final syllables, Waltermire (2004) and Eddington (2004a) corroborated Face's earlier results using written tests based on the same nonce words. Eddington also submitted the nonce words to an Analogical Model (Skousen, 1989, 1992, 2002) based on a database of the 4,970 most common Spanish words (see above, Eddington, 2000). The results from the algorithm

correlated most highly with native speaker intuitions when syllable weights and the final consonant were included as variables.

This does not necessarily contradict B ark anyi's results, because while her study did not support a categorical stress placement rule based on quantity sensitivity, she did find influence of lexically-based similarity networks on quantity sensitivity in native speaker judgments. The crucial finding is that phonological rules or constraints, such as those governing stress placement in response to syllable weight, are not categorical, but they are also not random. Rather, rules are probabilistic, possessing gradient productivity and consistency.

This is entirely compatible with an account in which the similarity of a nonce word to existing items in the lexicon on a variety of parameters influences the probability that the nonce word will match one pattern or another. Or to turn it around another way, the distribution of existing words in a multidimensional grammatical space (see Pierrehumbert, 2003b) creates statistical peaks that can act as attractors for new items, or alternately as schemas (Bybee, 2001b; Langacker, 2000) that are activated probabilistically based on the shape of the words or nonwords being processed.

With regard to quantity sensitivity then, quantity sensitivity and any final consonant appear to be among the relevant parameters for the productivity of Spanish stress patterns. Evidence for the importance of not only the final consonant, but also the final rime of a word comes from experiments by Aske (1990) and Face (2004). These studies capitalize on a subpattern of Spanish stress in which existing /n/-final nouns are stressed on the final syllable, but if the final rime is /en/, stress falls on the penultimate syllable between thirty and sixty percent of the time (depending on the corpus sample

used). In a nonce-probe task, Aske instructed native Spanish speakers to read aloud pseudonouns ending in /n/. Because the stress pattern for /n/-final verbs differs (generally penultimate regardless of the identity of the vowel in the final rime, except in the future tense) he embedded them in a sentential context to ensure their interpretation as nouns.

The results showed that Spanish speakers are sensitive to this probabilistic subpattern. They produced /en/-final pseudonouns with final stress 55.6% of the time, and with penultimate stress 43.5% of the time, but they produced other /Vn/-final pseudonouns with final stress in 96.8% of the cases. Face (2004) corroborated these results, albeit more weakly, in a perception task in which participants indicated where they perceived stress in synthesized pseudowords with no acoustic cues to stress (i.e. fundamental frequency, duration, and intensity). These studies provide additional support for a lexically-bound theory of phonology in which the probability of a given pattern (e.g. penultimate or final stress) depends on similarity networks based on multiple parameters, in this case the phonological content of the final rime and the syntactic category (noun vs. verb).

Networks of similarity across the lexicon such as those that constrain Spanish stress patterns have also been linked to long-studied and well-known phonotactic constraints. Two further examples will serve to close this section, showing that speakers' perceptions and judgments respond to the statistical characteristics of the lexicon with regard to phonological grammar.

The first example explored gradient effects of the morpheme-internal homorganicity constraint for nasal-obstruent clusters in English (Hay, Pierrehumbert, & Beckman, 2004). Hay et al. calculated the type frequency of a series of nasal-obstruent

clusters varying in the extent to which they satisfied the homorganicity constraint. They then cross-spliced auditory stimuli to create items with the various nasal-obstruent clusters. They found that participants' well-formedness judgments of these stimuli as well as reanalyses of nasal-obstruent clusters in their transcriptions correlated with the log frequency of the clusters. Most reanalyses also consisted of changes from less frequent clusters to more frequent clusters. This study provides evidence that nasal-obstruent homorganicity is not a categorical constraint, but rather that the well-formedness of a cluster varies according to its frequency and that of similar clusters in the lexicon.

A final example of the influence of statistics on phonological constraints comes from a study by Frisch and Zawaydeh (Frisch, Large, Zawaydeh, & Pisoni, 2001; Frisch & Zawaydeh, 2001). This experiment examined the obligatory contour principle for place of articulation (OCP-Place) in Arabic. Initially, Frisch and Zawaydeh found that Jordanian Arabic speakers' wordlikeness ratings of pseudoverbs differed significantly based on the presence of constraint violations, and they also found that the strength of the constraint varied according to the similarity of the consonant pair. However, they did not find a significant link between OCP-Place and the expected probability of the consonant pair. This supports the conclusion that OCP-Place is a psychologically valid constraint in Arabic, but it does not provide support for the influence of lexical statistics.

However, three factors led to a reanalysis of the data from this experiment (Frisch et al., 2001). First, a post-hoc analysis revealed that, while the pseudoverbs had been controlled for neighborhood density (see below), some tokens were highly similar to particular lexical items. Second, they noted that the expected probability of consonant pairs may not have been as appropriate as the transitional probability (the probability of a

second element, given a first) or the ratio between the observed frequency in the lexicon and the expected frequency (O/E), which can be used to distinguish between systematic and accidental gaps in the lexicon. Finally, Frisch and Zawaydeh (2001) had elicited wordlikeness ratings on a scale of 1-7, but in reality, participants relied on only the lower half of the scale (1-4), reducing the amount of variability in the data.

Frisch et al. (2001) therefore recoded the wordlikeness ratings by calculating the number of implicit rejections (ratings of 1) each token received (see Frisch, Large, & Pisoni, 2000). They then correlated this variable with OCP-Place, similarity of a nonword to a particular lexical item, neighborhood density, and the expected, transitional, and O/E probability measures. This reanalysis revealed significant effects for OCP-Place, similarity to particular items, transitional probability and O/E. A further regression analysis showed that O/E accounted for the most variance, but similarity to words and OCP-Place both increased the amount of variance accounted for by the model.

The results of this experiment indicate that patterns at a number of levels influenced the wordlikeness ratings, including frequency distributions in the lexicon. Further, the OCP-Place constraint itself applies gradiently to consonant groupings of varying similarity to one another. This highlights the importance of determining the specific features over which to calculate probabilities, and also shows that similarities at multiple levels may combine synergistically to produce grammatical patterns. Nonetheless, along with the research reviewed in this section on the lexically-bound and probabilistic nature of morphology and phonology, this study provides evidence that lexical statistics influence individuals' knowledge and use of grammatical patterns at many levels, including that of abstract symbolic rules and constraints.

This evidence is particularly relevant to this dissertation because it sheds light on the effects of frequency at the level of morphophonology, the level of the Spanish diphthongization pattern to be examined (cf. especially the studies by Albright et al., 2001; Bybee & Pardo, 1981; Eddington, 1996, 1998). In Chapter 2 I will lay out this pattern in detail. For now, we have seen strong support for the notion that grammatical structure in language is instantiated through frequency distributions across complex networks of similarity among linguistic entities in memory, and that humans' knowledge of grammar is keenly sensitive to these distributions. These networks of similarity encompass correspondences at all levels of structure.

This dissertation seeks to extend the findings reviewed here by exploring adults' sensitivity to the statistical characteristics of a particular pattern in Spanish morphophonology, that of diphthongization, and to do so using measures of lexical processing. These goals bring together research on the influence of statistics in synchronic and diachronic grammar with work on the effects of frequency in lexical processing, and relates these areas of inquiry to adult second language acquisition. If, as usage-based and cognitive theories claim, statistical patterns in grammar are learned through the interaction of general cognitive processes with language experience, then we would also expect individuals' knowledge of an L2 learned in adulthood to be probabilistic. Before turning to the question of second language acquisition, it is relevant to examine the kinds of lexical connections that are likely to affect the processing of alternating diphthongs in morphologically complex words.

1.3.3 Probabilistic phonotactics in language processing

There are three primary areas in which language processing research is relevant here. First, research on neighborhood density suggests that there are lexical links between words based on common segmental content, and that the number of highly similar items activated via these links has differential effects on processing depending on the size of the neighborhood and the frequency of the neighbors. Second, research on phonotactic probability indicates that speakers are sensitive to the probability of a given word as a function of the frequencies of its constituent parts, independently of the number of highly similar words. These two kinds of lexical relationships give an indication of the wordlikeness of novel items, such as those used in the present study, providing a much finer grained scale of the probability of the morphophonological composition of a possible word than do higher-level abstractions such as phonotactic constraints. Thirdly, there is abundant evidence that the relationship of an item to other morphologically similar forms, and to its own morphological family, has complex effects on the processing of that item. This research is highly relevant to the current investigation because the subpattern of Spanish diphthongization to be examined depends on the behavior of certain stems with particular affixes. Crucially, this involves knowledge of morphological correspondences between both stems and affixes across the lexicon.

All three of these factors are likely to affect the processing of both real and novel lexical items in language-specific ways. These effects are intricately tied to lexical, morphological, and phonological correspondences between specific lexical items, all of which must be acquired through experience of the language in use. For this reason, it is

important to understand how neighborhood density, phonotactic probability, and morphological correspondences have been investigated with native speakers, in order to contextualize the present study of language processing in L2 learners.

The effects of neighborhood density have been investigated extensively from a variety of approaches, yielding a complex set of results. Neighbors in the lexicon have generally been defined as the set of words that can be generated by changing one letter or phoneme of a given word (Landauer & Streeter, 1973). For example, *set* has a large number of orthographic neighbors including *get*, *pet*, *sit*, *sea*, but *apple* does not. Neighbors may also be calculated on the basis of common syllables (e.g. Álvarez, Carreiras, & Perea, 2004; Carreiras & Perea, 2002; Carreiras & Perea, 2004; Macizo & Van Petten, 2007).

Neighborhood density of various kinds has been shown to affect speakers' performance on a variety of tasks, but there has been disagreement over the direction of its effect under diverse conditions. Andrews (1997) reviewed a large number of studies and concluded that higher neighborhood density facilitates lexical decision and naming performance, but another review by Dell and Gordon (2003) suggested that the effect is facilitatory for production tasks, but that high density inhibits identification and repetition performance. A study by Carreiras et al. (1997) showed that dense orthographic neighborhoods had an inhibitory effect in several task paradigms, including lexical decision, progressive demasking, naming, and semantic categorization.

There is little doubt that the effects of neighborhood density depend on task conditions. Inhibitory effects have been attributed to competition between possible candidates that become active during processing, and facilitation has been linked to

possible excitatory connections between neighbors. The cognitive requirements of different tasks appear to indicate whether competition or facilitation will occur, with tasks involving selection from a set of candidates involving more competition, and tasks measuring lexical access benefiting from the additional activation present in larger neighborhoods (Dell & Gordon, 2003).

The investigation of neighborhood effects is further complicated by a number of factors. First, lexical decision latencies for nonwords created either by replacing one letter, or by transposing two letters in a real word, are slower if the real word is frequent than if it is infrequent. In essence, high frequency neighbors have a larger inhibitory effect on nonwords than do low frequency neighbors (Perea, Rosa, & Gómez, 2005).

In addition, Ziegler et al. (2003) manipulated phonological and orthographic neighborhood density and showed that, in auditory lexical decision, the presence of more phonological neighbors slowed response latencies, but dense orthographic neighborhoods speeded RTs. This result shows that, at least in French, orthographic and phonological neighbors have different effects on lexical decision, even in an auditory task. A further interesting implication of this result is that orthographic neighbors can be activated in the auditory modality, and that these similarity relationships also affect processing. While the current study does not directly examine orthography, this does have interesting consequences for the type of L2 learning examined here, since a great deal of instruction and study takes place in the written modality.

A study such as this, on second language learning, also raises the issue of crosslinguistic patterns in neighborhood effects. Languages may vary in the likelihood of neighbors at different locations in words, based on morphological, phonological, or other

patterns. For instance, while syllabic neighbors sharing a high frequency first syllable facilitate naming in Spanish (Carreiras & Perea, 2004), the first syllable seems to be not nearly as important in English (Macizo & Van Petten, 2007). A further example of crosslinguistic differences comes from a pair of studies that showed facilitatory effects of neighborhood density in English picture naming (Vitevitch, 2002) and inhibitory effects in Spanish picture naming (Vitevitch & Stamer, 2006). Vitevitch and Stamer attributed this contrast to differences in the morphological structure and the locus of neighbors between the two languages. This finding has particular implications for second language acquisition because learners may not create links between L2 lexical items in the same way as native speakers do, possibly changing the effective neighborhoods in their L2.

While decidedly complex, recent findings on neighborhood density might predict inhibitory effects of neighborhood density in the lexical decision paradigm employed in the present study. However, the relationship between the test items and the Spanish lexicon depends on many other factors (see below and Section **3.11.3.1**). Most relevant to the discussion of neighborhood density is the fact that neighborhoods play a role in lexical processing, affecting different kinds of processing differently. The activation of orthographically and/or phonologically related items is one way in which the likelihood of lexical form impinges on language processing. In other words, neighborhood effects show that the existence of similar words can affect speakers' performance with regard to particular lexical items. With regard to second language acquisition, this suggests that learners' performance will differ depending on both vocabulary size and the accuracy of form-based connections between lexical items.

Neighborhood density effects suggest that the particular characteristics and number of specific lexical items may affect how similar items are processed. This is one way in which the probability of certain lexical forms may be instantiated in language use. However, the traditional definition of neighbors is somewhat arbitrary, and is in any case highly specific. It is also possible to calculate the probability of a given lexical form based on the frequency of its components in the lexicon. Deciding on the scale of those elements is still a somewhat arbitrary decision, but robust effects for phonotactic probability have nonetheless been found at several levels of detail.

For example, two studies based on English (Coleman & Pierrehumbert, 1997; Frisch et al., 2000) calculated the positional probability of onsets and rimes in nonwords, taking the location of primary stress into account. Frisch et al. accomplished this by calculating the probability of each possible onset or rime in English in initial, medial, or final position (counting null onsets as possible onsets, such that neither initial rimes nor final onsets were possible), divided between stressed and unstressed positions. They then used these values to calculate the overall probability of a set of nonword stimuli. Higher phonotactic probability, by this measure, was associated with higher wordlikeness judgments by native speakers. This replicated the results of Coleman and Pierrehumbert (1997), who also showed that the overall probability was a better predictor of wordlikeness ratings than the probability of either the best or the worst element in a nonword. It appears that a generally high probability in the structure of a nonword reduces the negative impact of low probability components of the nonword.

Also interesting from a processing standpoint is the finding by Frisch et al. that speakers were more successful at recognizing high probability nonwords on a subsequent

memory test. This finding is particularly interesting because it suggests that speakers were able to use their knowledge of frequency patterns across the entire lexicon to boost their memory for nonwords (see also Gathercole, 1995).

It is not clear from these findings, however, how phonotactic probability might influence the timecourse of language processing. Furthermore, words from dense neighborhoods are likely to have higher probability structures, such that the two measures are often confounded. Both measures were investigated in a series of studies incorporating shadowing, lexical decision, and same-different (judging whether two stimuli were the same or not) tasks in the acoustic modality (Vitevitch & Luce, 1998, 1999, 2005; Vitevitch, Luce, Pisoni, & Auer, 1999). Phonotactic probability in these studies was calculated based on the positional frequencies of individual segments and on biphone frequencies. These studies suggested that higher neighborhood density had an inhibitory effect on response latencies, but that phonotactic probability generally facilitated responses except in lexical decision. They interpreted these findings, particularly the contrasting results in lexical decision, as indicating that phonotactic probability affects prelexical processing by facilitating the activation of word components, but that neighborhood density affects processing at the lexical level, where the presence of competitors slows response times. In a subsequent study, Luce and Large (2001) orthogonally manipulated neighborhood density and phonotactic probability and showed using a same-different task that there were simultaneous inhibitory effects for neighborhood density and facilitatory effects for phonotactic probability.

Research on both of these measures supports the notion of emergent networks of similarity-based connections across the lexicon, in these cases based on similar

phonological (or orthographic) content. In the case of phonotactic probability especially, these links provide a way of instantiating phonotactic constraints that is sensitive to the relative well-formedness or illicitness of particular phone sequences, as shown in research using well-formedness judgments. In addition, patterns of lexical frequency also impinge on lexical processing, modulating for example the dynamics of lexical activation and affecting both the timecourse and outcome of rapid, low-level processes during lexical access.

Links between similar elements at higher levels of analysis, such as those between morphemes, may also affect language processing, particularly respecting morphophonological alternations such as Spanish diphthongization. Two issues are relevant here. First, in what ways do morphologically related words affect processing? Second, and related to the first, to what extent are morphologically complex words stored and accessed as wholes?

There is evidence that access to morphologically complex words is modulated by the number of other words that exist with the same stem (morphological family size), but that there are also independent effects of stem frequency, affix frequency, and the frequency of the other family members (Burani & Thornton, 2003; de Jong, Schreuder, & Baayen, 2003). There is also evidence that, for existing morphologically complex words, the frequency of the whole word form is the best predictor of lexical decision latencies (Ford, Marslen-Wilson, & Davis, 2003). In addition to these effects, however, there is also evidence for influence of *apparent* morphological relatives, such that *corner* can prime lexical decision for words such as *corn*, even though no morphological relationship

exists between these words (Rastle & Davis, 2004). Words sharing initial sequences of segments, on the other hand, are not primed (e.g. *cornice* would not prime *corn*).

This is an interesting finding, because it suggests that there can be a phonological relationship between a derivational suffix, in this case *er*, and word endings with the same phonemic content but that do not constitute separate morphemes. This runs somewhat counter to the results of Meunier and Segui (1999), described above. Recall however that although they found evidence that frequent, morphologically complex words are not decomposed during processing under some conditions, there was also evidence that they could be decomposed under other conditions (stems primed high frequency derivations, but the reverse was not true).

Of further relevance to the current study is the finding that bilinguals are more likely to decompose high frequency inflected forms than are monolinguals (Lehtonen & Laine, 2003). In the study by Lehtonen and Laine, morphological composition predicted RTs in Finnish lexical decision for Finnish-Swedish bilinguals, but not for Finnish monolinguals.

From the research reviewed in this section it appears not only that statistical distributions of structure in language can be used to describe language structure, usage, and metalinguistic knowledge, but also that the probability of grammatical structure at a variety of levels impinges on actual language processing. Furthermore, it shows that speakers are sensitive to remarkably fine-grained detail at a variety of levels simultaneously, both in regular and irregular structures. This strongly suggests that those statistical characteristics play an important role in language acquisition as well as for language use and processing. These findings therefore reveal a very precise picture of the

task facing adult second language learners in acquiring morphophonological grammar, and also of the grammatical information available to them through their potential experience with the L2. In the next section I turn to a discussion of research demonstrating that the statistical distributions of various linguistic features can indeed be used by humans to acquire linguistic structure, both in first and second languages.

1.4 Human learning based on statistical patterns

The previous sections have shown that a great deal of information about the phonotactic and morphological structure of a language is available through networks of similarity between lexical items, both between lexical neighbors, which share most of their form in common, and in the probabilities of certain elements occurring in certain positions across the existing lexicon. This information has also been shown to impinge on human language processing, supporting the conclusion that these networks of similarity are instantiated in the cognitive representation of language.

Crucially, a growing body of evidence suggests that frequency distributions among similar entities, such as lexical neighbors and phonotactic patterns, play a role in language acquisition (e.g. Saffran, 2003). This has been shown both for children acquiring natural languages and in studies with both children and adults using artificial languages, providing a strong suggestion that emergent grammar in adult language acquisition depends on similar cognitive mechanisms to those relied on by children.

Regarding child language acquisition, for example, children's vocabularies have been shown to contain a large number of phonological neighbors, but that the overall

neighborhood density in their lexicon decreases as their vocabulary size increases (Coady & Aslin, 2003). It is possible to conclude, based on this trend, that connections between lexical neighbors aid in early vocabulary development.

Stronger support for the role of statistical patterns in the lexicon comes from research on the early acquisition of phonotactic patterns. A seminal study on phonotactic learning (Jusczyk, Luce, & Charles-Luce, 1994) revealed that 9-month-old infants learning English listened longer to high probability monosyllabic nonwords than to low probability nonwords, based on both the positional probability of phonemes and biphone probabilities. However, 6-month-old infants did not. This study provides evidence that, by 9 months of age, infants have learned which phonotactic patterns are more frequent in the language they are exposed to, and that they respond to this information in processing novel input.

More direct evidence that phonotactic development is highly language specific was obtained in a study of English coda acquisition by children (Zamuner, Gerken, & Hammond, 2005). Zamuner et al. hypothesized that, if language acquisition proceeded according to universal phonotactic constraints, then the most frequent codas crosslinguistically would be acquired first. If language acquisition is instead governed by the particular characteristics of the language being acquired, then codas should be acquired in the order of their frequency in English, which differs from the crosslinguistic ranking of codas. Zamuner et al. demonstrated that the latter was true, supporting the hypothesis that the statistical characteristics of children's L1 guide their acquisition of phonotactic constraints.

These studies of natural language provide strong evidence in favor of an emergentist view of language acquisition. However, natural language studies are complex and involve a large set of variables. Therefore, studies involving artificial languages in which the role of statistics and other factors can be more carefully controlled are of critical importance in demonstrating the relationship between statistical learning processes and language acquisition (for a review of earlier artificial language research with children, see Gomez & Gerken, 2000).

Artificial language studies have addressed a variety of levels of linguistic structure, from phonotactics to syntax. Of relevance to the phonotactic studies just mentioned, Chambers et al. (2003) trained 16.5-month-old children on monosyllabic nonwords in which specific consonants were restricted to initial or final position, e.g. [bæp], but not [pæb]. Head-turn results showed that the children were able to acquire these non-L1 phonotactic regularities. In a related study, Onishi et al. (2002) showed using a speeded repetition task that adults were able to learn first-order dependencies as in the above example, but they could also learn second-order dependencies in which the constraints on consonant position were also dependent on the intervening vowel, e.g. [bæp] and [pɪb], but not [bɪp] or [pæb]. The finding that both adults and children can acquire unfamiliar phonotactic patterns based on their frequency in exposure suggests that gradient patterns in L2 morphophonology may also be accessible to learners based on lexical experience.

These phonotactic studies are particularly relevant when we consider the calculations of the well-formedness of diphthongs in Spanish, based on detailed information about their segmental environments, in Albright et al. (2001). However, the

gradient subpattern in Spanish diphthongization examined in this dissertation requires more intricate knowledge than simple adjacency probabilities. It involves correspondences among phonotactics (diphthongs and monophthongs), paradigm relationships between the different forms of particular stems, and the behavior of certain classes of stems with particular derivational suffixes (see Section 2.2).

Further research on statistical learning in artificial languages suggests that even complex patterns relying on multiple levels of structure can be acquired based on statistical patterns in language experience, much as in the computational studies reviewed in Section 1.2. This research is grounded on a series of studies exploring the ability of both infants and adults to detect pseudoword boundaries given only transitional probabilities between syllables, employing both behavioral (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996) and ERP methodologies (Sanders, Newport, & Neville, 2002). At the syntactic level, similar research showed that if nonwords are ordered according to a finite-state grammar, infants can also discover possible syntactic structures based on statistical patterns, and can generalize these structures and apply them to strings of new nonwords (Gomez & Gerken, 1999).

Most importantly, statistical learning at these distinct levels is not independent, but rather information gained through statistical learning at one level of structure can be used to bootstrap further acquisition at that level, or acquisition at other levels. For instance, 9-month-old infants have been shown to use stress as a cue to word boundaries, in addition to their ability to segment words based on transitional probabilities. However, 7-month-old infants rely on statistics alone to segment words, suggesting that, statistical

learning can also serve to reveal other cues to linguistic structure, which may subsequently overtake statistics as a primary cue to structure (Thiessen & Saffran, 2003). In this instance, 9-month-olds appear to have discovered, based on transitional probabilities between syllables, that stress is a reliable cue to word boundaries in English. They then applied this knowledge to segment new language experience, even when it conflicted with statistical patterns.

Moreover, the information gained through statistical learning at one level, such as word segmentation, can also be used to learn patterns at higher levels of abstraction. Saffran and Wilson (2003) conducted a similar experiment to the word-segmentation studies discussed above, except that the words in their experiment were ordered according to a finite-state grammar. Given a continuous speech stream with no cues to word boundaries other than transitional probabilities between syllables, 12-month-olds were able to first segment the words, and then learn the word order rules of the finite-state grammar based on statistical patterns at the word level. Extending these findings to adults, Thompson and Newport (2007) found that adults could also induce syntactic rules working from cooccurrence frequencies among words to phrases and finally to larger syntactic relations. Adding another level of complexity, Thompson and Newport incorporated optional phrases, movement, repeated phrases, and different-sized form classes in the different experimental conditions. Interestingly, they found that adults performed best when all of these elements were included simultaneously, leading them to conclude that the addition of structured complexity may facilitate the learning of grammar.

All of these findings are encouraging for the exploration of emergent L2 grammar in adult learners. They suggest that adults and children have access to the same types of structural information via frequency distributions across the lexicon, and that they both have access to general learning mechanisms that enable them to acquire grammar from this statistical information. Moreover, these statistical learning mechanisms are not restricted to language, but are domain-general. Both adults and children are able to acquire similar kinds of structure for both adjacent and nonadjacent dependencies from tone sequences (Creel, Newport, & Aslin, 2004; Saffran, Johnson, Aslin, & Newport, 1999). This suggests that there is a basic continuity between the cognitive processes underlying language learning in adults and children. This opens the possibility of reframing the question of age-related differences in language acquisition in terms of long-term developmental differences in general cognitive mechanisms. One further strand of the statistical learning literature will serve to illustrate this possibility.

This strand of research stems from the observation that children may in some circumstances regularize probabilistic input, as when creole languages become more regular from one generation to the next. Evidence for this has been gained from the observation of deaf children learning ASL from parents who acquired the language in adulthood. The parents often produce certain ASL structures inconsistently, but there is ample evidence that their children acquire the same structures more consistently, in effect regularizing the inconsistent input they receive from their parents (Newport, 1999; Ross & Newport, 1996; Singleton & Newport, 2004).

Research using artificial languages has subsequently shown that children are more prone to regularize inconsistent input than are adults. Two studies (Hudson Kam &

Newport, 2005; Wonnacott & Newport, 2005) exposed children and adults to artificial language input in which the probability of a determiner appearing varied between experimental conditions. Interestingly, children tended to extend the most frequent structures, producing them more consistently, but adults did not exhibit this tendency as strongly. Adults tended to reproduce the statistical pattern in the input, although there was some evidence that adults do regularize when generalizing to new vocabulary.

The results of these studies suggest that there are differences in the ways that children and adults are able to make use of statistical information. One possible explanation is that adults' more developed cognitive resources, including perhaps working memory, allow them to track statistics in more complex ways than children. If this is the case, then task demands should also play a role in statistical learning. It is possible that adults learning a second language, a complex task requiring simultaneous processing on a number of levels, may tend to regularize certain structures, and that as their learning progresses, cognitive resources may be freed that will enable the acquisition of more detailed probabilistic patterns.

If this is so, adult L2 learners of Spanish may follow the general tendency with diphthongizing stems, which favors monophthongs in unstressed positions, and diphthongs in stressed positions. If, on the other hand, adults learning a second, natural language do reproduce input statistics more veridically, then it may in fact be easier for them to acquire the gradient subpattern in Spanish derivations with diphthongizing stems. There is, however, one difference between the studies of regularization cited here and the gradient morphophonological pattern that is the subject of the present investigation. In the artificial languages used by Hudson Kam and Newport and Wonnacott and Newport,

the probabilistic variability in determiner use was unsystematic. In Spanish diphthongization, the variability is systematic, depending on several levels of phonological and morphological knowledge. It is unclear whether children may still regularize systematic variability in some circumstances, but the research using artificial languages suggests that at least adults are unlikely to regularize this type of structure, except of course if the task demands involved in learning an additional language result in learning strategies that favor regularization.

The research reviewed in this section provides strong support for the hypothesis that both child and adult language learners extract grammatical information from the statistical patterns inherent in their language experience. Of further relevance to the present investigation is a significant body of research that shows how robust L1 phonetic categories interacts with the accumulation of L2 experience involving different phonetic categories. Based on the notion that phonetic categories are themselves comprised of probability distributions of exemplars in memory (Pierrehumbert, 2003b), this research further supports the hypothesis that statistical learning mechanisms undergird adult L2 acquisition. We now turn to a review of this literature.

1.4.1 Evidence for emergent language knowledge in a second language

This dissertation seeks to test for statistical sensitivity in adult learning of natural languages by experimentally examining points where learners may be sensitive to statistical information in the L2 Spanish lexicon. Until now, the statistical and emergentist approaches discussed above have been applied mainly to research in the L1

domain, although the research on adults using artificial languages provides strong evidence for a basic continuity in the cognitive processes that underlie certain aspects of language learning throughout the lifespan. There is also a lengthy literature on L2 phoneme discrimination and accentedness that, while not explicitly emergentist itself, provides further support for emergent structure in second languages learned in childhood or adulthood (cf. Pierrehumbert, 2003b), while also revealing the greater complexity underlying emergent structure in bilingualism (see review in Flege, 2003).

The central insight of this literature is that the cognitive representation of phonemes includes knowledge about variability in how they are encountered in actual language use. This does not refer merely to allophonic variation that is dependent on phonological or prosodic context or even just to registral variation, but it also refers to random variability between and even within speakers. That is, the sounds of language are not fixed points in acoustic or articulatory space, but they rather occur as frequency distributions of exemplars in memory. These exemplars of experience are not evenly distributed across the phonetic space but forms peaks and valleys, the peaks corresponding to prototypical allophones and the valleys occurring where articulations are rarely encountered in a language (see Section 1.2). In L1 research this corresponds to a number of phenomena including native speaker judgments of given sounds as “better” or “worse” examples of a certain sound, and to enhanced discrimination ability at the boundaries between peaks than closer to the peaks themselves (Maye et al., 2002).

In bilinguals and during L2 learning, these insights allow for detailed investigation of the accuracy with which L2 phones are perceived. Numerous examinations of bilingual performance via this fine-grained lens indicate that even

proficient and early bilinguals differ from monolinguals on measures of phoneme discrimination and production (Pallier, Colomé, & Sebastián-Gallés, 2001; Sebastián-Gallés & Bosch, 2002; Sebastián-Gallés & Soto-Faraco, 1999).

One recent example using errors in lexical decision showed that early sequential bilinguals in Spanish (L1) and Catalan (L2) failed to discriminate nonwords from real words when these differed only by a single vowel, which contrasted phonemically in Catalan but not in Spanish (Sebastián-Gallés, Echeverría, & Bosch, 2005). Simultaneous bilinguals only discriminated these items accurately if they were Catalan-dominant. The finding that even early and some simultaneous bilinguals fail to respond to subtle phonological contrasts in a second or non-dominant language, as measured in lexical decision error rates, is striking. It is possible, however, that language dominance is the crucial factor modulating discrimination performance in these studies, and not necessarily age (see also Sebastián-Gallés & Bosch, 2002).

On the other hand, there is evidence from ERPs that bilinguals do in fact learn nonnative contrasts, even when the L2 is learned after the onset of puberty. Winkler et al. (1999) showed based on the Mismatch Negativity (MMN) that fluent Hungarian-Finnish bilinguals who had learned Finnish relatively late discriminated between synthesized vowels pertaining to a single category in Hungarian, but to two categories in Finnish. Hungarian monolinguals, however, did not show a MMN for the same vowel tokens. This provides evidence that experience with L2 phonetic categories can and does alter low-level phonological processing in late bilinguals.

The notion that age does not necessarily preclude the acquisition of nativelike phoneme discrimination, but that other factors may inhibit this acquisition receives

further support from studies of Italian-English bilinguals living in immigrant communities. For example, Flege & MacKay (2004) showed that while early learners performed better in discriminating English contrasts not present in Italian, bilinguals who used their L1 infrequently also performed better, regardless of their age of L2 acquisition. (see also Flege, Frieda, & Nozawa, 1997; MacKay, Flege, Piske, & Schirru, 2001). These studies suggest that late bilinguals can learn to perceive nonnative phonological contrasts quite accurately, but that experiential factors such as L1 use may limit the extent to which this occurs.

The impact of L2 phonetic categories is not limited to perception. The studies reviewed by Piske et al. (2001) reveal that a variety of factors may affect accentedness, including continued L1 use, most of which directly affect the language input an individual receives in ways that are often confounded with age. Interestingly, the consequences of L2 phonetic experience impinging on L1 phonetic categories can have several results. On the one hand, bilinguals may end up with a hybridized system, as for example when Spanish-English bilinguals produce stop contrasts with Voice Onset Time values between those for each language (Flege & Eefting, 1987). On the other hand, Flege et al. (2003) showed that early Italian-English bilinguals produce the English diphthong [e¹] with a more exaggerated tongue gesture than English monolinguals, while late bilinguals tended to undershoot the monolingual target. Flege et al. attribute this contrast to the way in which new L2 categories are formed. Early bilinguals appear to have formed a new category, but then exaggerated the contrast between that new category and a similar L1 category. Late bilinguals have instead extended the L1 category, altering

its articulation towards a target between the L1 and L2 categories, as also seen in the Spanish-English VOT study.

This research suggests that L2 phonological acquisition is constrained by the L1 phonology, but that the interaction between first and second languages is bidirectional. This is consistent with the increasingly robust finding that L1 and L2 grammar and processing are inseparable, such that bilinguals' knowledge and use of their languages reflects the mutual influence of both the L1 and the L2, which interact in complex and synergistic ways (Cook, 2003; Hall, Cheng, & Carlson, 2006; Sebastián-Gallés & Kroll, 2003).

With respect to phonetic categories, the development of a bilingual system has been accounted for by various models including the Speech Learning Model (Flege, 2003), the Perceptual Assimilation Model (C. Best, 1995; C. T. Best, McRoberts, & Goodell, 2001; Hallé, Best, & Levitt, 1999), and the Native Language Magnet (Kuhl, 2000). These models accord a significant role to the accumulated experience of L1 phones in affecting the development of L2 patterns, similar to Langacker's notion of entrenchment (2000; see also Pierrehumbert, 2003b).

All of these models describe how the accumulation of exemplars of L2 experience may interact with the entrenched structure of L1 categories in parametric phonetic space. They differ as to the precise characteristics of this interaction, whether, for example, certain phonetic parameters are suppressed in perception or whether the existing structure affects the way new exemplars are categorized. Notwithstanding these differences, these models and the evidence on which they are based may be taken to support an emergentist view of L2 acquisition in which the statistical characteristics of both L1 and L2 interact.

They thus rely on a continuity in the cognitive underpinnings of child and adult language acquisition (Dehaene-Lambertz & Gliga, 2004; Flege, 2003).

While this dissertation focuses on diphthongization, a morphophonological alternation, and on stress, a suprasegmental phenomenon on which the general diphthongization pattern depends, the work reviewed here on the lower-level acquisition of phonetic contrasts lends strong support to the current hypothesis that, at least in certain instances, even adult learners may acquire gradient patterns in L2 grammar in a nativelike way. With this support, we now turn to a discussion of the cognitive processes underlying emergent grammar, paying special attention to certain cognitive abilities that may affect the course and outcome of the statistical learning of natural language grammar in adults.

1.5 Cognitive mechanisms in first and second language acquisition

As researchers have sought to explain how speakers acquire, represent, and use a probabilistic grammar that incorporates intricate networks of similarity between linguistic entities at all levels, a somewhat loose conglomeration of approaches has developed (Bybee, 2006) spanning what have been called emergentist approaches, usage-based theories, cognitive linguistics, and various connectionist and analogical models. While these theoretical perspectives are quite diverse, they generally seek to explain the learning, representation, and use of the full range of linguistic structure through general cognitive mechanisms, while avoiding the a priori stipulation of rules or structure (Langacker, 2000). The following subsection provides a general overview of the cognitive processes that have been proposed to fulfill this purpose. For the most part,

these proposals have been stated in general terms such that they describe cognitive processes that are operative throughout the lifespan. In the following subsection, three types of cognitive resources are discussed that may modulate the functioning of the more general processes put forth in usage-based theory and cognitive grammar, and which will be a focus of investigation in this dissertation.

1.5.1 Usage-based theory and Cognitive Grammar

From the emergentist perspective taken in this dissertation, linguistic structure is seen to emerge gradually and piecemeal (see Ellis, 2002a; 2002b and responses for a review relating this line of thinking to L2 acquisition) from the accumulation of language experience interacting with the particular characteristics of different aspects of human cognition (MacWhinney, 2001, 1999). As instances of language use, situated in concrete activity, are repeated and added to, patterns emerge based on the frequency of repeated aspects of experience (Tomasello, 2003). Emergence is thus an ongoing process of continual adaptation to the requirements of language use (Bybee & Hopper, 2001a; Hopper, 1998), and it is firmly rooted in the frequency of linguistic elements.

Central to this concept of emergence is the notion that “what may appear to be a coherent structure created according to some underlying design may in fact be the result of multiple applications or interactions of simple mechanisms that operate according to local principles and create the seemingly well-planned structure as a consequence” (Bybee & Hopper, 2001a, p. 10). As a consequence of this line of thinking, the distinction between grammar and the lexicon becomes increasingly blurred (Bates & Goodman,

1999). Grammar ceases to exist as a separate, autonomous system, and is reinterpreted as resonant patterns in structured memory of language experience. The lexicon is likewise no longer a list of words and idiosyncrasies, but a highly structured network of memories that function to support and constrain language production and perception.

The first implication of this approach is that usage-based grammatical knowledge relies heavily on memory (e.g. Nootboom, Weerman, & Wijnen, 2002). Frequency-based networks of similarity emerge gradually as successive exemplars are added to long-term memory from an individual's experience of language (Bybee, 2001b; Pierrehumbert, 2001a, 2003b). However, the structuration that takes place as the accumulating exemplars are integrated into networks based on their similarity with previous exemplars makes this storage much more efficient (Langacker, 2000). The repeated experience or activation of structural links strengthens their representation and reinforces the grammatical patterns instantiated in the lexicon (Bybee, 2001b).

Tomasello (2003) gives a comprehensive account of how this process might look in child language acquisition. The first requirement, in his model, is the ability to read intentions, based on the ability to share attention with others, to follow their attention to other entities, to direct the attention of others, and to imitate the intentional actions of others. More central to the present discussion are the pattern-finding, or categorization abilities that underlie the construction of grammatical knowledge. These include the ability to perceive similarity and create similarity-based categories, to form sensory-motor schemas based on repeated patterns of perception and action, to create analogies between complex entities, and to track the statistical distribution of perceptual and behavioral entities (Tomasello, 2003, pp3-4). This last ability is not meant to take the

form of direct statistical computation in the mind, but signifies that frequency of experience or use affects representations in memory (reminiscent of the strength of connections in connectionist modeling) or the entrenchment of neuromotor programs, or schemas (Bybee, 2001b; Langacker, 2000), or both.

Beginning with the ability to perceive and log similarity, the interaction of vocal and oral anatomy, human audition, and acoustic phonetics suggest a basic starting point for the emergence of structure in language. This starting point depends in part on structure found in the environment (minimally, acoustics), but it also relies on physical characteristics of the human vocal tract and brain, as determined in the human genome. As described above, repeated experiences of phonemes and other entities are logged in memory in great detail, leading to emergent clusters of experiences. These clusters have the effect of creating metrics of similarity (see Flege, 2003; Kuhl, 2000; Maye et al., 2002; Pierrehumbert, 2003b).

As experiences, both perceptual and behavioral, add up, sensory-motor schemas become increasingly entrenched (Bybee, 2001b; Langacker, 2000), and eventually become unitized (Goldstone, 2000). In other words, abstractions are formed over the statistical distributions of linguistic features at a given level through a process Langacker calls composition. For example, phonemes may emerge through functional similarities among phonetic categories, which are in turn based on distributions of phonetic exemplars in parametric phonetic space. In turn, similarities between compositional units at higher levels of abstraction may be logged and their frequencies tracked, allowing the emergence of still higher levels of structure. As this happens, higher levels of structure may interact with bottom-up emergent structure, affecting, for example, the perception of

new acoustic experiences (Kuhl, 2000; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), or the way in which new experiences are incorporated into structured memory (Flege et al., 2003). However, this does not necessarily represent a radical reorganization of the system, but rather a confluence and synthesis of different levels of representation (Pierrehumbert, 2003a). Exemplars of experience are still logged in memory and their frequency leads to continued evolution of emergent structures, but the emergence of higher-level structures adds to the complexity of the system, not replacing but rather interacting with lower-level processes.

This relates to the notion that as structure is abstracted from experience, the memory of that experience is not itself deleted. In other words, the instantiations of rules are not excluded from the cognitive representation of grammar (Langacker, 2000). For example, complete word forms such as *walked* may be stored whole in memory along with even the most regular rules that may be used to construct such word forms, in this case the regular English past tense rule VERB+*ed*. This contradicts what Langacker terms the *rule/list fallacy*, which stipulates that information derivable by rule may not be duplicated in the lexicon for economy's sake. On the contrary, given the assumption that "rules" emerge from frequencies in usage, forms must first be represented in memory for the brain to log their frequency, and it would be odd to posit some mechanism by which they were then erased, except for the general decay of memory over time (Bybee, 2001b).

This redundancy accounts for the fact that frequent linguistic expressions may become entrenched and function as lexical units while at the same time grammatical generalizations are made and used. Referring again to the emergence of the sounds of a language, the concrete exemplars that accumulate and give rise to the peaks and valleys

in phonetic space do not themselves vanish as a higher level of organization emerges (Bybee, 2001b; Pierrehumbert, 2001a). This is crucial to the notion of emergence, allowing for the evolution over time of lower levels of abstraction by the accumulation of new experiences to lead to change at higher levels of abstraction.

Crucially for the present investigation of L2 acquisition, the model of grammatical acquisition and knowledge sketched out here, which incorporates multiple, interacting levels of representation and structure, implies that language knowledge remains dynamic throughout the lifespan. Structures may be highly entrenched, and it may take a large amount of new experience to overcome the robustness of older statistical patterns, but grammar remains fundamentally flexible (Hopper, 1998).

Alongside this basic flexibility, however, there are many characteristics of adult cognition that may affect how new experiences are logged, categorized, and incorporated into already existing grammatical networks. The acoustic signal of speech encountered by adult L2 learners may not have changed, but the anatomy and cognitive resources of an adult have obviously developed substantially since childhood. This development alters the nature of adult cognition in complex ways, with inevitable consequences for emergent grammar (e.g. Newport, 1990). For example, the adult learner comes to the L2 with a highly granularized phonetic space based on their L1 phonology. This granularization has profound effects on perception, which subsequently affects the way a given acoustic signal will contribute to frequency distributions among exemplars in memory (e.g. Flege, 2003). In the same way, while the underlying process of language acquisition may be essentially similar between adults and children, the greater complexity of the system is likely to alter the course and outcome of that process.

With respect to the present investigation, it is difficult to predict the effect of L1 knowledge or entrenched neuromotor schemas underlying the perception and production of phonetic categories on the L2 acquisition of gradient subpatterns in Spanish diphthongization. However, there is a growing literature on individual differences in cognitive capacity and their effects on adult SLA. Given the reliance of usage-based theory on memory and categorization, certain of these individual difference measures have interesting implications for an emergentist theory of adult L2 acquisition, and these will be addressed here.

1.5.2 Working memory, phonological memory, and attentional control

Individual differences in various cognitive resources are likely to affect L2 acquisition because they affect how learners handle new experiences with language. This follows from usage-based theory, in which learners must store detailed information about surface forms in memory and then track statistical information within those representations. Individual differences that affect perceptual accuracy in an L2 or the level of detail that can be held in short term memory will necessarily also affect the patterns of frequency that emerge over a learner's accumulating L2 experience.

However, a wide range of results regarding various cognitive functions has been reported in the L2 literature, and it is unclear precisely how each of these contributes to this kind of learning. The present research therefore measures three different aspects of cognitive ability: phonological memory, central executive functioning (Baddeley, 2003), and attentional, or inhibitory control (e.g. Dijkstra & Van Heuven, 1998; Green, 1998), in

order to converge on the locus of the effects. Furthermore, this dissertation makes an innovative contribution by linking individual differences of this kind with an emergent notion of grammar acquisition.

Importantly, prior research suggests that phonological memory plays a significant role in lexical acquisition, and is therefore highly likely to affect the acquisition of lexical patterns. This is supported by the finding that children with higher nonword repetition scores used more complex vocabulary and longer utterances in their narratives (Adams & Gathercole, 2000). Phonological memory has also been linked to the ability of adults to extract grammatical gender in an artificial language (Williams & Lovatt, 2003).

Moreover, when adults learning a new natural language were allowed to rehearse L2 words, thus refreshing the contents of the phonological loop, they performed better on an array of linguistic and metalinguistic measures (Ellis & Sinclair, 1996).

However, most prior research in this realm has examined learners' ability to retain lists of nonwords or word-nonword pairs over the short term. While the short-term maintenance of phonological information is critical, this is in part because it undergirds the formation of long-term representations (Baddeley, Gathercole, & Papagno, 1998). The present study compares learners' phonological memory with their knowledge of the probabilities inherent in these networks of long-term lexical knowledge in order to enrich our understanding of how phonological memory enables the formation of long-term representations.

Another way that phonological memory is likely to affect the acquisition of statistical distributions in the L2 lexicon is through affecting the accuracy of short- and long-term representations of L2 items. At least at first, this takes place in an environment

of unfamiliar phonotactics and phonetic category boundaries. The traditional use of L1-like nonwords may therefore not be the most appropriate way to measure the influence of phonological memory on emergent L2 grammar. This is based on the findings of Gathercole (1995), who showed that the wordlikeness of pseudowords (in the L1) affected nonword repetition scores. Repetition of more wordlike items was more closely related to factors such as vocabulary size, suggesting that long-term knowledge of phonotactic and other patterns may have facilitated memory for these items. Repetition of less wordlike items was more closely related to measures of short-term memory. These results suggest that speakers exploit similarities to items or patterns in long term memory when holding novel material in memory, but are less able to use information in long term memory when the novel material is less familiar, as in L2 learning. Based on these findings, the present study relies on a nonword repetition task using words from an unfamiliar foreign language to measure phonological memory (see Section 3.11.17.1).

Second, the central executive is responsible for regulating attention, processing, and storage functions (Baddeley, 2003). Measures of working memory, such as the reading span used in the present experiment, address the storage and processing components associated with central executive functions. These measures are closely related to L2 reading, vocabulary, and grammatical measures and are thought to be linked to learners' ability to process larger chunks of the L2 in greater detail (see review in Miyake & Friedman, 1998).

Learners who differ in central executive capacity may be able to differentially attend to phonological information while processing L2 input, thus affecting their resulting grammars, but it is somewhat unclear how this will affect the acquisition of

detailed grammatical knowledge. On the surface, it appears that learners who can attend to more features of the L2 would be more likely to acquire such knowledge accurately. On the other hand, children have a lower memory span than adults, and it may be the case that adults' greater ability to attend to detail would complicate grammatical acquisition to the point of altering its outcome in adult learners (Newport, 1990).

Third, attentional or inhibitory control has been proposed as a mechanism by which bilinguals regulate activation of their languages. There are robust results showing that all of a speaker's languages are active to some extent during processing (Dijkstra, Grainger, & Van Heuven, 1999; Dijkstra & Van Heuven, 1998; Green, 1998), and speakers require some way of managing this information. Inhibitory control is thus likely to be crucial for the learners in this study, who must suppress the accumulated grammatical weight of their lifelong English experience in order to accurately track the frequency of the relevant L2 features. Greater control may thus facilitate the acquisition of fine-grained grammatical detail, as examined here.

Individual differences in these cognitive abilities are all likely to affect how learners deal with input. Since usage-based theory describes how grammar emerges based directly on the statistical patterns in the input, these factors can be expected to modulate learners' L2 grammars.

1.6 Summary

This dissertation draws on several strands of linguistic and cognitive inquiry, including probabilistic and usage-based approaches to grammar, statistical learning, SLA,

and cognitive resources. Situated at the confluence of these strands, it uses a sophisticated psycholinguistic methodology to bring new data to bear on the role of statistics in the L2 acquisition of fine-grained subpatterns in Spanish phonology, and on the role of certain types of cognitive resources in regulating that acquisition. These experimental techniques have been chosen for their sensitivity to subtle variations in the processing of probabilistic structures (Frisch et al., 2000; Pallier et al., 2001; Sebastián-Gallés & Bosch, 2002; Sebastián-Gallés et al., 2005), potentially before that variation is detectable in speaker intuitions (McLaughlin, Osterhout, & Kim, 2004).

This research thus opens avenues for the integration of new insights from L1 grammar and acquisition research with SLA, encompassing other areas of grammar as well as phonology (Achard, 1997; Achard & Niemeier, 2004a; Ellis, 2002a) and leading to a more nuanced understanding of variability and systematicity in adult SLA. This research will also pave the way for incorporating theories of cognitive resources with emergent or usage-based grammar. While the role of cognitive resources in language acquisition has received much attention in the past (Baddeley, 2003; Miyake & Friedman, 1998), including the link between phonological memory and vocabulary learning (e.g. Ellis & Sinclair, 1996; Service & Craik, 1993; Service & Kohonen, 1995), the implications of these abilities for the usage-based link between vocabulary learning and grammatical development have been neglected until now.

The remainder of this dissertation will be structured as follows. We now turn in Chapter 2 to a detailed discussion of the specific grammatical pattern examined here, diphthongization in Spanish derived words. The methodology is then outlined in Chapter 3, followed by two chapters analyzing the results of testing with native speakers of

Spanish, and with adult L2 learners, respectively. The final chapter reviews the results and discusses their significance in light of broader issues in language acquisition in general, and L2 acquisition in particular.

Chapter 2

The Diphthong/Mid-Vowel Alternation in Spanish Derived Words

2.1 Chapter overview

This dissertation sets out to test an emergentist view of adult second language learning by examining learners' processing of a complex morphophonological alternation that has been shown to vary in its predictability and well-formedness based on similarity relationships between existing words in the Spanish lexicon. Specifically, their sensitivity to a subpattern of Spanish diphthongization in derived words will be tested. If adult L2 learners can be shown to be sensitive to the subtle probabilistic nature of this alternation, this would provide evidence that adult L2 learning depends on a similar interaction of statistically sensitive learning processes with the statistical distribution of items and features in the lexicon to that underlie a usage-based conception of L1 acquisition.

As learners' experience grows in breadth and depth, both the type and token frequency of a wide variety of lexical items, features, and patterns also increases. This leads to the increasing availability of lexical data that may contribute to the growing reliability of generalizations, regardless of whether these generalizations are instantiated as networks of connections and correspondences between similar features (e.g. Bybee, 2001b), probabilistic distributions of exemplars in memory (e.g. Pierrehumbert, 2003b), sets of stochastic generalizations in the form of perceptual and neuromotor schemas

(Langacker, 1987, 2000) or symbolic rules (Albright et al., 2001), or, for that matter, stochastic Optimality Theoretic constraint rankings (Boersma & Hayes, 2001).

The question posed in this dissertation is, to what extent and to what level of detail are adult learners sensitive to the statistical distributions in their growing L2 lexicon, and to what extent does this sensitivity guide their grammatical development to approximate nativelike processing and behavior? In order to examine these questions, learners will be tested on their knowledge of lexically-based probabilistic variability in Spanish diphthongization.

Diphthongization has been chosen for two principal reasons. First, it occurs for the most part with a striking consistency in Spanish. However, its regularity is somewhat restricted to certain morphological contexts, whereas in other contexts its behavior is considerably more variable. These morphological contexts thus modulate the pattern in a probabilistic way. Second, several studies have established that native speaker behavior with regard to diphthongization is not only variable, but it reflects the statistical distribution of both regularity and exceptionality in the lexicon in surprisingly fine-grained detail. If second language grammar in adults is emergent in a similar way to first language grammar, then this earlier research on diphthongization provides both a strong indication that adult learners should be sensitive to its nuances and a clear prediction about how that sensitivity can be expected to shape adults' representation and processing of this structure.

This chapter provides a foundation for investigating Spanish diphthongization in second language learners by first outlining the descriptive details of its distribution in the Spanish lexicon and then reviewing experimental work investigating native speakers'

behavior. The chapter closes with a more precise description of the present research questions in light of this discussion, and details the predicted results of the experiments reported in this dissertation, both for native speakers and for learners.

2.2 The Spanish diphthong/mid-vowel alternation

Diphthongization in Spanish refers to the property of some stems by which different inflectional and derivational forms of those stems appear with either a diphthong or a monophthong. Diphthongizing stems occur in two flavors. In the first, the front mid-vowel [e] alternates with the diphthong [je], and in the other the back mid-vowel [o] alternates with [we]. The appearance of either a diphthong or monophthong in an alternating stem is highly consistent and predictable in verb paradigms, as exemplified in Table 2-1.

Table 2-1: Examples of diphthongizing verbs in Spanish

Sample forms with monophthong	Sample forms with diphthong	Gloss
<i>pensar, pensamos, pensé</i> <i>dormir, dormimos, dormí</i>	<i>piensa, piensan</i> <i>duerme, duermen</i>	think sleep

There are a large number of Spanish verbs that participate in this alternation, and with few exceptions the alternation correlates with stress position. If the alternating syllable is stressed in the inflected form, the diphthong surfaces (e.g. *piensa, duerme*), and if stress falls elsewhere, the monophthong surfaces (e.g. *pensé, dormí*). For this reason, most rule-based analyses have considered stress to be the trigger for diphthongization. The alternation is also stress-conditioned in many derived words

containing alternating stems. Since derivational suffixes generally remove main stress from the stem, alternating stems in derived words tend to contain monophthongs.

Nonetheless, there are also many attested derivations in which alternating stems appear with unstressed diphthongs, and there are even examples that sometimes appear with diphthongs, and sometimes with monophthongs, as illustrated in Table 2-2.³

Table 2-2: Derived words with diphthongizing stems

Base form	Attested derivation	Glosses
<i>niéve</i>	<i>nevóso</i>	snow, snowy
<i>vergüénza</i>	<i>vergonzóso</i>	shame, embarrassing
<i>puéblo</i>	<i>pueblíto</i>	village, village-diminutive
<i>viéjo</i>	<i>viejecíto</i>	old, little old man
<i>ciéno</i>	<i>cenóso</i>	mud, muddy
	<i>cienóso</i>	
<i>huéso</i>	<i>huesóso</i>	bone, bony
	<i>osóso</i>	
<i>caliénte</i>	<i>calentito</i>	hot, hot-diminutive
	<i>calientito</i>	

These forms illustrate that in many cases, alternating stems in derivations behave as predicted based on stress position (e.g. *nevoso*, *calentito*), but their behavior in other derivations (e.g. the relatively frequent word *viejecito*) contradicts the stress-based pattern. Still other derivations exist in both forms (e.g. *calentito*, *calientito*). Sometimes this variation is dialectal, but often these forms exist side by side as alternatives within a single language variety.

The existence of these words complicates the phonological analysis of the diphthong/mid-vowel alternation in Spanish. Furthermore, there are a large number of

³ Stress is marked with a diacritic in these examples to illustrate stress position, contrary to standard Spanish orthography

Spanish stems with mid-vowels that do not alternate at all (e.g. *comer* ‘to eat’, *beber* ‘to drink’), as well as several verbs with diphthongs that do not alternate (e.g. *adiestrar* ‘to train’, *amueblar* ‘to furnish’), although these examples do have morphologically related words with monophthongs (*destreza* ‘skill’, *moblaje* ‘furniture’). Whether a Spanish stem contains an alternating diphthong or not is thus opaque in Spanish.

2.2.1 Analyses of diphthongization in generative phonology

Following from the above facts from Spanish, there are two problems that must be solved in performing a linguistic analysis of this alternation, as well as in acquiring productive knowledge of the alternation. There must first be a way of determining which stems have a potential alternation, and then there must be a way of determining when the stem will contain a diphthong and when it will contain a monophthong. This information must be in some way available to speakers of Spanish, underlying their ability to both determine the correct form of known words (this could be accomplished in principle through rote memorization), and to use diphthongizing stems productively to create new derivations. This subsection reviews a series of traditional generative analyses of diphthongization, and the following subsection discusses a set of formal analyses of gradience and experimental work on the psychological reality and productivity of diphthongization in actual speakers of Spanish.

Spanish diphthongization has supplied linguistic researchers in a variety of traditions with significant descriptive problems for decades. A variety of generative

analyses of this phenomenon have been proposed (e.g. Harris, 1969, 1995; Hooper & Terrell, 1976; Roca, 1988), that provide a helpful exploration of the relevant issues here.

Harris reviewed several attempts to account for Spanish diphthongization and concluded that the existence of non-alternating monophthongs and diphthongs necessitates that alternating stems be specifically marked with a diacritic in the mental lexicon (Harris, 1969, 1977). This would serve to identify alternating stems. Harris then proposed that, given the presence of the diacritic, stress placement was the trigger for diphthongization, in keeping with the fact that, when the alternating syllable is stressed, it most often surfaces as a diphthong. Since derivational suffixes remove stress from the stem, derivations with alternating stems are thus predicted to contain monophthongs. Derivations with unstressed diphthongs could be accounted for by cyclic rule application, in essence by stipulating that sometimes the alternation is triggered after stress is removed to the suffix (resulting in an unstressed monophthong), and sometimes it is triggered before stress moves to the suffix (resulting in an unstressed diphthong).

Harris, however, did not provide a way of determining when the diphthong status of a stem would be determined before stress shift, and when it would be determined afterwards. A later attempt was made to correct this problem by making cyclicity a property of affixes (Halle, Harris, & Vergnaud, 1991). That is, the stress-based diphthongization rule would be applied after addition of certain suffixes (e.g. *ero*, *dad*), resulting in a monophthong, but not after the addition of other suffixes (e.g. *ísimo*, *ito*). This correctly predicts the form of words like *bondad* ‘goodness’, for example, because the suffix would first remove stress from the stem, *bueno*, and subsequently a monophthong would surface in the stem because that syllable is unstressed. The proper

form of *buenísimo* ‘very good’, on the other hand, is also predicted because the diphthongization rule would first apply to the stem *buen-*, yielding a diphthong because the syllable is stressed, and subsequently the suffix would be added, removing stress from the diphthong.

Halle *et al*'s account attempts to relate the nuances of the interaction between stress and diphthongization to the characteristics of specific suffixes. However, the fact that both monophthongs and diphthongs may occur in derivations with the same suffix is problematic. Furthermore, Eddington (1996) points out that the suffix *ista* occurs equally often with diphthongs and monophthongs in a large, dictionary-based corpus, making it difficult to tell which category this suffix pertains to. This exposes a further difficulty with the cycle-based analysis, in that there is little independent justification for the classification of suffixes as cyclic and non-cyclic other than the fact that they do or do not generally cooccur with unstressed diphthongs.

Like the diacritic analysis, the labeling of cyclic and non-cyclic suffixes captures the fact that alternating stems must be distinguished from non-alternating stems, and that specific suffixes influence the application of the stress rule, but they do not provide a way of explaining the distribution of these labels. Even with these stipulations, the inconsistent behavior of particular derivations continues to require that they be marked as exceptions in a rule-based analysis.

While unsuccessful in their ambitions of determining a rule or set of rules that categorically determines the behavior of the diphthong/mid-vowel alternation in Spanish, both of these analyses contribute insights toward a different type of analysis. Halle *et al.* worked with the notion that the behavior of certain suffixes modulated the behavior of

alternating stems, and Harris attributed the variability in diphthongization to the properties of particular derivations by marking them with brackets to indicate the timing of the diphthongization rule. Furthermore, Harris concluded that the only way to distinguish alternating stems from non-alternating stems was through marking them as such in the mental grammar.

These two insights tie both the identification of alternating stems and the behavior of those stems in different lexical and morphological contexts to the particular characteristics of individual lexical items (see Bybee & Pardo, 1981). The analyses reviewed above accept to certain degrees the lexically-bound nature of diphthongization, and then attempt to describe how those features interact with larger-scale generalizations to produce the patterns observed in Spanish grammar. This is simply the basic endeavor of grammatical analysis. The challenges in these analyses emerge with the attempt to integrate this interaction into a categorical symbolic rule or set of rules that determines in each case whether a stem should appear with a diphthong or with a monophthong. The frequent existence of situations in which either choice is both possible and attested in the language, whether in the behavior of particular suffixes or particular derivations, make this impossible without simply marking certain forms as exceptional.

This misses the important generalization that even exceptional forms have an impact on the processing and production of related forms. Grammar depends on both large- and small-scale generalizations over phonological and morphological features such as stress and suffixes. In a way, emergent grammar counts all attested forms as exceptional, since they may be stored and accessed whole if they are relatively frequent. At the same time, the lexicon of attested items is structured in a way that impacts both the

processing and production of new experiences of language. Patterns in the lexicon therefore vary in their consistency, responding to the presence of more and less exceptional forms, such that grammatical patterns such as Spanish diphthongization are gradient, rather than categorical. Gradience in itself does not necessarily preclude the involvement of symbolic rules (Hay & Baayen, 2005), but it suggests that the generalizations that may be made vary in their reliability, in the same way that semantic or syntactic cues of varying strength and reliability may interact to suggest one interpretation or another in sentence processing (e.g. Bates & MacWhinney, 1989; MacWhinney, 2005a). Another way of viewing gradience in diphthongization is that the particular phonological and/or morphological structure of a word and its relationship to the entire lexicon modulates the probability that a diphthong or a monophthong will surface. This view is at the heart of emergentist and usage-based theories of grammar (Baayen, 2003; Bybee, 2001b; Bybee & Hopper, 2001b; Langacker, 2000; MacWhinney, 2005b, 2007, 1999).

2.2.2 Descriptive and experimental accounts of gradience in Spanish diphthongization

Several researchers have examined gradience in Spanish diphthongization, addressing both the identification of diphthongizing stems and the factors that influence their behavior. These studies have the additional advantage, for the purposes of the present dissertation, of providing and testing highly detailed predictions about how gradience might impinge on the productivity of diphthongization for native speakers.

Albright et al. (2001) addressed the question of how stems with alternating diphthongs may be distinguished from non-alternating stems in both the formal and cognitive representation of Spanish grammar. They constructed a machine algorithm that compared the segmental content of pairs of Spanish stems and used that information to determine a set of rules that would identify alternating stems from non-alternating stems. The algorithm began by determining highly specific rules concerning the relationship between two individual stems based on their segmental content, and then proceeded to construct larger-scale generalizations over these rules.

The algorithm built by Albright et al. generated a set of 1,698 word-specific rules and 1,648 generalized rules, a very large grammar. The algorithm then calculated the adjusted reliability of each of these rules based on both the number of stems to which the rule could be applied and on the accuracy of the rule in classifying those stems. On the basis of this stochastic grammar, the likelihood of an alternating diphthong in a given stem could be determined based on the segmental content of the stem.

This represents an important step beyond the traditional generative analyses reviewed above by providing an independent way of identifying alternating stems. Crucially, this identification is not categorical but probabilistic. While speakers may simply know the status of existing stems based on experience, this algorithm makes predictions about how native speakers might treat novel stems based solely on their segmental content.

In the second part of the study, Albright et al. tested these predictions using a nonce word experiment. They constructed a set of nonce (verb) stems with possibly alternating mid-vowels (e.g. *lerrar*) that varied in their likelihood of alternating with a

diphthongized form based on the machine algorithm. They embedded forms of these nonce stems in short dialogs and then elicited forms that would contain a diphthong if the stem were classified as alternating. The possible alternating syllable was unstressed in the embedded forms (e.g. *lerramos*), but it would receive stress in the elicited form. The experiment thus relied on the highly consistent stress-conditioned nature of diphthongization in Spanish verbs. Following the elicitation, both possible forms of the nonce stem were offered and a metalinguistic judgment was collected about which of the forms was better.

Both the nonce verb forms produced spontaneously by native Spanish speakers and their judgments in the forced choice questions show that speakers are indeed sensitive to the detailed segmental information used in the computer algorithm. In short, nonce stems that “sounded better” with diphthongs, according to the model, were more likely to be produced with diphthongs spontaneously in the elicitation task, and speakers were also more likely to choose forms with a diphthong in the forced choice. This is an important finding for the present discussion because it indicates that speakers respond to fine-grained patterns of cooccurrence in extending diphthongization to novel lexical items.

This may help to explain why native Spanish speaking participants in another nonce verb study (Bybee & Pardo, 1981; replicated in Eddington, 1998) only followed the regular, stress-based diphthong/mid-vowel alternation for 77% of nonce verbs, even when given two allomorphs (e.g. *bercar~bierca*) in the stimulus sentences to show that the stems did in fact contain an alternating diphthong. Bybee and Pardo obtained this result by eliciting spontaneous production of nonce verb forms, and a replication by

Eddington (1998) using a forced choice paradigm did not produce significantly different results. In the latter study, participants chose the form predicted by the position of stress in only 81% of cases. While this indicates a clear tendency to follow the general, stress-based pattern, these results contrast strikingly with the fact that the alternation is almost completely regular within Spanish verb paradigms.

Bybee and Pardo as well as Eddington interpreted their results to indicate that diphthongization in Spanish is a morphologically-bound alternation. That is, it affects only particular stems, and is not easily extended to novel lexical items. On the other hand, participants in their study as well as in Eddington's replication did produce or select a majority of forms that conformed to predictions, suggesting that diphthongization is productive to a limited extent. In conjunction with the study reviewed above by Albright *et al.*, these findings support the conclusion that grammatical patterns exhibit gradient productivity.⁴

These studies, particularly Albright *et al.* (2001), suggest that speakers can identify alternating stems based minimally on the rate of occurrence of alternating diphthongs in similar segmental environments. This accounts for the first problem in analyzing or acquiring Spanish diphthongization, as discussed above. The second question, how it is that speakers determine which derivational and inflectional forms of a given alternating stem should contain a diphthong or not, was addressed in two studies by Eddington (1996; 1998).

⁴ Proponents of dual-process models hold that there are truly regular, default patterns that are not probabilistic but rather categorical, while exceptional patterns may be probabilistic. These two types of pattern are thought to depend on different cognitive mechanisms (Pinker, 1998, 1999; Ullman, 1999). Discussion of dual-process models is somewhat tangential to the present study, because it examines a demonstrably probabilistic pattern.

Eddington capitalized on the observation that the probability of an anomalous pattern (i.e. unstressed diphthongs in derived words with alternating stems) varies gradiently depending on the particular suffix used. He selected a set of ten derivational suffixes and compiled a list of all attested derivations with alternating stems based on these suffixes in the *Diccionario de la Lengua Española* (Eddington, 1996; Real Academia Española, 1970). He subsequently reduced this sample to the set of derivations most likely to be familiar to most adult Spanish speakers. For this sample, the proportion of words with stem monophthongs is shown for the less productive suffixes in Table 2-3. Eddington noted that the more productive suffixes (i.e. the diminutive and augmentative suffixes *ísimo*, *ito*, *illo*, *zuelo*, and *azo*) were likely to be poorly represented in dictionaries based on their considerable productivity, and he excluded them from this calculation. They are marked with a ‘*’.

Table 2-3: Percentage of attested derivations with monophthongs and percentage of monophthong forms of neologisms selected by native speakers (Eddington, 1996, 1998). Suffixes are ranked according to Eddington (1998)

Suffix	% of monophthongs in attested derivations	% of monophthongs chosen in neologisms	
		Eddington 1996	Eddington 1998
<i>ero</i>	83	75.4	72.5
<i>al</i>	100	62.7	66.7
<i>dad</i>	100	47.8	64.7
<i>oso</i>	100	86.2	59.8
<i>ista</i>	50	67.8	50
<i>ísimo</i>	*	19.2	49.5
<i>zuelo</i>	*	40.9	48
<i>ito</i>	*	4.7	25.5
<i>azo</i>	*	44.2	24.5
<i>illo</i>	*	6.1	22.5

All of these suffixes have the effect of removing stress from stems, yielding the general prediction that these derivations should contain stem monophthongs. This list corroborated earlier observations that certain suffixes tend to allow unstressed diphthongs in the stem more than others (e.g. Halle et al., 1991, above). Crucially, however, it gave an indication of the gradient variation in the probability of an unstressed diphthong associated with each suffix rather than dividing them categorically between suffixes favoring monophthongs and those favoring diphthongs. The frequency of unstressed diphthongs in the whole dictionary sample varied from 50% to 90%, depending on the suffix, and was similar in the reduced sample. Eddington hypothesized that speakers' preference for diphthongs vs. monophthongs in novel derivations would correspond to this observed probability of diphthongs or monophthongs in the dictionary corpus. For the most productive suffixes, for which he did not consider his corpus data to be reliable, he hypothesized that diphthongs would be preferred, based on the existence of relatively frequent derivations containing diphthongs, and on Halle et al. (1991).

To test this hypothesis, Eddington joined alternating stems with the ten suffixes to form previously unattested neologisms (Eddington, 1996, 1998). The crucial difference between this study and the nonce word studies reviewed above is that the stems used were common alternating stems known to participants, as opposed to invented stems that may or may not have contained an alternation. These studies thus sought to determine when speakers would prefer each alternate of the stem.

In both of these studies, Eddington used a forced choice paradigm to determine which form of each neologism would be selected by participants. In Eddington (1996) speakers read definitions of the neologisms and then selected the form they deemed to be

correct. In Eddington (1998) he further highlighted the fact that the stems could contain either a diphthong or a monophthong by using sentences including attested derivations containing both forms of each stem, as shown in the following example:

No nevó mucho, y la nieve fina que sí cayó no era más que una ____ ligera.

a. *nievita* b. *nevita*

‘It didn’t snow much, and the fine snow that did fall was nothing more than a light snow-diminutive.’

In this example, the forms *nevó* and *nieve* illustrate that the stem for snow may contain either a diphthong or a monophthong.

The results, also shown in Table 2-3, showed that native speaker preferences closely reflected the probabilities observed in the corpus. Native speakers chose monophthong variants of the neologisms more frequently when they used suffixes that cooccur most frequently with stem monophthongs in the corpus, and they chose variants with diphthongs more frequently with either the most productive suffixes, or with those suffixes that cooccur more frequently with diphthongs in the corpus (recall that Eddington assumed, based on prior studies and intuitions, that the most productive suffixes do in fact frequently cooccur with diphthongs, although the dictionary data did not show this pattern).

As can be seen, there was some variability among the dictionary results and Eddington’s two experiments with regard to the specific ranking of the suffixes. However, the rankings of the suffixes were relatively stable between these three measures, and there were no cases in which a significant preference reversed direction. Further research into the effects of methodology on this type of metalinguistic judgment,

using more accurate corpus data, might help to clarify the slight variability that was observed.

Together with the studies and observations reviewed earlier in this section, the results of Eddington's experiments show that speakers have sophisticated knowledge of statistical patterns that depend on knowledge not only of stem allomorphy but also of the gradient probabilities of a systematic alternation across many stems in particular morphological contexts. Emergentist theories hold that this knowledge arises through the recursive functioning of general cognitive processes as individuals accumulate more and more experience of a language. The synthesis of findings from these two strands of research strongly suggests that probabilistic knowledge of grammar and usage continually evolve throughout the lifetime. Although the precise cognitive resources that undergird speakers' sensitivity to patterns in language usage continue to develop, the basic process of the accumulation of linguistic exemplars in memory is likely to remain intact (see Pierrehumbert, 2001a).

This suggestion opens up fascinating possibilities for L2 research and creates the potential for explaining much of second language acquisition, including the observation that adults who learn a second language rarely achieve nativelylike proficiency, based on the interaction of language experience and the cognitive resources that individuals bring to the task. The emergentist hypothesis suggests that even adult learners are likely to develop fine-grained knowledge of statistical variability in grammatical structure, although it may differ from native speakers' knowledge of the same structures. Far from being 'deficient language learners', second language learners interact with language experience in highly sophisticated ways (Hall et al., 2006). By incorporating a detailed

analysis of the role of general cognitive processes in the development of this knowledge, emergentist theory provides a way forward into determining why, if adult second language acquisition is so sophisticated, its results do not match the results of first language acquisition. The findings reviewed here concerning diphthongization in Spanish provide firm reference point for this kind of work in the L2 domain.

This is the topic of the current research, but before describing the present experiments in greater detail, the results of an exploratory study with adult learners of Spanish will be described.

2.2.3 Spanish diphthongization in adult L2 learners: forced choice data

A preliminary study was conducted of the behavior of 45 English speaking learners in 2nd through 5th semester Spanish classes at a large US university (Carlson, 2006). The subpattern of the diphthong/mid-vowel alternation in which the likelihood of a stem diphthong is modulated by a series of derivational suffixes was selected, following Eddington (1996; 1998). This pattern was selected because while learners in a university context are routinely taught that diphthongization occurs regularly in alternating verb paradigms, the behavior of alternating stems with derivational suffixes is rarely mentioned. Learner behavior on nonce verbs may be affected by this kind of pedagogical intervention, but if learners are sensitive to the subtle patterns of diphthongization in other derived words, this should appear in their judgments of neologisms. Participant judgments were collected using a questionnaire modeled after that developed by

Eddington (1996; 1998). Both forms of each neologism were given simultaneously, but participants were asked to rate each alternative separately for likelihood of correctness.

Figure 2-1 shows the results of this exploratory study. The prediction for native speakers based on Eddington's results is that ratings for neologisms with monophthongs will be higher for those with suffixes that tend to cooccur with monophthongs in the corpus (shown to the left of the figure), and the opposite will be the case for suffixes with the opposite tendency (to the right), with more ambivalent ratings in the middle.

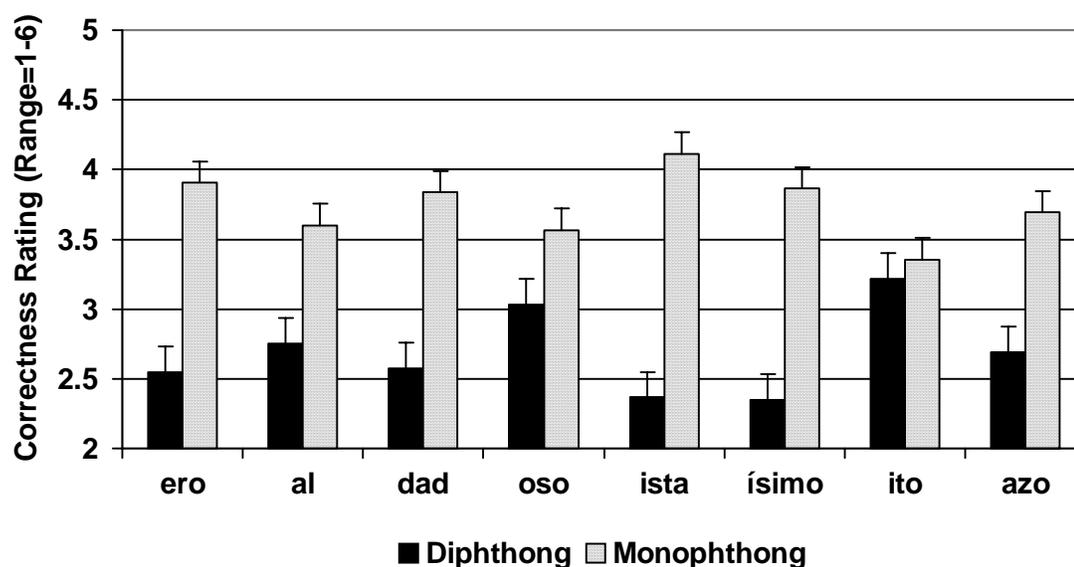


Figure 2-1: Mean correctness ratings for neologisms with unstressed diphthongs or monophthongs, by suffix (Carlson, 2006)

These data show that beginning to intermediate learners rate neologisms with monophthongs consistently higher than those with diphthongs, consistent with the general dispreference for unstressed diphthongs in Spanish. This suggests that they have acquired this general phonotactic tendency in Spanish. However, learners were significantly more ambivalent towards the presence of diphthongs for the suffixes *oso*

and *ito* than they were for *ista*, for which they most strongly preferred monophthongs. These results suggest that a lexically bound pattern is beginning to emerge, although it differs from that shown for native speakers. While a shift in preferences towards diphthongs was found for *ito*, neither the ambivalence found for *oso* nor the markedly higher monophthong ratings for *ista* were predicted based on Eddington's results.

The results of this exploratory study are promising, supporting the hypothesis that adult learners of Spanish apply morphophonological alternations gradiently depending on specific morphological context. They suggest that beginning to intermediate learners do treat some suffixes differently from others with respect to diphthongization, but that they have not attained a nativelike pattern of preferences. The research undertaken in this dissertation with more advanced learners seeks to reveal that adult learners more closely approximate the native pattern of diphthong bias as they gain experience.

2.3 The present experiments: Research questions and predictions

The present experiments were designed to assess the extent to which the gradient patterns of diphthongization in Spanish derived words, as described above, impact adult learners' knowledge and processing of Spanish morphophonology. In other words, it asked whether adults are able to learn the varying probabilities of alternate forms of diphthongizing stems in the context of specific derivational suffixes.

This work also sought to extend prior work on this pattern in two ways. First, it examined learners from a wider range and higher levels of proficiency, from intermediate through advanced. Second, it employed sensitive psycholinguistic tools to measure not

only judgments about neologisms, but also the impact of the specific structure of a neologism on lexical processing as measured in RTs.

This methodological innovation was made in order to determine what lexical information comes into play during the processing of alternating stems in previously unencountered derivations. The native speaker data reviewed above suggested that speakers rely on, among other things, the knowledge that a stem alternates (which itself has been shown to depend possibly on segmental patterns across diphthongizing stems) and knowledge of how other alternating stems behave with specific suffixes. The morphophonological structure examined here is predicted to impact both processing time, reflected in reaction times (RTs) and preferences for diphthongs or monophthongs in a given word.

This facet of Spanish diphthongization was chosen for a number of reasons. Most importantly, diphthongization follows a relatively consistent general pattern based on the location of stress, but that pattern varies probabilistically based on distinct morphological contexts. Those contexts (i.e. suffixes) are fairly transparent to learners once the morphemes themselves have been acquired, making the subpatterns of diphthongization at least available for learning. Given this availability, it is possible that learners will either learn the pattern veridically, or that they will tend to regularize it according to the most frequent pattern (e.g. Hudson Kam & Newport, 2005; Wonnacott & Newport, 2005). The behavior of diphthongizing stems in nonverb derivations is a good choice for the further reason that, as mentioned above, it is not generally taught explicitly. Learners' processing and behavior is therefore less likely to be influenced by pedagogical design, as would be likely in a nonce verb study.

The predictions for the current study may be stated in terms of the wordlikeness, or well-formedness of the neologisms based on the bias of suffixes toward monophthongs or diphthongs when used with alternating stems. A neologism is more wordlike if its diphthong status matches the direction and strength of its suffix bias. With native speakers, the wordlikeness of neologisms is expected to affect processing in a similar way to neighborhood density (e.g. Andrews, 1997; Carreiras et al., 1997; Dell & Gordon, 2003; Perea et al., 2005; Vitevitch & Luce, 1998, 1999; Vitevitch et al., 1999). In the same way that orthographic or phonological neighbors become active during the processing of words and nonwords, morphological relatives of the stem, and other forms with the same suffix may also be activated and affect the speed or efficiency of processing and the likelihood of errors (e.g. Baayen & Schreuder, 2003).

For the learners, the question is a bit more complex due in large part to differences in language experience. Classroom-based foreign language instruction may not provide the quantity of experience necessary for frequency-based patterns to emerge robustly. On the other hand, as learners become more and more advanced and are exposed to more and different kinds of Spanish usage, their experience should provide increasingly strong evidence for more nativelike patterns. Study abroad is particularly likely to provide language experience of the kind that can robustly undergird gradient morphophonological patterns (e.g. Collentine, 2004; Díaz-Campos, 2004). This study therefore included learners with both classroom and study abroad experience in order to ascertain whether the differential access to Spanish language experiences modulated their sensitivity to gradient in diphthongization.

In addition to experiential factors, the individual difference measures reviewed in Section 1.5.2 are also likely to modulate both the accessibility of certain patterns to learners and how the learners structure their language experience as they acquire L2 grammar. Different kinds of memory may allow them to attend to and integrate different levels of structure and different patterns of similarity.

The central hypothesis of the current research is that adult L2 learners are capable of learning fine-grained probabilistic aspects of their L2 grammar. This is based on the emergentist and statistical learning research reviewed earlier, and in particular on work indicating that adults are able to learn to segment speech and extract the rules of artificial grammars based on statistical relationships between segments, syllables, lexical items, and even nonlinguistic tones (e.g. Creel et al., 2004; Onishi et al., 2002; Saffran et al., 1999; Saffran, Newport et al., 1996; Thompson & Newport, 2007). There is even some indication that adults may learn gradient patterns such as this even more veridically than native speakers, who may have been more prone to regularize structures learned as children (Hudson Kam & Newport, 2005). Finally, the effects of this knowledge on rapid and early stages of lexical processing are likely to be visible, given evidence that bilinguals are more prone to decompose frequent morphologically complex words than are monolinguals (Lehtonen & Laine, 2003). It may also be the case that gradient effects will be visible in the some aspects of lexical processing, such as RTs, before they emerge in learners' behavior (McLaughlin et al., 2004).

This investigation of L2 learning thus relies on a complex morphophonological pattern that should be available to learners as their experience increases. It adopts a methodology capable of detecting the subtle consequences of lexical probabilities early in

lexical processing, and it applies this methodology to learners with a range of proficiency in Spanish, from intermediate to advanced. It thus attempts to track the emergence of gradience in Spanish diphthongization with a series of derivational suffixes as a function of language experience, proficiency, and cognitive resources. We now turn to a detailed presentation of the methodology used, followed by presentation of the results.

Chapter 3

Methodology

3.1 Overview

The experiments reported in this dissertation consist of three components, as shown in Table 3-1.

Table 3-1: Experimental Tasks

Lexical processing tests:	Auditory Lexical Decision (perception)	Auditory Conditional Naming (production)
Proficiency tests and assessment of language history:	Spanish picture naming	English picture naming
Tests of Cognitive resources:	Language History Questionnaire (experience, self-assessed proficiency, attitudes and motivations)	
	Foreign word repetition (Phonological memory)	Simon task (attentional or inhibitory control)
		Reading Span (executive control, working memory)

The first and primary component was divided into two tasks, testing phonological perception and production examined the central hypothesis about the statistically driven learning of variability in Spanish diphthongization patterns. An auditory lexical decision task (LDT) examined participants' perception of the Spanish diphthongization pattern described in Chapter 2. Similarly, an auditory conditional naming task tested the consequences of statistical variation in diphthongization in participants' language production. Participants were instructed to respond *palabra* 'word' if they heard a real

Spanish word and simply to repeat the nonword tokens. The use of both perception and production tasks was designed not only to provide converging evidence from two kinds of processing but also to reveal how the development of L2 phonological grammar may differ between perception and production. The within-participants design of the present experiment allowed for direct comparison of learners' current state of development in these two domains. Furthermore, perception and production tax cognitive resources differently. It may be that individual differences in cognitive functioning have different effects on language perception and production due to the different demands of these kinds of processing.

This was followed by a picture naming task designed to provide a measure of learners' proficiency in Spanish. The third component comprised three measures of cognitive resources. Together with the proficiency measure, this component was included to test the hypotheses that higher proficiency is associated with the emergence of more stable and native-like patterns and that cognitive abilities modulate learners' sensitivity to statistical patterns in the lexicon. Differences in cognitive resources may facilitate or hinder the construction of a nativelike lexical network by allowing more or less information to be extracted from experience and attended to at a given time. However, due to the wide range of results reported in the literature, it is unclear exactly what memory characteristics and other attentional factors may facilitate SLA as examined here. For this reason, phonological memory, inhibitory control, and executive control were tested separately in order to converge on the learner characteristics most closely associated with the acquisition of probabilistic grammar. These tests also provided a

means of equating the groups at different proficiency levels, thereby minimizing the possibility that other variables might account for differences across proficiency levels.

The remainder of this chapter will be structured as follows. In Sections **3.2.1** through **3.2.3** I describe the perception and production experiments, including a detailed explanation of how the word and pseudoword tokens were selected and constructed. In Section **3.2.4** I outline the picture naming task and LHQ, which were used to assess participants' experience, attitudes, and proficiency in Spanish. Section **3.2.5** explains how the tests of cognitive resources were constructed. Following this description of the experimental tasks, Section **3.3** details the participants who took part in the experiment, and Sections **3.4** and **3.5**, respectively, outline the procedures followed in administering the experiment and in preparing the data for analysis.

3.2 Tests of phonological processing

I now discuss the characteristics common to the perception and production tasks that comprise the main component of this study, reserving the particulars of each task for the following subsections. The lexical processing tasks were chosen for their sensitivity to subtle structural differences in very early stages of language processing. They thus complement the critical findings from previous research, including the pilot study, in which metalinguistic judgments have revealed statistical effects in speakers' conscious intuitions. Psycholinguistic studies indicate that sensitivity to certain linguistic features can be detected before their effects are visible at this later stage of processing (see

Jurafsky, 2003, for a review of probabilistic inquiry in psycholinguistics). For example, event-related potentials revealed a sensitivity in learners to the word/nonword status of items after only 14 hours of instruction in French, although the same learners performed at chance in their lexical decisions (McLaughlin et al., 2004). The psycholinguistic lexical processing tasks used here therefore target sensitivities that may not yet be available for conscious reflection, and also examine more precisely the impact of morphophonological structure on language processing apart from metalinguistic intuitions. The use of these instruments to examine knowledge of statistically variable structures in L2 phonology is also in line with the neighborhood and phonotactic studies described in Section 1.3.3, which have successfully employed similar online tasks (e.g. Andrews, 1997; Carreiras et al., 1997; de Jong et al., 2003; Luce & Large, 2001; Meunier & Segui, 1999; Pallier et al., 2001; Perea et al., 2005; Sebastián-Gallés et al., 2005; Vitevitch & Luce, 1998, 1999; Vitevitch et al., 1999).

This experiment specifically targets subtle structural variation in phonological patterns. Therefore, the central experimental tasks were conducted in the auditory modality in order to most directly measure phonological knowledge. While the effects of orthography may affect language processing even in auditory tasks (Ziegler, Ferrand, & Montant, 2004; Ziegler et al., 2003), this also served to minimize the effects of the incomplete overlap of L1 and L2 sound-symbol correspondences. With respect to the current research, Spanish diphthongs always and only correspond to digraphs, whereas both English diphthongs and monophthongs may be written with one or more letters (e.g. *go* [go^w], *ceiling* [silŋ]). It is thus likely that some learners may not interpret Spanish orthographic conventions properly, which may obscure learners' phonological processing

on written tasks. L2 reading is, of course, an interesting and complex problem in its own right, but a deep investigation of the interaction of L1 phonology and orthography on L2 perception is beyond the scope of this dissertation.

3.2.1 Perception experiment

The LDT was used to investigate gradient L2 patterns in learners' perception of Spanish neologisms containing an alternating diphthong. This task has been used extensively in psycholinguistic research to reveal structural and processing constraints within the lexicon, including the sensitivity of bilinguals and late L2 learners to aspects of the L2 (e.g. Pallier et al., 2001). Furthermore, it has been the most frequent choice in the literature on neighborhood effects (Andrews, 1997) and has yielded the most robust results. Nonetheless, the choice of the LDT for the present project must be carefully considered because, like any particular task, it makes specific demands on processing that do not necessarily line up with the demands of other tasks (e.g. Berent, Shimron, & Vaknin, 2001; Luce & Large, 2001; Vitevitch & Luce, 1998, 1999; Vitevitch et al., 1999).

For instance, increased neighborhood density and probabilistic phonotactics, which are highly correlated (but see Luce & Large, 2001, who attempted to unconfound these variables) have been associated with increased reaction times (RTs) on the LDT. On the other hand, the same-different task, in which participants hear two stimuli and decide whether they were the same or not, showed the reverse effect for pseudowords, and no effect when words and nonwords were mixed (Vitevitch & Luce, 1999). These

results suggest that processing in this task does not necessarily involve accessing lexical representations, but may only rely on lower-level phonotactic generalizations, which facilitate responses to more probable items. The LDT, on the other hand, requires speakers to access lexical representations in order to make accurate decisions. Items with more neighbors are thought to activate more competitors at this level, resulting in the observed inhibitory effect.

Of particular relevance to the current experiment, the requirement to access lexical representations makes the LDT better suited to detecting sensitivity to the statistical patterns of Spanish phonology examined here. In particular, knowledge of gradient diphthongization tendencies in Spanish derived words (nonverbs) depends crucially on the knowledge that a particular stem contains an alternating diphthong, and on sensitivity to variation in the appearance of diphthongs in the presence of particular derivational suffixes. Given this dependence on specific morphemes, the detection of effects is contingent on activation of the lexical representation of the morphemes, and these are unlikely to appear on a task in which, for example, only adjacency and positional probabilities of phonemes come into play.

There are a number of considerations to be made in using the LDT with learners. First, previously unencountered real words are effectively pseudowords for learners. However, there is also evidence that the responses of lower proficiency learners tend to be biased towards “word” (see Kroll, Michael, Tokowicz, & Dufour, 2002, Table 4). Furthermore, since the probability of diphthongization is based on alternating stems with particular suffixes, real stems and affixes must be used to create the test items. While the neologisms created in this way are nonwords in that they are not attested in existing

corpora, the presence of real morphemes in plausible neologisms may increase the natural tendency of learners, and possibly also native speakers, toward false positive responses still further. However, the prediction that more wordlike neologisms, that is, neologisms that more closely match the patterns in the lexicon, will induce longer RTs is matched by the prediction that they will also induce more false positives. Moreover, given that inhibitory effects for higher probability items are also robust for words, RT analyses may also be conducted on positive responses, if these are present in sufficient number, treating novel derivations as words, rather than as pseudowords. At the design stage the extent of the learner bias toward false positives was not known, although pilot trials indicated that there would likely be a substantial number of both positive and negative responses in the lexical decision.

The evidence for inhibitory effects of wordlikeness (taken as the probability of a word's phonotactic shape) in lexical decision led to the prediction that RTs would be slower on the LDT for more probable word structures (Vitevitch & Luce, 1999, and others). That is, responses should be slower for neologisms containing unstressed diphthongs when they contain suffixes that occur more often with unstressed diphthongs in the lexicon (see Section 3.2.3 for how tokens were constructed). Such higher probability items were also predicted to yield more false positives.

3.2.2 Production experiment

An auditory conditional naming task was used to examine whether learners' language production reflects the lexical statistics of Spanish diphthongization in derived

words. Participants listened to a series of real and invented words, as for the LDT. They were instructed to say *palabra* ‘word’ if they were confident that they had heard a real Spanish word, and to repeat the item if they thought they had heard an invented word. This task design was selected because it depends on a lexical decision and thus requires participants to access the lexical-level representations on which the grammatical pattern in question crucially depends. The parallel nature of the LDT and naming tasks enabled a direct comparison of learners’ perceptive and productive performance.

This modified naming task yielded three kinds of data for analysis. First, RTs and error rates were available for analysis as for the production experiment. The initial predictions were the same as for the LDT (but see Dell & Gordon, 2003; Vitevitch & Luce, 1999, for a discussion of the contrasting effects of neighborhood density on perception and production). Further, the recording of participant responses made it possible to analyze the kinds of errors made. For suffixes that highly disfavor unstressed diphthongs in the existing Spanish lexicon, the likelihood that neologisms with diphthongs would be corrected by substituting a monophthong in the critical syllable was expected to increase.

3.2.3 Item construction and selection, LDT and Conditional Naming

This study examines sensitivity to emergent probabilities based on an intricate network of phonological, morphological, and semantic connections between specific lexical items, relying on a small subset of this complex, multidimensional phonological and morphological space. Consequently the test items must be carefully constructed.

Table 3-2 lists the types of stimuli and the counts of each type included in each task. Each participant completed two tasks, an LDT and a conditional naming task, and therefore two parallel lists were constructed. Half the participants performed the LDT on List 1 and the naming task on List 2, and the other half performed the tasks on the opposite lists.

Table 3-2: Stimulus counts for LDT and Conditional Naming tasks

Test items:	Item type	Test list (each task)
	Neologisms (diphthong)	28
	Neologisms (monophthong)	28
Fillers	Nonwords with real suffixes	24
	Nonwords	27
	Spanish nouns	33
	Spanish nonnouns	29
	Suffixed words (no alternating stems)	45
Total		214

The 56 neologisms on each list were formed by concatenating Spanish stems containing alternating diphthongs (e.g. *dormir/duermo*; ‘sleep’, infinitive/first person singular present) with a series of derivational suffixes. Half of the neologisms in the experiment appeared with a monophthong in the alternating syllable, and the other half with a diphthong. In order to control for any possible effects due to the characteristics of particular stems, two versions of each list were constructed as mirror images of each other. These mirror versions of each list contained the same stimuli, except that neologisms appearing with a monophthong in one version appeared with a diphthong in the other. For example, List 1 contained a neologism formed by joining the stem from *abierto* ‘open’ with the suffix *-ista* to form a word with the possible meaning ‘a very open person’. In order to control for the particular characteristics of this stem, one version

of List 1 contained this neologism as *abiertista* and the other version contained it as *abertista*. Distractors formed by concatenating nonsense stems with derivational suffixes (e.g. *timiesazo/timesazo*) were distributed in the same way, but other distractors remained the same across the two mirror versions of each list. List 1 and List 2 themselves, of course, consisted of entirely different sets of stimuli. The different versions of each list were counterbalanced across participants.

The following subsections describe the construction of each type of stimulus in greater detail.

3.2.3.1 Test neologisms

The 112 neologisms (56 per experimental task) were constructed by concatenating existing Spanish stems with a series of derivational suffixes. The list of suffixes comprised seven of the ten examined by Eddington (1996; 1998): *ero*, *al*, *oso*, *ista*, *ísimo*, *ito*, and *azo* (see Table 2-3). The exclusion of these suffixes also reduced the number of test stimuli, ameliorating the possibility of fatigue during the experiment.

The adjectival suffix *dad* was excluded because its presence often triggers the insertion of an epenthetic vowel (e.g. *ansiedad* ‘anxiety’), which might have affected RTs in some way. Two of the diminutive suffixes used by Eddington, *cillo* and *zuelo*, were excluded because they are seldom taught in university language classes and are thus unlikely to be familiar to the participants. Of course, particularly lower-level learners may be unfamiliar with some of the remaining seven suffixes, and might therefore treat words containing these suffixes as monomorphemic. Nevertheless, this should not

adversely affect the experiment, as part of learning a morphologically dependent phonological pattern is learning the morphemes themselves, and this study seeks to determine if and when specific idiosyncrasies in diphthongization emerge in learners.

The selection of stems for use in the neologisms was accomplished by listing all diphthongizing stems above the 70th percentile in the Buscapalabras lexicon (Davis & Perea, 2005). Buscapalabras is a program for calculating various lexical statistics, including, for example, word frequency, bigram frequency, biphone frequency, and orthographic and phonological neighborhoods. The lexicon used by the program is a subset of the LexEsp corpus (Sebastián-Gallés, Martí, Carreiras, & Cuetos, 2000), a corpus of over 5 million words comprising texts from a variety of genres, including novels, interview transcriptions, scientific publications, and magazines taken from Spain and other areas in the Spanish speaking world. The Buscapalabras lexicon is a subset of about 31,000 lexical entries created by filtering the LexEsp corpus by several criteria. Items not appearing in the dictionary of the *Real Academia Española* (Real Academia Española, 1995, electronic edition) were removed, thus excluding most inflectional variants of words, along with foreign words, abbreviations, and so on. Items with a frequency of zero were also excluded, and words shorter than 3 and longer than 12 segments were also excluded.

The resulting lexicon provides a fairly good approximation of the lexicon of an educated native speaker in the form of a lemmatized list. Its use as a starting point in this study is in line with practices in the L2 vocabulary learning literature (e.g. Laufer & Nation, 1999). The frequency cutoff at the 70th percentile corresponded to the top 9479 entries in the Buscapalabras lexicon, or all entries with a LexEsp frequency over 3.03 per

million words. This cutoff was established as a point over which many words are still accessible to university-level learners of Spanish. It corresponds roughly to the suggestion that the 10,000 most frequent words is an appropriate cutoff for university L2 study (Hazenberg & Hulstijn, 1996; Schmitt, Schmitt, & Clapham, 2001), and as such provides a reasonable upper bound for the language experience of learners in this study.

As stated above, diphthongizing stems with frequencies greater than 3.03 per million were extracted for this study, being careful to avoid stems that always appear with a diphthong (e.g. *riesgo* ‘risk’). This list was then reduced to include only those stems appearing in both diphthongized and monophthongized form above the frequency cutoff, to increase the likelihood that learners would have been exposed to both forms. In other words, only stems for which learners were likely to have evidence for an alternating diphthong were selected.

These stems were then crossed with each of the 7 suffixes to form neologisms. Each neologism was created twice, once with a diphthong in the alternating syllable, and once with a monophthong. Thus, for example, *diente* ‘tooth’ was crossed with *oso* to form both *dientoso* and *dentoso*, roughly meaning ‘toothy’. In like manner, it was crossed with the remaining suffixes, and likewise for all of the stems. This list was then compared with the LexEsp corpus in order to identify and remove any real words that were inadvertently created through this process. For example, *dental* ‘dental’ would also have been created, but it already exists in the Spanish lexicon. In such cases, both the real word and its alternate, in this case, *diental*, were deleted. This procedure was also used in the case of homophones accidentally created, as when, for example, the joining of *fiebre* ‘fever’ and *ero* created *febrero* ‘February’. Stems such as *cielo* ‘sky’, which have close

homophones in one of their forms, were also removed. In this case, neologisms such as *celazo* may have been as likely to activate a phonologically related stem such as that in *celoso* ‘jealous’ and the intended stem.

In this way a list of 1334 neologisms was compiled by crossing the 115 stems with the 7 suffixes. However, closer inspection revealed that some of the neologisms created were semantically more plausible than others. To control for the plausibility of neologisms as well as to check that the target stems would be reliably identified, a norming study was conducted with 85 native Spanish speaking participants from the University of Jaén, in Spain. An additional benefit of this norming study was that it served as a second way of detecting neologisms that had in fact already been created and entered common usage. Due to the large number of stimuli, 15 separate lists of 115 stimuli were created and distributed randomly among participants.

Participants were instructed to give two responses to each item. To assess whether the target stem had been perceived accurately, participants were asked to give a possible definition of the word. This was then scored as a binary variable, depending on whether the definition given reflected the meaning of the target stem or not. In order to assess the plausibility of the neologisms, participants were asked to give each item a rating from 1 to 4 based on how likely they thought it would be that someone would invent the word.

To ensure the reliability of results across lists, 14 stimuli (one per suffix with a diphthong, and one with a monophthong) appeared on all 15 lists. This facilitated comparisons between the mean ratings for each list. The plausibility rating showed moderate reliability across participants, Cronbach’s $\alpha=.761$, and the rate of detection of the target stems showed good reliability, Cronbach’s $\alpha=.883$. However, while reliability

was acceptable, some groups gave generally higher ratings than others. Therefore, the plausibility and identification rates were normalized by list, to facilitate comparisons between items on different lists.

The results from this norming study were then used to filter the group of possible experimental stimuli. In order to do this, neologisms with plausibility ratings more than one standard deviation below the mean, and more than two standard deviations above the mean were removed. The same criterion was applied to the accuracy scores for target stem identification. If after this process only one form of a neologism remained (i.e. it only the form with a diphthong or only that with a monophthong), it was also removed.

From the remaining neologisms, eight items per suffix (7) were selected, four with diphthongs and four with monophthongs, for a total of 56 neologisms. The number of front and back diphthongs was balanced. This process was repeated, yielding two lists of 56 experimental stimuli. Participants performed the LDT on one of these lists, and the conditional naming task on the other list, counterbalanced across subjects. As all participants listened to both lists in the course of the experiment, an effort was made to ensure that participants heard a given stem only once in the experiment. However, the aforementioned selection process left too few stems to meet this goal fully. Therefore, 36 stems were repeated from one task to the other, but they appeared with different suffixes in each task. The small number of stems remaining for stimulus construction also made it difficult to control for stimulus length or for its relative frequency in diphthongized and monophthongized form. These variables were recorded for post-hoc analysis.

One further comment regarding stimulus length is relevant here. The suffixes used in creating the neologisms range from one to three syllables in length. This rendered it

difficult to equalize stimulus length across suffixes. This may not affect comparisons within suffixes, but token length may affect performance on the LDT itself, apart from affecting the statistical patterns. Stimulus length will thus be included as a covariate in the analysis to control for its effects.

As stated above, this process yielded two lists of 56 neologisms. Participants performed one experimental task with each of these lists. In addition, two versions of each list were created as mirror images of each other. Neologisms appearing with a monophthong in one version (e.g. *abertista*) appeared with a diphthong in the other (e.g. *abiertista*), and vice versa. These versions were counterbalanced across subjects to ensure that both possible realizations of a given neologism appeared.

3.2.3.2 Distractors

In addition to the 56 neologisms described above, each experimental task included 158 distractor items for a total of 214 items per task. Among the distractors were 107 attested Spanish words and 51 nonce words. Thus, half of the items in each task were real words, and half were not, counting the neologisms as nonwords.

The real word distractors comprised three groups, containing 33 real Spanish nouns per task, 29 Spanish words from other categories, and 45 derived words using the seven experimental suffixes. The first two of these groups contained words two to four syllables in length and were divided equally between high and low frequency words. The low frequency group was defined as those words having frequencies of 10-40 per million

in the LexEsp corpus, and the high frequency group as those having frequencies between 90 and 500 per million.

The nonword distractors comprised two groups. Each task contained 27 “monomorphemic” nonce words not interpretable as having any derivational affixes, and 24 “derived” nonce words containing a nonce stem and a derivational suffix. The nonce words and stems were constructed such that none of them contained a real Spanish word of more than one syllable in length. There were no violations of Spanish phonotactics.

The “derived” nonce words were similar to the test neologisms in that they all contained either a diphthong or a mid-vowel. These were distributed evenly among 12 derivational suffixes (including the seven test suffixes). The inclusion of these stimuli was intended to prevent the presence of a derivational suffix from becoming a cue to lexical status by creating three categories of words with suffixes. This group of nonce ‘derivations’ contained a real suffix but an invented stem, in contrast with the derived word distractors which contained real stems and real suffixes in attested Spanish derived words. Both of these distractor groups contrasted with the neologisms, which consisted of real stems and real suffixes forming possible, but unattested Spanish words. As with the neologisms, two versions of each list were constructed as mirror images.

3.2.3.3 Recording of auditory stimuli

All stimuli for the LDT and conditional naming tasks were recorded by a male native speaker from Valencia, Spain in a soundproof booth. Several sessions were conducted in which the speaker was given training and time to practice, in order that both

the nonwords and words would be produced in a natural way. Each word was recorded three times at a sampling rate of 44100 Hz and then band-pass filtered between 100 and 5500 Hz. One of the three repetitions was then selected in which all segments were clearly produced with no reductions, additions, omissions, or final devoicing of vowels. Extraneous clicks were excised by removing the one or two affected periods from the waveform. The root mean square pressure was then normalized to 0.3254 Pascal (approx. 63.2 dB) in order to equalize the intensity of the stimuli.

Finally, all stimuli were resynthesized so that there would be no intonational cues that could be used to determine whether an item was a real Spanish word or not. This was accomplished by selecting prototype exemplars with canonical intonation contours, extracting the pitch tier using PRAAT software (Boersma & Weenink, 2006), and placing the new pitch tier onto the remaining stimuli. Since Spanish intonation is dependent on a word's length and stress pattern, separate prototypes were selected from among the real word recordings for each stress pattern and syllable length represented in the stimuli. The prototypical intonation contours were then applied to all stimuli of the same length and stress pattern. Because the duration of syllables varied from one item to the next, the temporal dimension of the pitch contour was adjusted to match the timing of syllables in each item. The stimuli were resynthesized with the new pitch contour such that the intonation contour was identical for all items of a given length and stress pattern. Coupled with the training of the speaker, this ensured that the nonword stimuli sounded canonical, and not unusual or forced, which might otherwise provide a cue that could be used in distinguishing between real and nonce words.

3.2.3.4 Procedures for LDT and conditional naming tasks

Participants completed the LDT first, followed by the conditional naming task. This ordering was selected to present the perception task before the production task. As stated above, there were two separate lists of stimuli. Half the participants completed the LDT with one list and the naming with the other, and the lists were reversed for the other half of the participants. There were also two versions of each list, created by changing the test items with diphthongs to monophthongs and vice versa. Each participant heard only one version of each list, and the versions were counterbalanced across participants.

Within each task, stimuli were presented in random order, with the condition that test neologisms would not appear among the first four items. The task proceeded as follows. Participants saw a fixation (+) in the center of the screen for 1000ms, after which they listened to a stimulus through headphones. They were given 5000ms after onset of the stimulus in which to respond. Reaction time was measured from stimulus onset until the participant either pressed a button (LDT) or began a verbal response (conditional naming). On the LDT, participants pressed a button labeled “yes” with the right hand to indicate that they were confident that what they had heard was a real Spanish word, and they pressed a button labeled “no” with the left hand to indicate that they thought it was not a real Spanish word. On the naming task, they responded *palabra* to indicate that a stimulus was a real word, and they were instructed to repeat items they thought were made-up Spanish words as accurately as possible. Following either their response, or 5000ms if no response was given, the fixation disappeared and there was a 1000ms pause, followed by another fixation, beginning the cycle again.

3.2.4 Proficiency measures

Two different tasks were used in order to classify learners according to their Spanish proficiency: a self-assessment and Spanish and English picture-naming tasks. The self-assessment formed part of a language history questionnaire (LHQ, see Appendix 4). Given that the target population in this study represents a range of proficiency levels and may therefore construe questions and standards of comparison differently (Moritz, 1995), the self-assessment utilizes questions that refer to specific activities within their common realm of experience as well as a general, four-skills self-rating.

The picture naming task was used alongside the more subjective self-assessment to provide a more objective measure of proficiency. With performance on the English version of the task as a baseline for each participant, RTs in Spanish picture naming provide a measure of efficiency of lexical access, and non-responses and errors give an indication of vocabulary size differences among the participants. The picture naming task thus yields a two-dimensional measure of participants' proficiency in Spanish.

To construct the picture naming task, two lists of 40 line drawings each were compiled using the UCSD norms (Szekely et al., 2004). Each list contained 20 high and 20 low-frequency nouns. High frequency items were those with log frequencies (\ln) over 3.5 in both languages, and low frequency items had log frequencies less than 2.5 in both languages. Ten items on each list were cognates. These list characteristics allow for finer-grained analysis of picture naming performance based on earlier findings. High frequency items should be named faster and more accurately than low-frequency items, and as

proficiency increases, a cognate facilitation effect should begin to emerge in English (L1) as well as Spanish.

In addition to these criteria, the two lists were compared on a range of measures available in the UCSD norms. These variables were, for both languages, proportion of valid RTs, degree of response agreement between subjects, the number of alternate names given, error rates, RT for target responses, visual complexity, number of objects in the picture, syllable count, grapheme count, and log frequency (ln based on CELEX). The lists did not differ on any of these measures.

Participants performed the picture naming task first in Spanish and then in English, so that all L2 tasks in the experiment would be completed first. The stimulus sets were counterbalanced across subjects, one list of 40 items per language. Pictures were presented in random order and participants were given unlimited time to respond. Participants were instructed to name each picture as quickly and accurately as possible. If they did not know the name for a picture, they were instructed to respond either “no sé” or “I don’t know”, depending on the language of the task. After responding, they advanced to the next item by a keypress. Reaction times were measured between the appearance of the picture and the onset of a response, collected on a voice key. Accuracy was determined by digitally recording the responses.

This lexically-based measure is likely to be closely linked to probabilistic knowledge of exceptional patterns, because patterns of frequency depend crucially on the quality, structure, and detail of the connections between items in the lexicon, whether words and morphemes (Bybee, 2001b), or constructions and formulaic sequences (Langacker, 2000; Tomasello, 2003). The size and structure of the lexicon is thus likely

to exert a major effect on the grammatical knowledge that is possible in that it must be large enough for the pattern in question to be statistically robust (Pierrehumbert, 2003b).

In addition to these measures of proficiency, aspects of learner history, predispositions or attitudes that may affect the learning of Spanish morphophonology were also assessed using the LHQ. Special attention was paid to the nature and duration of any study-abroad experience, because the types of experience available during study abroad were expected to affect the availability of the pattern in question for acquisition. The LHQ also addressed learners' goals (e.g. to sound nativelylike, to be conversant, to pass the course), contexts of use (e.g. interlocutors, social settings, purposes, habits), and attitudes toward Spanish phonology. Elliott (1995) found that self-reported attitudes toward pronunciation and accentedness in Spanish were the most significant factor in predicting pronunciation accuracy. While pronunciation and accentedness are not the same as knowledge of morphophonological grammar, they may affect the perception of morphophonological patterns by, for example, affecting how diphthongs and monophthongs are perceived. Post-hoc analyses may reveal a role for these variables as well.

3.2.5 Cognitive resources

As noted earlier, a variety of differing results have been reported regarding the role of cognitive resources in SLA. Therefore, three experimental tasks were used in order to converge on the abilities that best predict performance in phonology: a foreign word repetition task, a Simon task, and a reading span task.

3.2.5.1 Foreign word repetition

First, a foreign-word repetition task using Korean words was constructed to measure participants' ability to repeat nonwords with unfamiliar phonotactics, allophony, and phonetic category boundaries. Korean was selected because of its accessibility to the researcher, because its non-Indo-European phonotactics are clearly different from both Spanish and English, and because it was likely to be completely unfamiliar to participants.

Several characteristics of Korean make it a good choice for this task. Among the phonetic categories in Korean that are not present in Spanish or English are the unrounded back vowels, particularly the high back unrounded vowel /ɯ/. Korean also makes a three-way distinction between aspirated, lax, and tense stops, (e.g. /p^h, p, b/), where Spanish and English each distinguish only voiced and voiceless stops (/p, b/), albeit with different VOT boundaries. There are also several notable differences in articulation. For example, adjacent vowels in Korean are generally pronounced with hiatus, where hiatus is generally lacking in English, and occurs only in certain contexts in Spanish. There is also evidence that Australian English speakers have difficulty learning to pronounce intervocalic liquids in Korean (Kim & Park, 1995). Palatals in Korean are stops, whereas English palatals are affricates or fricatives, and alveolar-dental consonants are laminar, rather than apical. All of these differences may affect individuals' abilities to form accurate representations of foreign words in short-term phonological memory, and consequently to form accurate long-term memory representations that might contribute to fine-grained knowledge of grammar.

Foreign-word repetition tasks such as this have been used to examine vocabulary learning among certain populations (e.g. Service, 1992; Service & Craik, 1993) and to evaluate bilinguals at different proficiency levels in an fMRI study (Chee, Soon, Lee, & Pallier, 2004). If the emergence of detailed probabilistic knowledge of L2 structures depends on vocabulary learning, then the stability of foreign phonotactic shapes in a learner's memory should also be connected to the development of probabilistic knowledge of lexical distributions, especially in the case of phonological patterns. Words with foreign phonotactics and category boundaries were used here because nonword repetition scores for words rated low in wordlikeness have been found to correlate more highly with phonological memory, and scores for more wordlike nonwords are associated more closely with long-term memory (Gathercole, 1995). The use of this type of stimuli targets the extent to which short-term phonological memory is linked to the development of L2 phonological knowledge.

The stimuli for this task were compiled by selecting one three-syllable and one four-syllable word from each page of a Romanized dictionary of Korean designed for use by English speaking learners (Lee, Fouser, & Baxter, 1995). This yielded a list of 97 items which were recorded in a soundproof booth by a female native speaker of Korean from Seoul. Recordings were made at a sampling rate of 44100 Hz and band-pass filtered between 100 and 6000 Hz. The upper limit was higher than that set for the Spanish stimuli discussed in Section 3.2.3.3 in order to capture the more complex differences between fricatives, affricates, and stops in Korean. Each word was recorded 5 times, with the most canonical and complete repetition being selected, avoiding utterance-final pitch drops and reductions of segments or syllables. The root mean square pressure was

normalized to .3254 Pascal (approx. 63.2 dB), as for the Spanish stimuli, and clicks were removed by excising one or two complete periods from the wave form.

The list of 97 recorded stimuli was then reduced to 50 by excluding items that resembled English or Spanish words and by selecting a set of test items that were balanced in length (25 three-syllable and 25 four-syllable words), phonemic content, initial onsets, syllable structures, and final syllables. The test contained 50 stimuli in keeping with several studies using nonword repetition paradigms (Gathercole, Willis, Baddeley, & Emslie, 1994; Michas & Henry, 1994; Service & Craik, 1993; Service & Kohonen, 1995), although studies with fewer stimuli have also been conducted (Ellis & Beaton, 1993).

The procedure in this task was the same as that for the LDT and naming tasks, except that participants simply repeated each item as quickly and accurately as possible.

3.2.5.2 Simon task

A version of the Simon task (Simon & Ruddell, 1967) was used to measure attentional or inhibitory control. This task required participants to press one of two buttons corresponding to which color square appeared on a screen. The position of the square, however, varies from one trial to the next, such that sometimes it appeared on the side corresponding to the button for its color, and sometimes its position conflicted with that of the response button. Participants had to ignore the relatively strong spatial cue in order to respond correctly.

The Simon task has been used to demonstrate the positive cognitive consequences of bilingualism, a result hypothesized to be linked to the greater cognitive demands of managing more than one language (Bialystok, Craik, Klein, & Viswanathan, 2004). Arguably, learners in this experiment may also benefit from increased inhibitory control as measured on this task, because they must inhibit L1 phonological information in order to incorporate the L2 patterns into their developing grammars. Moreover, a nonlinguistic task such as this measures general cognitive capacities, which from an emergentist perspective are taken to form the foundation of language learning, both first and second (Ellis, 2002a; Langacker, 2000; Tomasello, 2003). A relationship between performance on this task and L2 phonological development would provide evidence to support this position. Conversely, the lack of a relationship would provide evidence that the development of a frequency-based phonological grammar may depend more on specifically phonetic and phonological abilities than on general attentional or inhibitory control. This task may therefore provide a useful metric on which to group participants.

For this task, learners were instructed to position the index finger of each hand over the left- or rightmost buttons on a button box. On each trial they saw either a blue or a red square measuring a little less than one degree on each side, and were instructed to press the rightmost button (with the right index finger) if the square was red, and the leftmost button (with the left index finger) if the square was blue. The experiment consisted of a practice block consisting of 24 trials and three experimental blocks of 42 trials each for a total of 126 trials. Each experimental block contained seven trials in each condition, i.e. red or blue in the center, 2° to the right, or 2° to the left. Trials were presented in random order. If a participant made an error, the word ERROR appeared on

the screen for 1500ms, to encourage accuracy. Both accuracy and RT data were recorded for each trial.

3.2.5.3 Reading span task

The reading span (Miyake & Friedman, 1998) has been extensively used in the psycholinguistic literature to measure simultaneous processing and storage capacity under a relatively complex load of reading, comprehending, and the maintenance of items in memory. In it, participants read sentences and judge their plausibility. After each set of two to six sentences, they are asked to recall the final word of each sentence. Differences in learners' capacity for this kind of processing may modulate their learning of phonology given the demands of phonological perception and comprehension of lexical items, discourse, and managing communicative situations in the L2.

For this reason, the present experiment employed a reading span task following Waters & Caplan (1996). This task was administered in English, as a Spanish version would likely have been too difficult for all but the most advanced learners. A Spanish version of the task might therefore have yielded results that reflected Spanish proficiency more than working memory. As the goal was to measure general working memory capacity, English was chosen as the language of the reading span task.

The task was divided into three practice blocks followed by 20 experimental blocks. The blocks contained from two to six sentences, with four experimental blocks of each size, presented in pseudorandomized order such that no two consecutive blocks contained the same number of sentences.

Participants read each sentence and then responded by pressing a button marked “yes” (right hand) if the sentence made sense, and “no” (left hand) if it did not. The sentences disappeared after five seconds, but participants had unlimited time to respond. At the end of each block of sentences the word RECALL appeared in the center of the screen, at which point participants were instructed to type the last word of each sentence in the block, in order.

3.3 Participants

A total of 58 participants took part in this study. The central group included 40 learners from intermediate through advanced levels of proficiency at Penn State University. This provided a cross-section of proficiency levels such that the developmental course of probabilistic knowledge of subpatterns in Spanish diphthongization could be observed. Native-speakers of languages other than English (3), heritage speakers of Spanish (2), and one participant with significant hearing loss were excluded, as well as one participant who had spent 12 years in a Spanish speaking country. No participants were proficient speakers of other Romance languages. Eighteen of the remaining participants reported study abroad experience, ranging from 7 weeks to one year.

In addition, 20 native speakers of Spanish with minimal English proficiency from the University of Granada, Spain, were tested to replicate the earlier results with the present methodology (see Section 2.2.3) and to establish baseline native speaker performance on the experimental tasks. One participant was excluded, because while she

had been educated almost entirely in Spanish from the age of 5 and self-rated her Spanish at the maximum end of the scale on all measures, she had lived in Spain for only three years and her native language was Cherja.

All participants in the learner group signed consent forms and were paid for their participation. Participants in the Spanish native speaker group also signed consent forms, but received course credit for their participation.

3.4 Procedure

Data were collected in individual one-hour sessions. All tasks were presented on a computer using E-Prime software (Schneider, Eschman, & Zuccolotto, 2002) for all but the LHQ. Auditory stimuli were played through headphones at a comfortable listening level. Responses were collected either via keypress on a button box or, for the oral tasks, via a microphone connected to a voice key. Oral responses were also recorded digitally using a headworn microphone. In the reading span task, responses regarding the plausibility of the sentences were recorded through a button box, and the words were entered on the computer keyboard.

After signing consent forms, participants in the learner group completed the LDT followed by the conditional naming task. They then completed the Spanish and English blocks of the picture naming task in that order. This was done in order to present the critical tests of morphophonological grammar first, and to group all of the tasks in Spanish together before the tasks in English or Korean. Following the picture naming tasks, participants completed the Korean word repetition task, the Simon, and the reading

span tasks. The LHQ was completed last, after which participants were paid (learner group) and given an opportunity to have any questions about the study answered.

The Spanish native speaker group followed the same procedure, except that they did not complete the English picture naming task or the reading span, and the LHQ was altered to collect data about participants' experiences learning English and any immersion experience they may have had in order to detect and exclude any participants with substantial experience in English. The detailed self-assessment of proficiency and questions about motivations and attitudes towards foreign language study were omitted for the Spanish native speaker group. The instructions for all tasks, as well as the LHQ, were translated by the Author, a near-native speaker of Spanish, and verified by a native speaker of Spanish.

3.5 Data coding and analysis

For the conditional naming task, response accuracy was determined by comparing the test stimuli with the digital sound recordings of the task. Correct responses (*palabra* for words, accurate repetition of the item for nonwords) were distinguished from incorrect responses and errors in repetition. Segmental errors (i.e. substitution, omission, or addition of one or more phonemes) were transcribed using standard Spanish orthography. For the nonword distractors these were not counted as critical errors, but neologisms and real word distractors reproduced with segmental errors were excluded because such errors result in a different word or stem from the target being produced. This might indicate that the item was not heard correctly, or that the lexical item was not

activated. Finally, errors in which the diphthong or monophthong in the test neologisms or suffixed nonwords was reversed (e.g. *conciertito* was produced in response to *concertito*, or vice versa) were given a unique coding and analyzed separately.

Data from the LDT and naming tasks were then cleaned separately according to the following criteria. Absolute outliers were defined as those items with reaction times less than 300ms or greater than 3000ms for native Spanish speakers, and less than 300ms or greater than 3500ms for learners. Items with reaction times smaller or greater than these values were excluded. The remaining distribution of RTs was positively skewed ($p < .001$), and therefore the RTs were log (base 10) transformed to conform to a normal distribution. Log RTs were then normalized by length in syllables using z-scores, to allow for the comparison of items of different length. Mean z-scores were then calculated for each subject, separating words and nonwords. Z-scores greater than 2.5 standard deviations above or below the means for each subject were excluded from analysis. The log RTs corresponding to the excluded z-scores were also removed, so that stimulus duration and length in syllables could be included as covariates in an ANCOVA based on the log transformed RTs. These analyses on the transformed RTs provide a finer-grained assessment of the influence of stimulus length than would the comparison based on z-scores. However, the z-scores are useful for identifying anomalous trials that are likely to represent microphone errors or other events not related to the experiment itself.

The picture naming tasks were coded in a similar way. Based on the sound recordings of each participant, responses were coded as correct if the target name was produced, incorrect but semantically appropriate if another, related word was produced, and incorrect if either no word or a different, unrelated word was produced. For the RT

analyses, trials with RTs less than 300ms or greater than 3000ms were excluded as absolute outliers. Then, mean RTs were calculated for each subject in each language, and trials with RTs greater than 2.5 standard deviations above or below the means for each subject were also excluded.

Coding for the conditional and picture naming tasks was performed by the Author, a near-native speaker of Spanish, and by a native speaker of Spanish. The data for 5 participants were coded by both raters and checked for interrater reliability. For the conditional naming task, Cronbach's $\alpha=.828$, and for the picture naming task, Cronbach's $\alpha=.816$. Both values are within the acceptable range.

The results from the foreign word repetition task were scored by three native speakers of Korean. A repetition was counted as correct if there were no obvious substitutions, omissions, or additions of phonemes, and the scores were measured as the percentage of words repeated accurately. Nonresponses were counted as incorrect, but responses where the voice key was tripped early by an extraneous sound, cutting off the stimulus, were excluded. The raters also provided a global rating of the participants' accentedness, as a way of measuring their ability to mimic accent in a foreign language. Raters used a scale of 1-6, where one corresponded to a grossly inaccurate accent causing significant problems for comprehension, and six corresponded to a near-native accent. In order to assess interrater reliability, the data for 10 participants were scored by each of the three raters. Cronbach's α for repetition accuracy (number of words correctly repeated) was .965, indicating a high degree of reliability, and for the global accentedness rating it was .784, indicating an acceptable degree of reliability. In order to determine the scores for the 10 participants scored by all raters, the middle score of the three raters was

selected or, in cases where two raters agreed on the same score, the agreed upon score was used.

Scoring for the Simon task was carried out by first excluding all trials on which an error was made, all trials immediately following an error, and all trials with $RT > 1500$ ms. The mean RT was then calculated for each participant for the congruent and incongruent trials (i.e. trials for which the spatial cue corresponded to the hand used for response and trials for which it conflicted, respectively). The “Simon effect”, or the difference between these means, $RT(\text{incongruent}) - RT(\text{congruent})$, was used as a between-subjects variable to assess its influence on participants’ sensitivity to the lexical subpatterns in Spanish diphthongization.

The reading span task was scored by counting the number of words correctly recalled throughout the experiment (Friedman & Miyake, 2005). Finally, based on responses to the LHQ, the number of months spent in study abroad was calculated, and a self-assessed proficiency score was calculated by summing participants’ self ratings on the parameters included on the questionnaire.

Chapter 4

Results: Native Speakers

4.1 Chapter overview

This chapter presents and discusses the results of the Spanish diphthongization experiments for the native speaker control group. Recall that there were two components to the experiment, one examining native speaker perception of gradient patterns in the wordlikeness of derivations with unstressed diphthongs, depending on the particular derivational suffix involved, and the other examining native speaker production of the same patterns. The perception experiment utilized a lexical decision task, and the production task utilized a conditional naming task in which participants were instructed to name only the nonwords. This latter task included a perceptual component in that participants first had to make a decision regarding the lexical status of an auditory token, and then produce the appropriate response based on that decision. The results of each of these experiments are reported separately and subsequently compared in the discussion.

Each experiment yielded two kinds of data, reaction times and error rates (henceforth, false alarms, in which neologisms were mistakenly taken by participants to be real words). The conditional naming experiment includes the further possibility that speakers may reverse the diphthong status of the test neologisms by replacing diphthongs with monophthongs and vice versa. These three types of data provide different ways of evaluating the hypothesis that Spanish speakers are sensitive to the lexical probability of

an unstressed diphthong with a given suffix. To restate the hypothesis in greater detail, the wordlikeness, or well-formedness of neologisms containing an alternating diphthong in the stem will vary depending on the behavior of existing Spanish derivations involving alternating stems. Neologisms containing an unstressed diphthong are predicted to be more wordlike when they involve a suffix that occurs more frequently with unstressed diphthongs in the existing Spanish lexicon, and less wordlike when they involve suffixes that co-occur infrequently with unstressed diphthongs. Neologisms containing a monophthong in the alternating syllable are predicted to show the converse pattern of wordlikeness. Furthermore, this variation in the acceptability of neologisms is predicted to be probabilistic and gradient, being more pronounced for suffixes that show a more pronounced tendency in the existing lexicon, and less pronounced for suffixes that show a less pronounced tendency (Eddington, 1996, 1998).

In order to illustrate this prediction and facilitate comparison with earlier work, the results of Eddington's (1998) forced choice task, for the suffixes used in the present study, are reproduced here in Figure 4-1. The suffixes are presented, here and in all subsequent Figures, ranked in the order obtained in Eddington's results. To the left are the suffixes for which his participants most frequently selected the variant with a monophthong, following the traditional rule that monophthongs surface when stress is removed from the alternating syllable.⁵ To the right appear those suffixes for which participants most frequently selected the variant with a diphthong. Note that these data show a decreasing bias towards monophthongs beginning with *ero*, and an increasing

⁵ Recall that all of the suffixes used in this experiment remove stress from the stem.

bias towards diphthongs at the opposite end of the ranking. This result closely reflects the pattern shown in the most familiar derivations in Eddington's corpus (see Section 2.2.3).

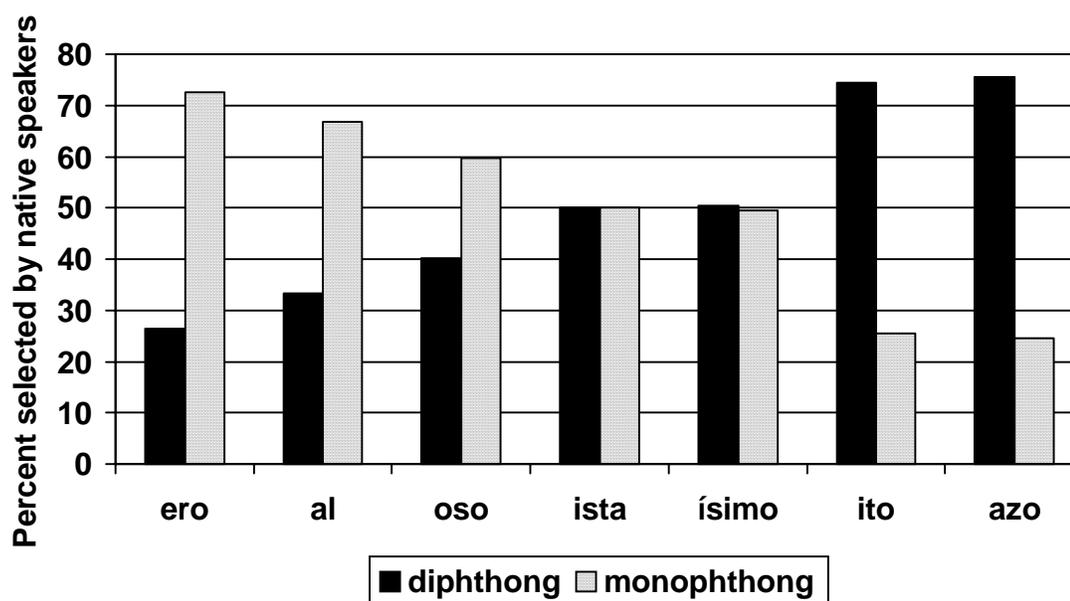


Figure 4-1: Native speaker selection of neologisms with and without diphthongs from Eddington (1998) for the suffixes in the present study

The present methodology contrasts with the forced choice paradigm employed by Eddington in several important ways. The current task required participants to listen to a single item, decide whether it was a real Spanish word or not, and respond accordingly. The test neologisms were all invented, but possible Spanish words based on existing Spanish morphemes. As stated above, this yielded two kinds of results, false alarm rates and RTs, as well as errors in the production task in which the diphthong status of the item was reversed.

The predictions made for each type of data are as follows. First, neologisms that are more wordlike should induce more false alarms than less wordlike neologisms.

Second, following the results of Vitevitch and colleagues regarding neighborhood and

word probability effects in lexical decision (see Section 1.3.3), more wordlike neologisms should be associated with longer decision latencies reflecting the increased competition between related forms. Finally, in cases where either monophthongs or diphthongs are distinctly preferred over the alternative, less wordlike nonwords may be “corrected” in production. That is, if a form such as *duermero* is highly dispreferred (as predicted by Eddington’s results), then it may be produced as *dormero* in the conditional naming task. This may happen if, in the interval between hearing the item and producing it, participants rely on their knowledge of the Spanish lexicon, including possible forms of the target stem and the behavior of the suffix (Gathercole, 1995).

The following sections present the results of first the LDT and then the conditional naming tasks. Within each section, the false alarm rates are presented first, followed by the RT results. The subsequent sections discuss these results and their implications for native speakers’ knowledge and processing of Spanish.

4.2 Results: Lexical decision task

4.2.1 False alarm rates: LDT

There was a significant interaction between suffix and diphthong status in the LDT analysis by subjects, $F(3.231, 54.919)=3.643, p<.05$. This interaction was also

significant in the analysis by items, $F(6, 30)=3.426, p<.05$.⁶ Neither of the main effects for suffix or diphthong status were significant. The results are shown in Figure 4-2.⁷

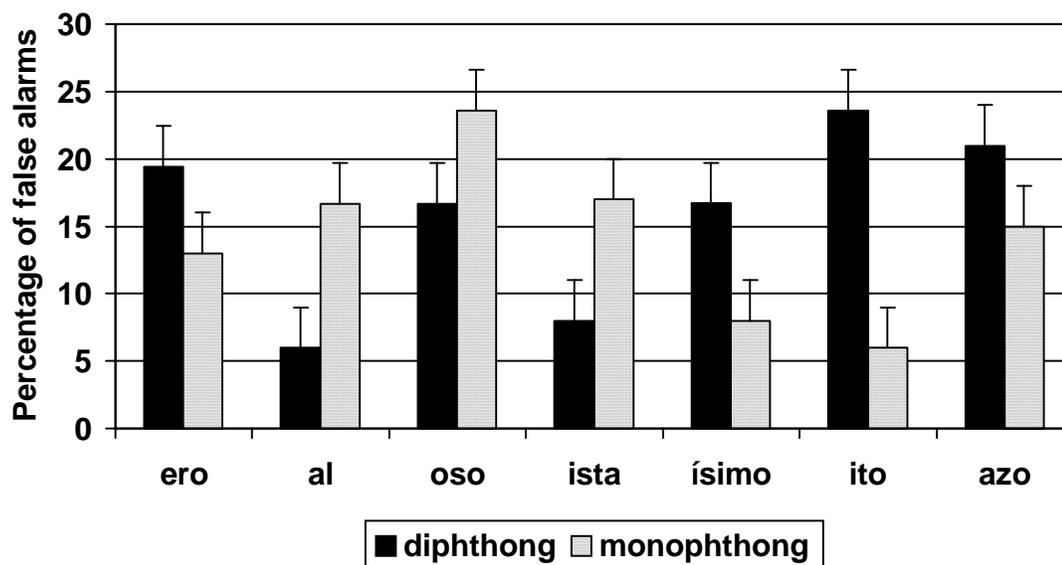


Figure 4-2: False alarm rates for neologisms in LDT by native speakers

Recall that a higher false alarm rate is predicted to reflect greater wordlikeness.

These results suggest that neologisms ending in *al*, *oso*, and *ista* are more wordlike when the stem appears with a monophthong than when it appears with a diphthong. For the remaining suffixes, *ero*, *ísimo*, *ito*, and *azo*, this pattern appears to be reversed.

Furthermore, the tendency within some suffixes appears to be more ambivalent than within other suffixes. This trend is roughly in line with the Eddington's results shown in Figure 4-1, although *ista* and *ísimo* appear less ambivalent in the current results. The present results also depart somewhat from the gradient and progressive changes in bias in

⁶ Repeated measures ANOVA were used in the analyses by subjects, and ANOVA were used in the items analyses

⁷ Error bars in all graphs represent the standard error of the mean by diphthong status.

the ordering of suffixes shown in Figure 4-1. The only marked deviation from the prior results occurs with *ero*, in which neologisms with diphthongs appear to be at least, if not more, wordlike than those with monophthongs.

4.2.2 Reaction time analyses: LDT

The RT analyses, based on the logRT,⁸ show a similar trend to that shown in the false alarm rates, but while the main effect of suffix was significant in both the subject and item analyses $F(6, 102)=6.276, p<.001$ in the analysis by subjects, $F(6, 205)=2.296, p<.05$ in the analysis by items, the crucial interaction between suffix and diphthong status was not.⁹ However, the comments of some participants following the experiment suggested that the target stem may not have been recognized accurately in neologisms with monosyllabic stems. For example, the stem from *cerrar* ‘to close’ may not have been recognized in the neologism *cerrazo*. Since diphthongization in Spanish depends crucially on the knowledge that specific stems alternate under certain conditions, RTs are unlikely to reflect biases dependent on that knowledge if the stem is not recognized.

For this reason, the analyses were run a second time, excluding the neologisms with monosyllabic stems ($n=29$). This time, the interaction between suffix and diphthong status was significant, both in the analysis by items $F(6, 88)=3.054, p<.05$, and in that by

⁸ Recall that the RT data were significantly and positively skewed. The analyses of RTs are therefore based on the log (base 10) RTs (see Section 3.5). The log RTs are the values given in this chapter. The RT figures are reproduced in Appendix 5 based on mean RTs in milliseconds

⁹ For obvious reasons, the length of a stimulus has an effect on RTs, especially in an auditory task such as this. Since the neologisms by necessity varied in length, the duration of stimuli (in milliseconds) was entered as a covariate in all RT analyses by items.

subjects $F(3.089, 27.797)=3.901, p<.05$. However, the exclusion of neologisms with monosyllabic stems in the subjects analysis resulted in empty cells for the suffixes *al* and *azo*, such that too many cases were excluded from the analysis. The analysis by subjects reported here therefore included only the remaining five suffixes. This did not affect the items analysis in the same way, because, due to the counterbalancing of stimulus sets, enough stimuli remained with multisyllabic stems with *al* and *azo* to include these in the analysis. The results are given in Figure 4-3

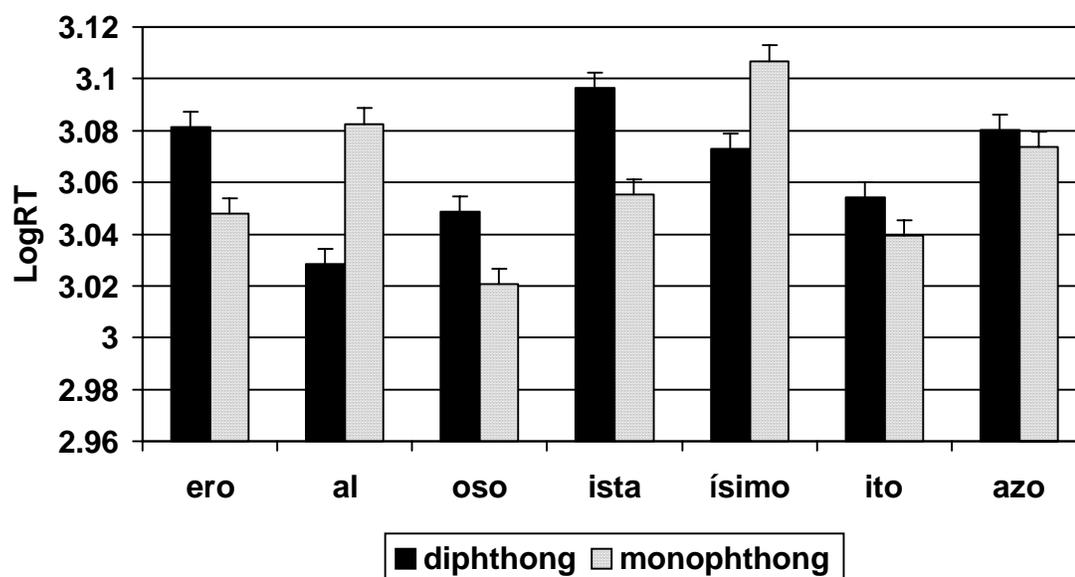


Figure 4-3: Log RTs for neologisms with multisyllabic stems in LDT by native speakers, analysis by items

These results differ more markedly from the results of Eddington. Note first that three of the suffixes at the end of the ranking where a preference for monophthongs was expected, *ero*, *oso*, and *ista*, show longer latencies for their diphthongized forms than for the variant with a monophthong. The remaining suffix on this side, *al*, shows the reverse pattern, but these means are based on fewer stimuli than those for the other suffixes. The

prediction made above was that longer latencies would occur for more wordlike stimuli. However, it seems unlikely that RT and false alarm data would yield opposite results, especially given the similar distribution of trends between suffixes in the false alarm data and the fact that the same trend (albeit nonsignificant) is observed in the RT analysis of all neologisms. Instead, this suggests the possibility that the predictions of inhibition were incorrect. The prediction of an inhibitory effect of increased wordlikeness was based on results related to neighborhood density and phonotactic probability in lexical decision tasks. It is possible that the morphologically complex nature of these stimuli, combined with their reliance on real Spanish morphemes, may instead produce facilitation in lexical decision.

A second observation is that the RT results for *ito* and *azo* show ambivalence regarding the diphthong status of the neologism, whereas prior data suggest that the diphthongized variant should be more wordlike. It should be noted that no monophthong bias is observed in these suffixes either.

The analysis of false alarm rates was also performed excluding neologisms with monosyllabic stems. However, while the trends are similar for some suffixes, the overall interaction between suffix and diphthong status was not significant, suggesting no significant bias towards either monophthongs or diphthongs for the remaining stimuli. It may be the case that the reduction in the number of stimuli renders an effect more difficult to detect. It is also possible that the contrasting results of the RT and false alarm analyses for the complete set of test items vs. only those with multisyllabic stems may reflect the sensitivity of these measures to different aspects or stages in lexical

processing. Possible explanations for these discrepancies will be discussed below in Section 4.4.

4.3 Results: Conditional naming task

4.3.1 False alarm rates: Conditional naming

The conditional naming paradigm, in which participants were instructed to respond *palabra* if they heard a real Spanish word and to repeat the item if they heard an invented word, allows for several kinds of errors to be recorded in addition to RTs. As with the LDT, both false negatives and false positives are possible, the false positives being the relevant type of error for the test neologisms. However, in cases where the item was repeated, being correctly identified as a nonword, the digital recordings were also analyzed to determine other types of error that may have occurred.

There are three possibilities in this experiment. On the one hand, the item could be repeated accurately. If it is not repeated accurately, two categories of error exist in this experiment. First, a word may have been reproduced with a change to its segmental structure. This poses no difficulty for the examination of nonce words, but if a real word or neologism was repeated incorrectly, this might indicate that the word was misperceived. In such cases, the item was excluded from both the error and RT analyses, because it is not clear whether the target stem or suffix, both of which are crucial to the hypothesis being tested, was actually recognized correctly. Second, a special case of segmental error on the test neologisms is the possibility that a diphthong may have been

reproduced as a monophthong, or vice versa. This type of error would suggest that recall of the neologism for production may have been affected by grammatical features of Spanish. Consequently, corrections in the direction of greater wordlikeness would be expected.

The false alarm rates for the conditional naming task were somewhat different from those for the LDT. There was a significant main effect for suffix in both the analysis by subjects $F(4.074, 69.258)=4.921, p<.05$, and by items $F(6, 209)=4.05, p=.001$. There was also a significant interaction between suffix and diphthong status in the analysis by subjects $F(5.652, 96.086)=2.423, p<.05$, but although the trend in the analysis by items was similar, it did not reach significance. The results are shown in Figure 4-4.

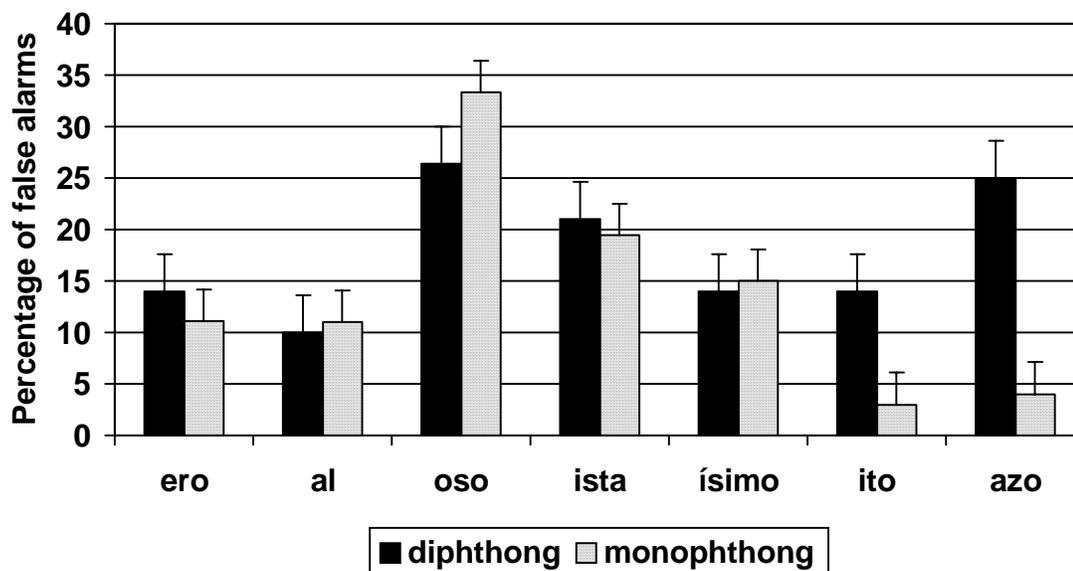


Figure 4-4: False alarm rates for neologisms in conditional naming by native speakers, analysis by subjects

Planned contrasts showed that this effect was due to a higher number of neologisms ending in *ito* and *azo* being misclassified as real Spanish words when they included unstressed diphthongs in the stem than when the stem appeared in monophthongized form. This effect is expected based on prior results if a higher error rate indicates greater wordlikeness. However, the lack of any significant bias towards monophthongs for suffixes at the other end of the ranking (*ero*, *al*, and *oso*) contrasts with the earlier results. Analyses of false alarm rates in the naming data in which neologisms with monosyllabic stems were excluded were not significant.

4.3.1.1 Reversals of diphthong status: Conditional naming

As stated above, the repetition of neologisms involved in the conditional naming task opened the possibility that lexical and grammatical knowledge of Spanish could influence the correct recall of nonwords. If participants relied on activation of lexical representations of stems and suffixes rather than simply maintaining the phonetic shape of a neologism in short term memory, then those lexical representations could lead to a correction of less wordlike neologisms in the direction of lexical biases in Spanish. This would result in less wordlike forms being rendered more wordlike, either by converting diphthongs into monophthongs (for suffixes with a monophthong bias in the lexicon), or vice versa.

Overall, the native speakers were highly accurate in their repetitions of neologisms and nonce words, and the number of reversals of diphthong status was too few to run inferential statistics. Nonetheless, the number of reversals, summed across all

native Spanish speaking participants, is reported in Table 4-1. Reversal errors were made on neologisms with six of the seven suffixes tested. Of this group, errors in which a diphthong was reproduced as a monophthong occurred with the suffixes for which monophthongs are most highly preferred in the lexicon, *ero* and *oso*. Errors in both directions were made with the suffixes *ista* and *ísimo*, which are associated with greater ambivalence in Eddington's study as well as the current LDT results, as well as for *azo*. Finally, there is one example in the native speaker data of a monophthong being reproduced as a diphthong for the suffix *ito*. While the total number of reversals is quite low, their distribution does reflect the predictions based on earlier corpus and experimental data.

Table 4-1: Reversals of diphthong status in conditional naming by native speakers

Suffix	Reversal direction	Number of reversals
<i>ero</i>	D(iphthong) → M(onophthong)	3
<i>oso</i>	D → M	2
<i>ista</i>	D → M	2
	M → D	1
<i>ísimo</i>	D → M	2
	M → D	4
<i>ito</i>	M → D	1
<i>azo</i>	D → M	3
	M → D	1

4.3.2 Reaction time analyses: Conditional naming

There was a significant interaction between suffix and diphthong status in the logRT analysis by items $F(6, 205)=2.542, p<.05$, but not by subjects. No other effects

were significant. Planned comparisons revealed that neologisms ending in *ito* were responded to more slowly when the stem contained a diphthong, and that neologisms ending in *azo* were responded to more slowly when the stems contained a monophthong. While there is a suggestion of a trend towards slower reaction times with diphthongs for most of the remaining suffixes, this trend was not significant. The results are shown in Figure 4-5 .

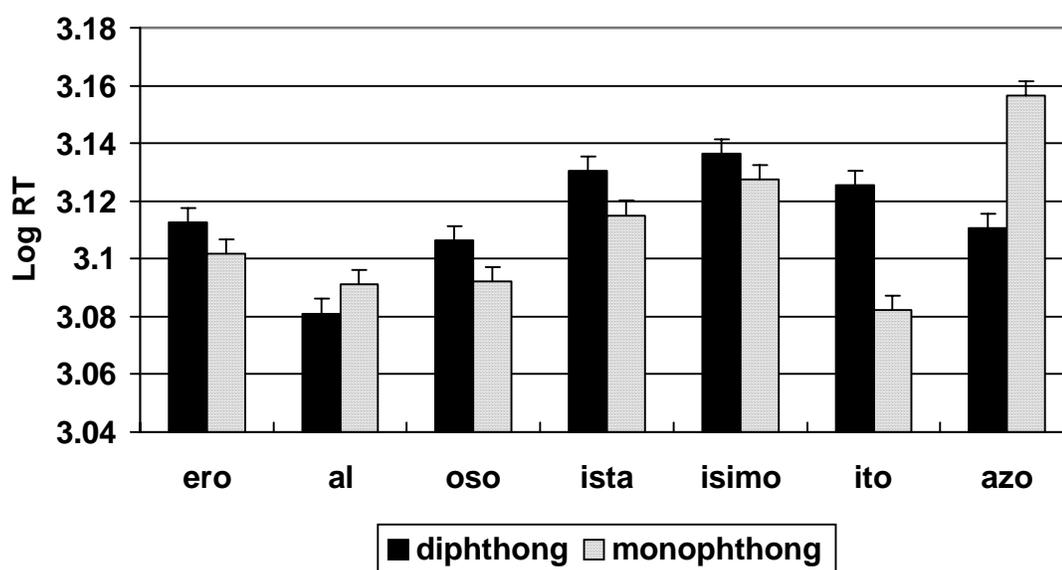


Figure 4-5: Log RT data for neologisms in conditional naming by native speakers, analysis by items

The logRT results from this task generally resemble the false alarm rates in that the general pattern shows little bias towards either monophthongs or diphthongs, except in the case of *ito* and *azo*. In contrast to the false alarm data, however, the trends for these two suffixes are opposite. Neologisms ending in *ito* are produced with longer latencies when they contain diphthongs in the stem, but neologisms ending in *azo* are produced with longer latencies when they contain monophthongs.

This complicates the interpretation of these results. If, as was suggested in the LDT, shorter latencies are associated with greater wordlikeness, then the RT results for *azo* parallel the bias towards diphthongs shown in the false alarm rates. This interpretation is also in line with the apparent though nonsignificant trend toward shorter reaction times for neologisms with monophthongs shown here for most of the remaining suffixes. However, this would leave unexplained the apparent preference for monophthongs in neologisms ending in *ito*. If, on the other hand, wordlikeness has an inhibitory effect on naming latencies, in contrast to the facilitatory effect observed in the LDT (see Section 4.2.2), the trend observed here for *ito* would be in line with that observed in the false alarm rates and in Eddington's results. However, this would require an explanation for the opposite predictions regarding RTs for the LDT and naming paradigms. Furthermore, it would leave unexplained the apparent preference for monophthongs in the neologisms ending in *azo*. Possible reasons for these contradictory results for two suffixes predicted to behave in a similar way will be considered further in the discussion.

4.4 Discussion

The results of the present experiments for native speakers of Spanish may be summarized as follows. Eddington's corpus and experimental results predict a gradient pattern in which neologisms with stem monophthongs are more wordlike with some derivational suffixes, but this monophthong bias becomes neutral for other suffixes, and it reverses for still other suffixes, such that neologisms with stem diphthongs are more

wordlike in these latter cases. Ranking the seven suffixes examined here according to these predictions, the general converse trends of decreasing wordlikeness with stem monophthongs and increasing wordlikeness with stem diphthongs is supported by the present experiments.

This support is strongest in the analysis of false alarm rates in lexical decision, where, with the exception of the suffix *ero*, the bias at one end of the ranking is toward monophthongs and at the other end it is toward diphthongs. This pattern also obtains in certain production errors, in which participants convert dispreferred diphthongs into monophthongs and vice versa, although participants produced relatively few of these errors.

In the (log) RT data for lexical decision the trend is generally preserved (excepting the case of *al*), but while there is a likely monophthong bias at one end of the ranking, the other end shows ambivalence with regard to the status of the diphthong alternation. However, this is only true if wordlikeness facilitates lexical decision under these experimental conditions. On the conditional naming task, both false alarm rates and RTs reflect a general ambivalence regarding diphthong status, except that a bias towards diphthongs is evident at the predicted end of the ranking. However, the two suffixes at this end of the ranking, *ito* and *azo*, show anomalous results with respect to naming RTs.

Finally, there appears to be an effect of the length of the stem whereby monosyllabic stems affect the results of the LDT differently with respect to false alarm rates and RTs. While the trends are similar for both measures regardless of whether neologisms with monosyllabic stems are included, the crucial interaction between suffix and diphthong status is only significant for false alarm rates when these items are

included, and it is only significant for RTs when these items are excluded. Exclusion of these items rendered the conditional naming results nonsignificant.

These results and their interpretation will be discussed in the following sections. First, the predictions for the effect of wordlikeness on RT data will be discussed, and an argument made for a facilitation effect of wordlikeness in straight lexical decision. Second, the variability in the extent to which the present results match the predictions based on earlier corpus and behavioral research will be discussed. Particular attention will be paid to the anomalous behavior of certain suffixes, the effects of stem length, and the distribution of monophthong and diphthong bias as well as ambivalence across the two dependent measures and the two tasks.

4.4.1 Effects of wordlikeness on lexical processing: Facilitation or inhibition?

As noted above, the RT data for the LDT show an interesting pattern with respect to the ranking obtained in Eddington's forced choice experiment. Three of four suffixes (excepting *al*) at one end of the ranking received longer latencies when the neologism contained a stem diphthong, *ísimo* received longer latencies with a stem monophthong, and the remaining two suffixes (*ito* and *azo*) received more or less equal RTs in both diphthongization conditions. Recall also that the interaction between suffix and diphthong status was only significant when items containing monosyllabic stems were excluded, although the trends were similar without the exclusion of these items.

Leaving the neologisms ending in *al* and the two ambivalent suffixes for the next section, these results are aligned with a division of Eddington's ranking in two halves,

one in which monophthongs are preferred, and the other showing a preference for diphthongs. However, the prediction made based on prior research on the effects of neighborhood density and phonotactic probability of nonwords on lexical decision (Dell & Gordon, 2003; Vitevitch & Luce, 1998, 1999; Vitevitch et al., 1999) was that more wordlike stimuli, and those with more related forms in the lexicon, would inhibit lexical decision. In the current results, faster latencies are associated with monophthongized forms for *ero*, *oso*, and *ista*, and with diphthongized forms for *ísimo*, which were all predicted to be more wordlike by Eddington's corpus and experimental results. Excluding the ambivalent results for *ito* and *azo*, the results appear to be the reverse of what was predicted. This raises the possibility that greater wordlikeness may have a facilitatory effect for this particular type of stimulus, and that the inhibitory effects of phonotactics and neighborhood density are reduced or absent.

The interpretation of a facilitatory effect for wordlikeness in the current experiment receives further support from the more robust effect observed in the false alarm data from the LDT. The prediction of increased false positives for more wordlike items is clearer, and is consistent with Eddington's results. Similar, albeit somewhat weaker, support can be seen in the reversal and false alarm data from the conditional naming task. Given the similarities between the RT data from the LDT and these other measures, and despite small variations that will be discussed in the next section, the results may be interpreted as consistent with facilitation of lexical decision for more wordlike stimuli.

This change in predictions is not without independent justification. With regard to phonotactic probability, Vitevitch and his colleagues (Luce & Large, 2001; Vitevitch,

1997; Vitevitch & Luce, 1998, 1999; Vitevitch et al., 1999; Vitevitch & Sommers, 2003) have reported robust facilitation effects on a variety of tasks apart from the LDT paradigm. Furthermore, in a comprehensive review of the lexical neighborhood literature, Andrews (1997) concluded that increasing orthographic neighborhood size generally facilitates processing, and she attributed most of the inhibitory effects reported to variations in the experimental tasks used.

The strongest support for reversing the predictions for the RT data comes from the literature on morphological processing. The current study relied on a standard LDT paradigm, but with a crucial difference in the nature of the test stimuli. The test items were all neologisms constructed from existing (and relatively frequent) stems and suffixes. Thus while they were not attested in corpora of Spanish, they were nonetheless possible, and indeed likely Spanish words. The use of real morphemes to create the nonwords resulted in a stronger resemblance to real Spanish words than in traditional nonword experiments. In the current data this was reflected in two ways. First, while false alarm rates were well below chance, ranging roughly from 3% to 30% (see Figures 4-2 and 4-4), they were significantly higher ($p < .001$) than the false alarm rates obtained for attested Spanish words and for nonce words, including the items formed using nonce stems and real suffixes. Second, they also received the slowest reaction times of any stimulus type. Reaction times for the test neologisms were longer than those for nonce words, suffixed nonce words, and suffixed words, which were in turn longer than those for monomorphemic real words. These results strongly suggest that the neologisms were more difficult to process than ordinary nonce words or real words. Moreover, the current experiment presented both the LDT and the conditional naming tasks in the

auditory modality. Under these unique task conditions, it may be that the verification process for more wordlike items would proceed more easily than for less wordlike items.

Most crucially for the present discussion, the use of the auditory modality entails the unfolding of the test stimuli over a period of several hundred milliseconds (neologism durations ranged from 400 to 1100ms). Unlike a visual LDT, in which the entire word may be perceived more or less instantaneously, these temporal constraints result in stems being heard before suffixes (see Giraud & Grainger, 2003 for a study contrasting the processing of derived words involving prefixes and suffixes). This has the consequence of allowing a significant amount of processing to take place before the suffix is heard.

Since all of the stems in the test items were attested forms, the neologisms in this study were effectively real Spanish words up until the onset of the suffix. The only cue to the contrary that is available in the hearing of the stems would be that diphthongs in these stems generally are stressed. While the stress pattern of an auditory stimulus may not be fully apparent until the whole token is heard, there are still intonational and durational cues to stress that may be perceived earlier, although this may not be enough in some cases to form a reliable prediction about the stress pattern of the item. Participants may thus be biased towards a WORD response prior to onset of the suffix, and then revise this response upon perceiving the suffix. This account is consistent with the predictions for false alarm rates, but it is still unclear how this might affect RTs.

Three recent studies help to make the step from this observation to a prediction of facilitation in lexical decision for more wordlike neologisms. First, Spivey et al. (Spivey, Grosjean, & Knoblich, 2005) used a mouse tracking paradigm to track listener's goal-directed hand movements in response to auditory input. In this way, they were able to

track the unfolding of lexical activation as a function of the temporal dynamics of speech (see also the commentary by Magnuson, 2005). Their results showed a continuous uptake of sensory input and a dynamic progression of competition among a continually changing set of simultaneously active representations. Similar and even more relevant results were obtained in a study combining auditory speech input with eye-tracking data in a visual environment (Magnuson, Dixon, Tanenhaus, & Aslin, 2007). Listeners' eye movements were facilitated in response to more frequent words, and words from dense neighborhoods showed an early facilitation and late inhibition effect. Magnuson et al. interpreted these results as indicating that the set of active competitors changes over time as speech unfolds.

Applying this reasoning to the current experiment, this suggests that as a neologism unfolds over time, a set of active candidates will evolve over which the lexical status of the neologism will be verified. According to the results of Spivey et al. and Magnuson et al., this evolution occurs continuously, as each segment or acoustic gesture is perceived. Since the stems are all real Spanish stems, a set of phonologically and morphologically related words will become active as the stem is perceived. In addition, expectations regarding the length of the word may also be formed based on phonemic and subphonemic cues, including the presence or absence of a diphthong and cues suggesting the presence or absence of stress (see again Magnuson et al., 2007). The candidate set may thus include not only specific candidate words, but also predictions about the complete shape, stress pattern, length, and other characteristics of the item, against which verification proceeds.

The manipulation of suffixes and diphthong status in the present experiments is likely to interact with this evolving set of candidates and expectations in the following way. When a stem with a diphthong or monophthong is heard, predictions are made about the length and possible morphemic content of the word. The ensuing suffix may be either consonant or dissonant with those predictions, based on the likelihood of co-occurrence of that suffix with either diphthongized or monophthongized stems. If the suffix is consonant with those predictions, the candidate set may not be altered much, and verification proceeds more quickly. If on the other hand the suffix conflicts with expectations, the candidate set must first be revised before the verification process can be completed. This revision is likely to increase the response latency.

By this account, the predictions for the auditory tasks used here, involving neologisms composed of real Spanish morphemes, should be that higher wordlikeness will facilitate lexical decision. This prediction is further supported by another study by Soto-Faraco et al. (2001) in which participants listened to sentences in which the last word was truncated (e.g. *princi-*), with no cues as to the meaning of the truncated word. Following the sentence, a word was presented visually for lexical decision. The results showed that lexical decision was facilitated if the word on the screen matched the fragment, but the decision latency was longer if the word and fragment differed either by one vowel, one consonant, or in stress pattern.

According to the predictions based on the gradual and dynamic evolution of candidates against which neologisms are verified, the facilitation observed for test items that more closely match lexical patterns in Spanish is fully expected. However, the data are not as clear in the conditional naming task as in the LDT. For most suffixes, there was

no significant difference in logRTs between neologisms with diphthongs and those with monophthongs, although there were significant and opposite differences for the two suffixes for which a preference for diphthongs was predicted. These differences may have been due to the production aspect of the naming task.

For the conditional naming task, participants were instructed to say the word *palabra* if they were confident that they had heard a real Spanish word, and to repeat the stimulus if they thought it was not a real word. This response was, of course, based on a lexical decision which should follow the same dynamic as the LDT, which relied on a button press. However, the language production component introduces additional processing that is likely to unfold simultaneously with the processing underlying the lexical decision itself. Specifically, as the real Spanish stem is perceived, a set of candidate words that fit closely with the stem is activated. The evidence at this stage points to a real Spanish word, and thus the response of *palabra* may be prepared. As the suffix unfolds, it will again be either consonant or dissonant with the hypothesis that the item is indeed a real Spanish word. If it is consonant, then verification proceeds as usual, and at the point that the listener decides it is not a real word, the response will be revised to a repetition of the item instead of the production of *palabra*. If, however, the identity of the suffix is dissonant with the hypothesis of a real word, that is, if the neologism is less wordlike, this will provide earlier evidence that the affirmative response must be revised, and a repetition of the item will be prepared. New candidates for verification may still be activated, delaying the ultimate response as suggested above for the LDT, but the revision in neuromotor programs underlying the response may begin earlier with the less wordlike neologisms.

Since the identity of the suffix may provide an earlier cue that less wordlike neologisms are not real Spanish words, the preparation of the appropriate response may begin earlier. In the case of more wordlike neologisms however, this early cue is not present, and the preparation of the negative response is likely to begin later in the verification process.

There are thus two conflicting influences at work. The verification process should take somewhat longer for less wordlike neologisms based on the activation of new candidates for verification, but the neuromotor programs underlying articulation of the appropriate response may be activated earlier for the same items (or at least the inhibition of the articulatory program for *palabra* may begin earlier), thus shortening RTs. On the other hand, verification may be faster for more wordlike neologisms, but switching from the neuromotor programs for *palabra* to those for a repetition may slow RTs. This suggests that there are differences in the timecourse of perceptive and productive lexical processing, although more focused inquiry will be required to clarify this possibility.

The opposing influences of lexical verification and preparation of an articulatory response in the conditional naming task may thus explain the leveling observed in the RT effects for most suffixes. On the other hand, two suffixes for which a diphthong bias was predicted, *ito* and *azo*, showed significant differences in RTs based on diphthong status, although the direction of the difference was different in each case. The explanation for the behavior of these two suffixes may lie in the fact that they show the strongest lexical bias against the more general pattern in Spanish that alternating diphthongs most often surface as monophthongs when the alternating syllable is unstressed, according to Eddington's corpus data. The pattern in existing derivations with these suffixes thus

conflicts with the prevailing pattern in Spanish, whereas the trend for the remaining suffixes is more in line with that pattern.

For suffixes with a diphthong bias, the more general, stress-based pattern may serve to ameliorate the difference in wordlikeness between the form with a diphthong and that with a monophthong. Another way of viewing this possibility is to consider the pronunciation of the stems in the stimulus recordings. All stems, regardless of their diphthong status, appear without stress in the auditory stimuli, because the suffixes used all attract stress. Based on the general bias against unstressed diphthongs in Spanish verbs and other forms, stems with monophthongs should be intrinsically more wordlike, prior to onset of the suffix during listening. The onset of a suffix with a diphthong bias should thus cause relatively little dissonance with the candidate set, regardless of the diphthong status of the stem. In other words, the evolving candidate set may be relatively ambivalent regarding the diphthong status of a stem for the suffixes with the strongest diphthong bias, *ito* and *azo*.

This asymmetry between suffixes with a monophthong bias and those with a diphthong bias is consistent with the observation in the RT data from the perception experiment that these two suffixes showed relatively ambivalent ratings. The predictions for the false alarm data, however, would not necessarily be affected by this asymmetry, because, arguably, the ultimate decision regarding lexical status would be made in light of the entire neologism.

In the conditional naming paradigm, the leveling effect due to timing differences in the preparation of an articulated response would most strongly affect suffixes with a monophthong bias, since the general, stress-based rule would not alter the evolution of a

candidate set. In the remaining cases, however, the verification process may be relatively unaffected by the diphthong bias of *ito* and *azo*. If this is the case, then factors relating to the articulation of the specific neologisms may play a greater role. This would be expected to pattern according to wordlikeness, a prediction which makes the contrasting results for these two suffixes particularly puzzling. The reason for this contrast may be due to other characteristics of these particular suffixes such as the frequency with which they are produced by individuals (as opposed to encountered in print or other media), but further study will be required if this contrast is to be explained.

These observations regarding the predictions concerning false alarm and reaction time data underscore the importance of triangulation in examining complex questions about grammar and processing such as these. Further light may be shed on the issue by replicating these experiments in the visual modality, where the temporal characteristics of speech are collapsed. In the visual modality, words are recognized much more quickly than they can be articulated, and thus the evolution of a candidate set against which to evaluate and identify words follows a much shorter time course. Furthermore, with visual presentation the stem and suffix can be perceived simultaneously, and the activation of candidates based on the stem and suffix is unlikely to be sequential, as is the case with auditory presentation.

For this reason, visual LDT and naming tasks should be used to complement the findings of the present work. Electrophysiological measures of the time course of lexical processing may also shed light on the dynamics of activation and verification in the auditory processing of morphologically complex words with phonological alternations such as those examined here.

One further piece of evidence is relevant to this discussion of the processing of morphologically complex words in the auditory modality. In order that morphological complexity itself would not serve as a simple cue to the lexical status of test items, the experimental design included a set of attested Spanish derivations and a set of nonword stems with real derivational suffixes. In the former case the experimental manipulation of diphthong status was naturally impossible, given the small number of attested derivations that may appear in either form. With the second group, the nonce stems were constructed to contain either a diphthong or the corresponding mid-vowel, and concatenated with the seven experimental suffixes.

As these nonce derivations unfold, an evolving set of candidates for verification is activated, as before. However, while even nonwords have phonological neighbors and a certain probability of form, this set is unlikely to be as extensive or robust as that activated during the audition of real stems, at least with respect to the activation of semantically or morphologically related forms. Diphthong status is thus highly unlikely to play a significant role in this activation, because Spanish contains many words with diphthongs and monophthongs that do not alternate, and there is thus no evidence for any phonological alternation in the nonce stems. Moreover, on hearing a nonce stem, speakers are likely to anticipate a NONWORD response (either a button press or a repetition, in these experiments). The onset of a real suffix may cause some reevaluation of this hypothesis, but it is unlikely to affect processing as strongly as if the item contained a real stem.

The present experiments yielded no significant results for the set of nonce derivations, either in RTs or in false alarm rates. Neither the diphthong status nor the

suffix, nor their interaction affected the dependent measures. False alarm rates for the nonce derivations were significantly lower than those for neologisms with real stems and suffixes, and the RT data also indicated that they were processed faster than the neologisms, but slower than unsuffixed real words (all $p < .001$). That these stimuli were easier to process than neologisms is consistent with the account presented here, and the contrast with unsuffixed real words is predicted simply on the basis of the presence of two morphemes.

There were, however, no stimuli in which a real stem was concatenated with a nonce suffix. This kind of stimulus should also produce a null result, because a nonce suffix would provide unambiguous evidence against the classification of the item as a real word, without regard to any alternations inherent in the stem. However, care must be taken that nonce suffixes would be recognized as separate from the stem. This could be done by using longer and more frequent stems, such that both the identity of the stem and the following morpheme boundary are relatively unambiguous before onset of the nonce suffix. In addition, it would be interesting to observe the effects of repeated presentation of nonce suffixes with real stems on processing over the course of a task, although this is less relevant for the current research questions.

While derivations with real stems and nonce suffixes may provide a useful set of control stimuli for experiments such as the present one, the overall outcome of this innovation in the lexical decision paradigm is not entirely clear, and piloting using other task types, including offline metalinguistic tasks, may help to both ensure that nonce suffixes are indeed perceived as suffixes and to clarify how such stimuli might affect lexical processing.

4.4.2 Gradient patterns in corpus data, forced choice, false alarms, and reaction times

Overall, the results of the present experiments support the hypothesis that Spanish speakers' knowledge and processing of the diphthong/mid-vowel alternation is gradient, rather than categorical. Analysis of the interactions between derivational suffix and diphthong status revealed that in some instances neologisms with stem monophthongs are more wordlike, at other times those with stem diphthongs are more wordlike, and at still others there is ambivalence toward the two possible forms of the neologism. Moreover, the general trend is in keeping with the corpus and experimental results obtained by Eddington (1996; 1998). Suffixes that co-occur more often with stem monophthongs in the existing Spanish lexicon generally show a monophthong bias in neologisms, those that co-occur more often with diphthongs generally show a diphthong bias, and stems with no lexical preference generally show ambivalence.

While the hypothesis of gradience receives support from the stable pattern of variability in diphthong bias between suffixes, there are several discrepancies shown in the data. First, while Eddington's corpus of the most frequent words in several dictionaries revealed a categorical bias for monophthongs with the suffixes *ero* and *al*, neither his forced choice task nor the present experiments matched this bias. Instead, the monophthong preference with these suffixes was high, but not categorical. This is relatively easily explained if derivations with other suffixes are allowed to influence suffix-specific biases.

Another source of discrepancy stems from the fact that all of these suffixes are also productive, and it is entirely possible that speakers' experience of diphthongizing

stems with these suffixes may contrast somewhat with those words that have been accepted into the official Spanish lexicon and thus appear in dictionaries. It was in fact for this reason that Eddington chose not to even report the dictionary statistics for diphthongization in words with the most productive suffixes (in this study, *ísimo*, *ito*, and *azo*).

Apart from this slight mismatch with the dictionary corpus data (and the lack of precise, corpus-based predictions for the most productive suffixes), other inconsistencies between the various methodologies and dependent measures must be addressed. First, there are some instances of individual suffixes with anomalous behavior with respect to the predictions. Also of interest is the observation in the LDT results that monosyllabic stems may be associated with different outcomes in the false alarm and RT data. Finally, the exact distribution of monophthong bias, diphthong bias, and ambivalence varies among the dependent measures, although the larger-scale trend for each dependent measure is always in the predicted direction.

The false alarm data from the LDT most closely and dramatically parallel the results from Eddington's forced choice task (see Figures 4-1 and 4-2), with the exception of *ero*. There is a clear boundary between the suffixes *ista* and *ísimo*, to the left of which monophthongs induce more false positives and to the right of which diphthongs induce more false positives. The boundary suffixes themselves show preferences in the present data, but ambivalence in Eddington's data, but they nonetheless serve to divide the monophthong bias from the diphthong bias. The fact that this dependent measure most closely parallels the forced choice results is not surprising, because both tasks, in essence, require a decision regarding the lexical status of the test item. The main

difference was that the forced choice task involved presentation of both possible forms of the neologism simultaneously (as well as other, attested derivations of the stem with both a diphthong and a monophthong), while the LDT presents one form at a time.

This contrast in task conditions is likely to accentuate participants' preferences in the forced choice paradigm, which may explain the more ambivalent results for *azo* and *ero* in the LDT error rates, although other explanations must be considered. Planned contrasts revealed that the patterns of results for these two suffixes differed significantly only from the monophthong bias observed for *al*. It thus appears that the bias for these suffixes is relatively neutral, although it trends towards a diphthong bias (based on the nonsignificant contrast between these suffixes and *ito*). One way in which the experimental conditions may have led to this result lies in the presentation by Eddington of both possible forms of each neologism, whereas in the present experiment only one form was given. In Eddington's experiment this may have increased the role played by information about stem alternation by making this information explicit. While alternate forms of the test stems in the present experiment are likely to have been activated during processing, this information may not have been foregrounded to the same extent, reducing its impact on processing time. The observed variability in experimental results can thus be attributed at least in part to differences in the relative activation of different kinds of lexical and grammatical information occasioned by contrasting task demands.

An additional source of variability in results is possible based on the nature of Eddington's corpus data. As reviewed in Section 2.2.3, Eddington calculated his lexical statistics based on data from the *Diccionario de la Lengua Española* (Real Academia Española, 1970) and other electronic corpora (Eddington, 1996). He extracted all derived

words with alternating diphthongs and elicited familiarity judgments from a group of native speakers, which yielded a list of what he considered to be the most frequent, or most familiar derivations. Excluding the most productive suffixes, the distribution of diphthongs among these derivations formed the basis for Eddington's predictions, and mapped well to his forced choice data.

There are several problems, however, with the use of this sample. Eddington acknowledged that dictionary data are unlikely to provide a clear indication of the behavior of highly productive affixes, but all of these suffixes are productive to a certain extent. It is also unclear how best to reduce the dictionary sample to most closely reflect the lexicon of an average speaker. Eddington's use of a subjective familiarity measure is a good start, but more detailed work is required to more closely converge on the set of derivations that are likely to be represented whole in memory (Bybee, 2001b).

Comparison of the distribution of diphthongs across the entire dictionary corpus and the distribution among the most familiar items reveals that, in the larger sample, the monophthong bias for *ero* was much weaker (77% of attested derivations contained a monophthong, vs. 100% in the smaller sample). The larger sample also revealed a strong preference for monophthongs (90%) for *azo*, and the value for the smaller sample was not given. It is clear, then, that the determination of the correct subset of a large corpus is not a trivial matter, given the contrasting lexical biases in these two samples.

Task modality may also play a role. Eddington's experiment used a written task, and the present study used auditory tasks. Speakers' familiarity with lexical items and the biases emergent over that familiarity may vary between these two modalities, due to differences in the frequency of particular lexical items in different genres of text and

speech. More detailed corpus sampling and closer examination of visual and auditory experience with language are needed to clarify this matter further.

The data for one further suffix, *al*, deserve further comment. In the items analysis of the RT data from the LDT, items in *al* with diphthongs received shorter latencies, implying a diphthong bias where a monophthong bias was predicted. However, this analysis was only significant when neologisms with monosyllabic stems were excluded, and this exclusion affected *al* and *azo* more than the other suffixes. In fact, these two suffixes had to be excluded from the analysis by subjects in order to preserve enough cases to be analyzed. The means observed for these suffixes are therefore based on fewer items, and may therefore be less reliable. Recall that the division between neologisms with monosyllabic stems and those with longer stems was made based on post-hoc participant reports that indicated that the shorter stems may not have been reliably identified, in which case the diphthong alternation would not have played any role in processing.

The exclusion of monosyllabic stems creates an interesting pattern in the LDT results based on particular characteristics of the stems used, in this case their length in syllables.¹⁰ Specifically, the interaction between suffix and diphthong status in false alarm rates was significant only when including all test items, and the interaction in logRTs was significant only when neologisms with monosyllabic stems were excluded. It

¹⁰ Several other statistics were calculated for the stems used in these experiments, including the number of times it occurred (above a frequency of 3.03 per million in the LexEsp corpus) in monophthongized and diphthongized form (type frequency of each variant), the frequencies of the most common derivations with a monophthong and a diphthong, respectively, the difference between these two frequency values, the number of derivations in which the stem occurred (total type frequency) and the total frequency of the stem in all derivations (total token frequency). None of these variables yielded a significant result in any of the present experiments.

is understandable that the exclusion of half the test items might sufficiently reduce the statistical power to render an effect nonsignificant, as in the false alarm data, and this is made more likely by the fact that false alarm rates were relatively low to begin with (between roughly 5% and 22% on the LDT). There was no comparable ceiling effect in the RT data (average RTs for neologisms were slightly higher than 1100ms, and the average stimulus duration was only a little over 700ms), but this still does not explain why a significant effect was observed in the RTs only with the reduced set of stimuli.¹¹

The reason for this pattern of findings in the LDT may lie in the time course of activation during the unfolding of the auditory stimulus, which was discussed above. A shorter stem would have several consequences for processing. First, a larger candidate set is possible based on the hearing of one syllable than on two, similar to the observation that shorter words have more possible neighbors. Alternately, the candidate set may not be as well defined for shorter stems. In either case, a monosyllabic stem does not contain as strong a cue to its identity as a longer stem, and consequently the bias towards an affirmative response in lexical decision is likely to be weaker. Second, following from the larger and/or more diffuse candidate set for monosyllabic stems, the prediction that the following syllable will initiate a new morpheme is likely to be weaker. This may result in delayed recognition of the suffix, which might in turn force a later reanalysis of the initial syllable as a complete morpheme in its own right.

¹¹ It should further be noted that the conditional naming experiment did not show this discrepancy based on the length of the test stems. Both the false alarm and RT analyses yielded significant suffix X diphthong status interactions across all stimuli (in the latter case only for the analysis by items), but not after exclusion of the shorter stems. Closer examination of the RT data, however, revealed that the nonsignificant effect was in large part due to the effect of the suffix *azo*, for which very few stimuli remained after removal of those with monosyllabic stems.

The effect of this additional processing would be to delay RTs for neologisms with monosyllabic stems in relation to the total duration of the stimulus. This was, in fact, observed in the LDT results. While response latencies for test items with monosyllabic stems were significantly shorter than for items with trisyllabic stems, $p < .05$,¹² controlling for duration in milliseconds, they were not significantly different from those for items with disyllabic stems. This supports the hypothesis that verification is delayed for neologisms with monosyllabic stems, which in turn may obscure effects due to the diphthong status of these items.

The greater cue strength for the identity of longer stems is likely to have the further effect of boosting the error rate for the less wordlike neologisms in relation to shorter stems, thus making the likelihood of false positives more equal across less and more wordlike stimuli. Again the present results support this account. False alarm rates for the more wordlike stimuli are relatively unaffected by the exclusion of neologisms with monosyllabic stems, whereas there are comparatively more such errors for the less wordlike stimuli. The time course of processing thus predicts that the inclusion of shorter stems would accentuate the effects of wordlikeness on false alarm rates, and ameliorate those effects for RTs, yielding the current results.

One final result deserves comment here. The more general ambivalence in the RT data from the naming experiment has been commented on above as a possible result of the reversal of early activation of a *palabra* response. This revision in neuromotor

¹² Overall, the ANCOVA with logRT as the dependent variable, stem length in syllables as the independent variable, and stimulus duration in milliseconds as a covariate revealed a significant main effect for stem length, $F(2, 216)=3.688, p < .05$.

activation patterns was predicted to conflict with the time course of lexical decision, with the result of leveling effects due to the interaction between suffix and diphthong status. However, the RT data from the LDT, regardless of the presence or absence of the stimuli with shorter stems, suggest a greater ambivalence in RT data in general than is shown in the false alarm data.

The crucial difference between these two dependent measures is that the false alarm rates reflect the ultimate polarity of the lexical decision, regardless of the processing that led up to it, and the RTs reflect the cumulative latency of processing. It is thus not surprising that these measures yield different results. The precise results in this experiment show that the false alarm data (LDT) yielded significant biases for almost all stimuli, but the RT data showed almost no bias for some suffixes, especially for *ito* and *azo*.¹³

It has already been stated that the forced choice paradigm may have accentuated what were in reality only slight, but reliable, preferences. In the same way, the false alarm data reflect a categorical, binary choice that may be affected more by the reliability of the lexical bias of a given suffix than its magnitude. In contrast, processing, as reflected in RTs, may be more sensitive to the magnitude of the bias.

Furthermore, the eventual direction of participants' decisions is determined after hearing the whole stimulus, but there are a variety of factors relating to the stem that may affect processing before the identity of the suffix is revealed. This aspect of the RT measure would have the effect of accentuating general patterns in the grammar, such as

¹³ It is difficult to compare the quantitative strength of the biases between these two measures.

the general bias against unstressed diphthongs, and delaying the processing effects of elements that occur later in the audition of a stimulus. The fact that relative ambivalence in the LDT response latencies is observed for those suffixes with the strongest bias against the general, stress-based pattern is entirely consistent with this account.

Finally, the suffixes with the strongest diphthong bias also tend to be the most productive suffixes. This may affect perception more than production by increasing the likelihood that speakers will accept a wider range of variance in the realization of productive morphological processes than they might themselves produce. This may also have the effect of leveling the effect of lexical biases on RTs in the case of highly productive suffixes. In short, speakers are more willing to accept a word that doesn't exactly fit their existing lexical patterns, provided it is interpretable, but they may be unlikely to produce them.

4.5 Conclusions

The results of these two experiments allow for a detailed evaluation of the perception and production of neologisms containing alternating diphthongs by native speakers of Spanish. All of the suffixes used to create the neologisms attract stress, removing it from the alternating syllable. The general pattern in Spanish disprefers unstressed diphthongs in alternating stems, but there are many words in Spanish that violate this general pattern. Earlier corpus and experimental studies (Eddington, 1996, 1998) have shown that the probability of an unstressed diphthong occurring in derived

words varies gradiently depending on the relative co-occurrence of unstressed diphthongs with specific derivational suffixes in the existing Spanish lexicon.

The results from the current lexical decision and conditional naming tasks both replicate and extend these earlier findings. The pattern of gradience shown in the earlier results appeared most strongly in the false alarm data in the perception experiment, but was also evident to varying degrees in response latencies and in conditional naming.

However, the more sensitive and multifaceted nature of the online and auditory experimental tasks, compared to the earlier forced-choice paradigm, reveal a more complex interplay of a number of variables. In particular, certain characteristics of the alternating stems, namely, length in syllables, came into play and may have influenced the time course of the lexical decision before the onset of the derivational suffix. In addition, the need to make and revise decisions online about an articulatory response in the production task may have interacted with the verification process. The dynamic evolution of and competition within a candidate set as an auditory stimulus unfolds helps to make sense of the varying results between the perception and production tasks, and between the false alarm and RT measures. It also finds support in recent eye-tracking, hand-movement (mouse-tracking), and lexical decision studies. More convincing evidence could be obtained by comparing the present results with those from a visual version of the same experimental tasks. The prediction is that the effects due to the time course of lexical activation in the auditory modality would disappear in the visual modality, in which the entire stimulus may be perceived more or less instantaneously. A further variation could make use of priming or an asynchrony between visual presentation of stems and suffixes to reintroduce the temporal dimension into the visual modality.

Finally, there were occasional discrepancies between the present results and the corpus-based predictions for certain suffixes. This may have been due to inaccuracies in the dictionary samples used to make those predictions, as well as to the suitability of dictionary data given the auditory modality of the present experiment. Finer-grained corpus analyses based on different kinds of texts and genres may help to clarify the picture by measuring differences between context, register, and modality. While visual and auditory processing are inseparable for literate individuals (Ziegler et al., 2004; Ziegler et al., 2003), words that are primarily encountered in one modality or the other might affect grammatical biases in different ways.

Notwithstanding the remaining questions about native speaker processing of alternating diphthongs and morphologically complex words in Spanish, these experiments provide evidence that the variable patterns in the diphthong/mid-vowel alternation found in extant Spanish derivations have measurable consequences for lexical processing. These results therefore constitute a suitable baseline against which to compare the performance of adult learners of Spanish in the remaining experiments in this dissertation.

Chapter 5

Results: Adult Learners

5.1 Chapter overview

This chapter examines and discusses the results for the adult learners of Spanish. As with the native speakers, the experiment consisted of two central components, each carried out with the same group of subjects. An LDT was used to investigate learners' sensitivity to gradient subpatterns in Spanish diphthongization in perception, and a conditional naming task was used to explore that sensitivity in language production. In addition, the learners were measured on a number of parameters to assess differences between subjects in Spanish proficiency, experience, and several cognitive characteristics. These tasks are described in detail in Chapter 3, but the specific measures are reviewed again here.

Proficiency was measured using a picture naming task yielding two different indicators of proficiency. First, participants' overall accuracy in naming pictures was measured as a percentage of words correctly named. Items that were given semantically appropriate, but nontarget names were excluded from the analysis, the target names for each item having been determined according to the UCSD norms (Szekely et al., 2004). This yielded an indication of participants' vocabulary size. The second measure used was a lexical fluency measure calculated by subtracting the RT for the correct trials in English from that for the correct trials in Spanish. This calculation controlled for participants'

speed of lexical access in their L1. A larger value indicated relatively slower naming in Spanish as compared to learners' L1 English and, consequently, less lexical fluency. A smaller value indicated relatively faster naming in Spanish compared to English, and thus higher fluency.

Data concerning learners' experiences learning Spanish abroad was also collected. Fourteen participants reported having studied in a country where Spanish was the primary language. Of these, most had spent a semester abroad, roughly 3-5 months, and one each had spent 2, 6, 8, and 12 months studying abroad. With few exceptions, there was little variance in time spent abroad, and therefore study abroad was operationalized as a binary variable indicating whether a learner had or had not had this experience.

The remaining experimental tasks measured individual differences in phonological memory, working memory, and attentional or inhibitory control. Phonological memory was measured using a Korean word repetition task, a language that was not familiar to any of the participants. As described in Section 3.2.5.1, two scores were calculated based on this task. The first counted the percentage of words repeated with no obvious substitutions, omissions, or additions of phonemes. The second score was obtained by asking the native speaker raters to assess participants' degree of accentedness in Korean on a scale of one to six. The former will henceforth be referred to as accuracy in Korean word repetition, and the latter as accentedness in Korean. These scores were correlated, $r=.649$, $p<.001$, and are therefore not fully independent. Both of these measures were designed to measure participants' ability to remember and reproduce previously unencountered words under unfamiliar phonetic and phonotactic conditions meant to simulate the task of learning new L2 words.

Working memory was assessed using a reading span task in which participants judged the plausibility of sentences while attempting to remember the last word of each sentence in a set. This measure was scored by taking the number of words correctly recalled in the entire task, and will be referred to as memory span in the following discussion. Attentional or inhibitory control was measured on a Simon task, which required participants to respond to the color of a square on the screen, but ignore its position. The position of the square was either congruent with (on the same side) or incongruent with the hand used to respond. This task was scored by subtracting the mean RT for congruent trials from that for incongruent trials. A higher value for this “Simon effect” indicates lower inhibitory control, and a lower value indicates higher control.

For the purpose of the present analysis, participants were divided into high and low groups on each of these measures by a median split, except for study abroad, which was already binary, and accentedness in Korean, for which the raters only used four points on the scale.

The LDT and conditional naming tasks each yielded two dependent measures, which were analyzed in separate, mixed repeated-measures ANOVAs, with suffix (7 levels) and diphthong (2 levels) as the repeated measures, as for the native speaker group. First, RTs were measured from the onset of each stimulus until the initiation of a response. However, the RT curves exhibited significant positive skew. For this reason, analyses will be carried out on log (base 10) transformed RTs, based on correct trials. Second, given that the test items were neologisms not (yet) existing in Spanish, the percentage of false positive responses was calculated. Finally, the conditional naming task yielded a number of responses in which either diphthongs were converted to

monophthongs, or vice versa. This type of error is also of interest, because it offers a clear indication of learners' preferences regarding the form of a neologism. For this reason, although there were relatively few errors of this kind, they will be reported, as they were for the native speaker control group.

One further note regarding the analysis of the LDT and conditional naming tasks is relevant. In the native speaker group, there was some indication that neologisms with monosyllabic stems were potentially problematic. Due to their length, it was not clear that the target stems were being reliably recognized, although they appeared to impact the LDT more than the naming data. In the experiments with adult learners, two scenarios make the comparison between stems of differing length difficult. First, error rates were markedly higher for learners than for native speakers. As a result, there were so many empty cells in the RT matrix that statistical analysis was rendered impossible for the reduced stimulus set. Second, while trends were similar, there were no significant effects when neologisms with monosyllabic stems were excluded. This manipulation will therefore not be carried out here.

The remainder of this chapter will proceed as follows. First, the results of each experiment will be reported, first for the LDT, and then for the conditional naming task. Within each of these sections, the primary interaction of interest, that between suffix and diphthong status, will be reported first, followed by any three-way interactions between these variables and measures of proficiency, experience, or cognitive ability. Following the presentation of the data, the results of the various tasks will be discussed in a similar order. First, the interactions between suffix and diphthong status in the different

experiments will be discussed. Subsequently, issues concerning proficiency, and finally, cognitive resources will be examined.

5.2 Results for adult learners: Lexical decision task

5.2.1 Diphthongizing stems in neologisms with derivational suffixes

This section describes the interaction between derivational suffix and the presence or absence of a diphthong in the stem, apart from three-way interactions with between-subjects characteristics of individual learners. The results of the two dependent measures, logRT and false alarm rate, are presented.

In the logRT data from the LDT, there was a significant interaction between suffix and diphthong status, $F(6, 42)=2.975, p<.05$. The main effects for diphthong and suffix were not significant. As shown in Figure 5-1 and discussed later, there are marked similarities with the results for native speakers, but the results are not identical. Apart from the suffixes *ero* and *al*, there appears to be a distinction between suffixes to the left and to the right (note that the suffixes are ordered, from left to right, according to descending preference for monophthongs as predicted by the results of Eddington, 1998). For *ísimo*, *ito*, and *azo*, reaction times are shorter for neologisms with unstressed stem diphthongs (recall also that all of the stems in the test items are unstressed, by virtue of the attraction of stress to the suffix). Neologisms ending in *ista* are responded to more quickly when the stems contain monophthongs, and those in *oso* are ambivalent to diphthong status.

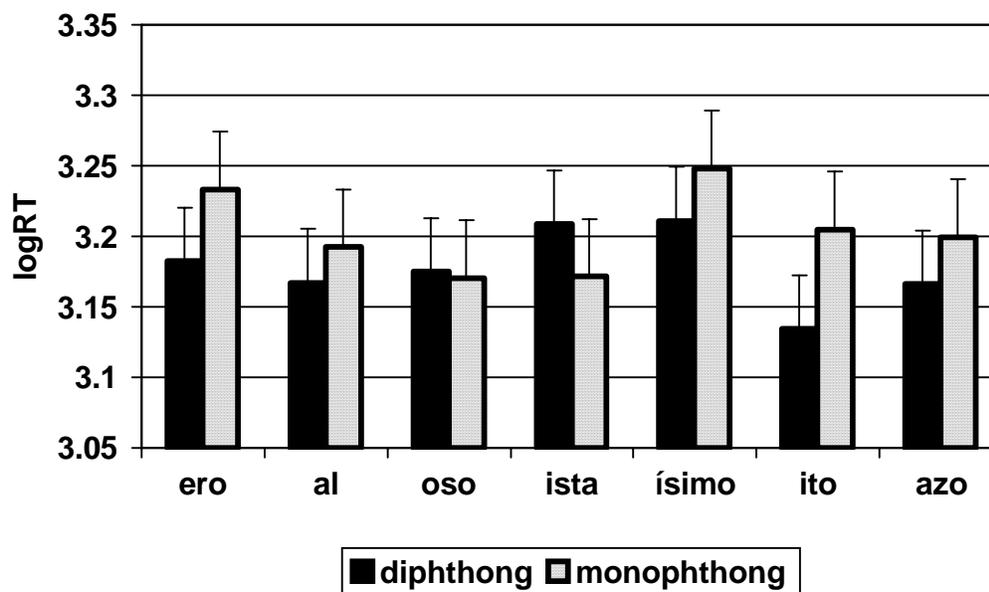


Figure 5-1: logRTs for neologisms in LDT by learners

Based on the prediction of facilitation for more wordlike neologisms (see Section 4.4.1), that is, for neologisms that better fit the pattern established in existing derived words with alternating stems, these results are generally consistent with Eddington's findings and the results from the native speaker control group, with some differences. The results for *oso* are more ambivalent than Eddington's results would imply, but this is consistent with the native speaker RT data from the LDT, though not the false alarm rates. The remaining two suffixes, *ero* and *al*, contrast with Eddington's data, in showing shorter reaction times for the diphthongized variant, although the difference is smaller for *al*. However, native speaker LDT performance with these suffixes was also inconsistent. Native speaker response latencies were shorter for neologisms ending in *al* when they contained unstressed diphthongs, although they made more false alarm errors on those

with monophthongs. The native speaker RTs for *ero* were more in line with the prediction of a monophthong bias, but in this case the false alarm data were inverted.

The false alarm rates in the LDT for the adult learner group differed from the native speaker group. While the main effect for suffix was significant, $F(6, 138)=2.597$, $p<.05$, indicating longer reaction times for neologisms ending in *ero*, neither the main effect for diphthong status nor the interaction between suffix and diphthong status was significant. It is relevant to note that false alarm rates were significantly higher ($p<.05$) for the critical neologisms than for any of the distractor types, including real words with and without suffixes, and nonce words with and without (real) suffixes. This was also seen in the native speaker results and indicates that neologisms created from real stems and real suffixes are in general more difficult to process. Error rates were about twice as high (between 35 and 55% by suffix) for neologisms as for other stimulus types.

This effect is unsurprising based on the use of real morphemes in invented words. It is also not surprising that learners' false alarm rates with neologisms were higher than those for native speakers. Learners in general tend to have higher false alarm rates in lexical decision. Of relevance to the present results is the fact that the high error rates resulted in a large number of empty cells in the logRT matrix for learners. Excluding cases listwise, this reduced the number of learners included in the analyses from 33 in the false alarms analysis to 18 in the logRT analysis. In order to further support comparison between the two dependent measures, the false alarm analysis was run a second time, including only those participants with valid RTs for all suffixes, i.e. the same subgroup of learners as the RT analysis. The results of this second analysis were nearly identical, the only difference being that the main effect for suffix was no longer significant.

5.2.2 Individual differences between learners

5.2.2.1 Considerations for data presentation

The results presented in this section depend on three-way interactions between suffix, diphthong status, and measures of Spanish proficiency, experience, or cognitive resources. In order to facilitate the interpretation of these interactions, a transformation was carried out on both the logRT and false alarm data. In the case of the RT data, the logRT for neologisms with stem monophthongs was subtracted from that for neologisms with stem diphthongs, separately for each suffix. Based on the prediction that greater wordlikeness would result in decreased response latencies, this transformation should yield a positive value for suffixes with a monophthong bias, and a negative value for those with a diphthong bias. Thus, a positive number indicates a bias that is in line with the general pattern in Spanish, in which unstressed diphthongs are dispreferred, and a negative number is expected for those suffixes whose behavior in existing Spanish derivations runs counter to this general trend.

This “monophthong bias” was also calculated for the false alarm data, such that a positive score would indicate greater wordlikeness for neologisms with stem monophthongs, as with the logRT data. However, in order to obtain values with the right polarity, false alarm rates for neologisms with stem diphthongs were subtracted from those for neologisms with stem monophthongs, because greater wordlikeness was predicted to result in higher false alarm rates. For clarity, Eddington’s (1998) forced choice results for native speakers are reproduced in Figure 5-2 using a similar

transformation. For comparison, these are the same data shown in Figure 4-1, except that the selection rate for diphthongs has been subtracted from that for monophthongs.

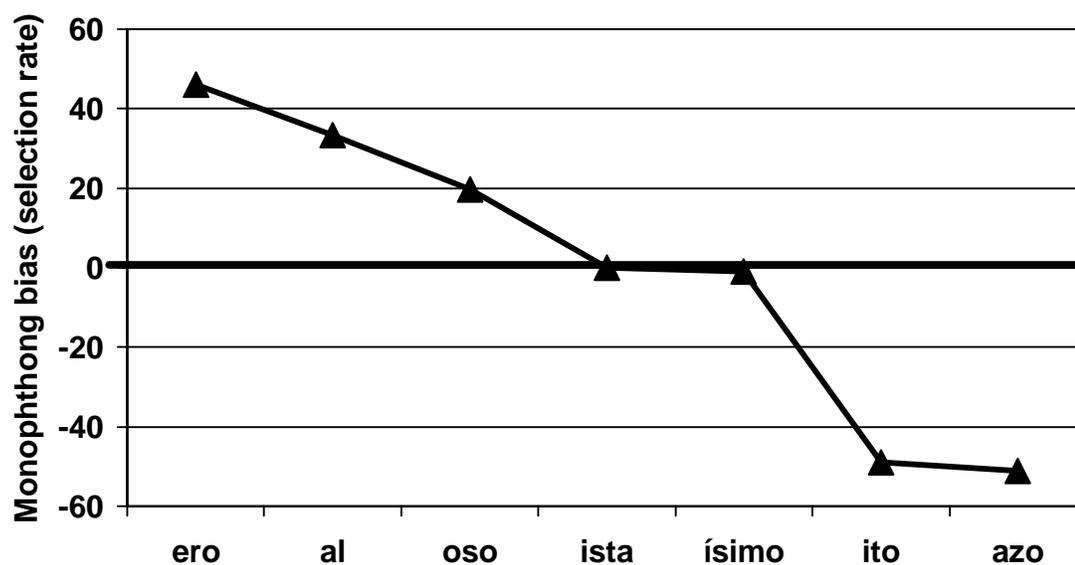


Figure 5-2: Monophthong bias in Eddington (1998) for native speakers on a forced choice task

To facilitate clear comparisons between the data for the learner and native speaker groups in the present experiment, the data from the native speaker groups are reproduced in Figure 5-3, transformed to show monophthong bias as discussed above. These data represent those presented in Figures 4-3 and Figure 4-2, except that the logRT data here include the monosyllabic stems, unlike in Figure 4-3.

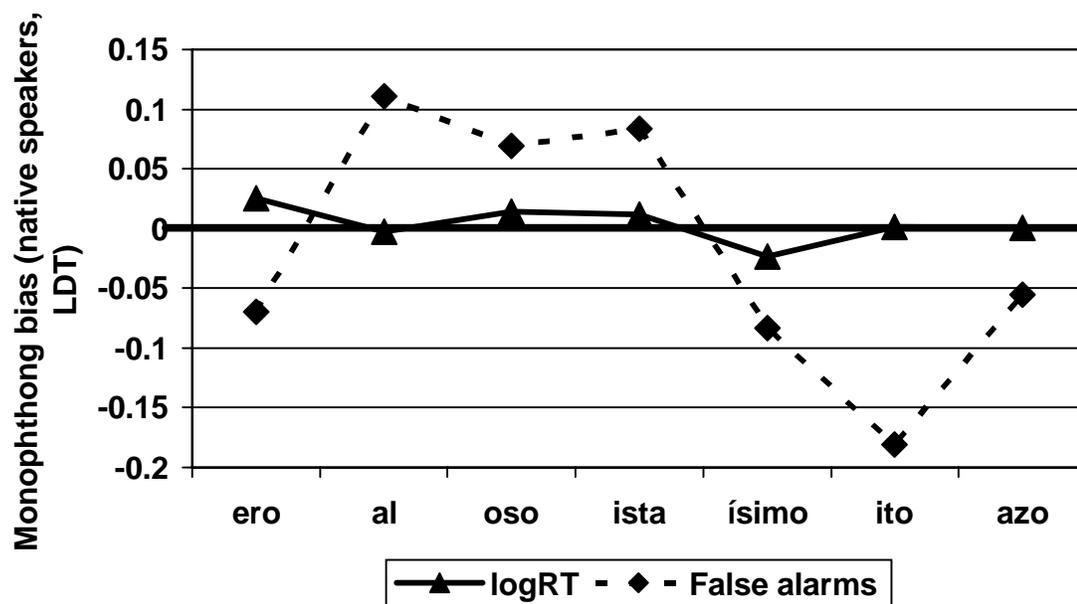


Figure 5-3: Monophthong bias for native speakers on the LDT, all neologisms

It should be emphasized that the values for monophthong bias represent the difference between the error rate or logRTs for neologisms with diphthongs vs. those for monophthongs, by suffix. As the absolute value grows (i.e. as values become further from zero), the difference between the diphthongized and monophthongized versions becomes greater, and closer to zero the two versions resulted in similar results on the dependent measures. For example, the value in Figure 5-2 for *ito* is strongly negative. This indicates that diphthongs were strongly preferred over monophthongs for this suffix. That is, neologisms like *abiertito* were much more wordlike than their counterparts with monophthongs, in this case *abertito*. Note that the values for monophthong bias give no indication of the actual value of the dependent measure (RT or error rates) for *abiertito* or *abertito*, the bias simply measures the difference in results for these two categories of item. To summarize the interpretation of the graphs showing monophthong bias, values

near zero indicate ambivalence regarding the presence of a monophthong or diphthong in the stem, and values far from zero indicate a strong bias in one direction or another.

The figures illustrating the results of the individual difference measures will incorporate these “monophthong bias” calculations, labeled to indicate whether the bias is based on logRT or false alarm scores. In some cases, individual differences measures revealed an interaction with diphthong status only. In these cases, the bias scores are not used, because they reflect the interaction between diphthong status and suffix.

As a final note regarding this transformation, monophthong bias in logRTs is plotted as a difference between two logarithms, based on the use of logRTs in the ANOVAs to correct the positive skew in the raw RT data. While the significance of the effects may be clearly seen in this transformation, it is difficult to gauge precisely what these values correspond to in milliseconds. This is because a difference in logRTs of 0.02, for example, will vary as a function of the original magnitude of the log latencies. Therefore, the relevant plots are reproduced in Appendix 5 based on actual RTs, in order to indicate the magnitude of the bias in milliseconds.

5.2.2.2 Effects of Spanish proficiency and experience.

Interestingly, there were no significant interactions in the logRT analysis between suffix, diphthong status, and either of the dependent measures from the picture naming task. In the false alarm data, however, there was a significant interaction between diphthong status and Spanish picture naming accuracy, $F(1, 23)=10.054, p<.05$ (this was only marginally significant, $p=.054$, in the reanalysis based on the subset of participants

with valid RTs for all test item conditions, see above). Based on picture naming accuracy, the less proficient learners made more false alarm errors on neologisms with stem monophthongs than on those with stem diphthongs. The more proficient learners made equally many false alarm judgments in both of these conditions, as shown in Figure 5-4.

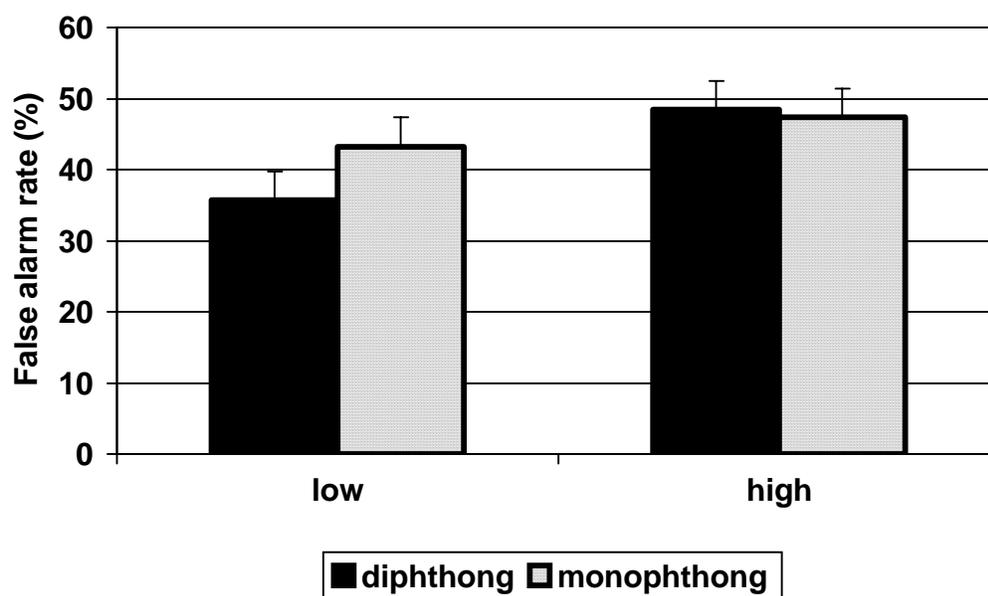


Figure 5-4: False alarm rates by diphthong status in the LDT by learners, by Spanish picture naming accuracy

Despite the lack of proficiency effects in the RT data, there was a significant interaction between suffix, diphthong status, and study abroad, $F(6, 42)=2.821, p<.05$.

The results are illustrated in Figure 5-5.

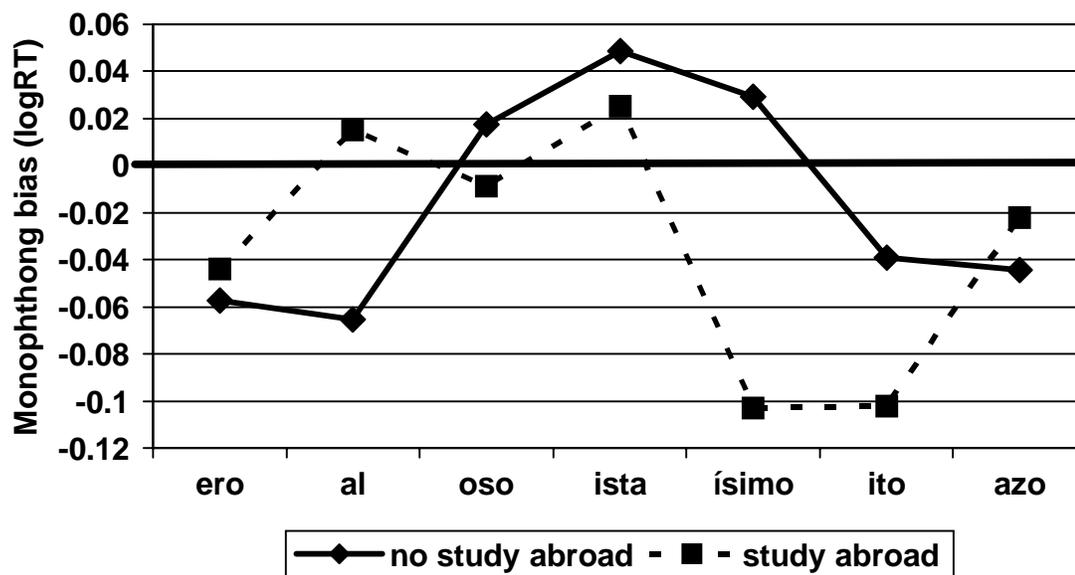


Figure 5-5: Monophthong bias (logRT) by suffix on the LDT for learners with and without study abroad experience

The results for these two groups of learners differ primarily with regard to the slight monophthong bias for *al* and the strong diphthong bias for *ísimo* and *azo* in the study abroad group. Apart from the diphthong bias for *ero*, therefore, learners with study abroad experience show a more marked diphthong bias for suffixes that cooccur more frequently in the lexicon with unstressed diphthongs, and a more stable, albeit weaker, monophthong bias for the suffixes at the other end of the scale. Given this observation, it appears that the deviation from predictions in the RT data for *al*, shown in Figure 5-1, was due to the behavior of the group without study abroad experience. It thus appears that the data from the study abroad group better match the predictions based on earlier data.

5.2.2.3 Individual differences in cognitive resources

There were no effects of memory span, attentional control, or either measure of foreign word repetition for the false alarm rates. In the RT data, on the other hand, there were significant interactions between suffix, diphthong status, and accentedness in repetition of Korean words, $F(18, 42)=1.931, p<.05$, and between suffix, diphthong status, and memory span, $F(6, 42)=4.361, p<.05$. The interaction with Korean accentedness is reported in Figures 5-6 and 5-7. Since this measure contained four levels, the two more accented groups (i.e. those with less success in reproducing a native Korean accent) and the two less accented groups are separated for clarity. Note again that these learners had no prior experience with Korean, and thus this accentedness measure is an indication of their ability to accurately mimic the words they have just heard.

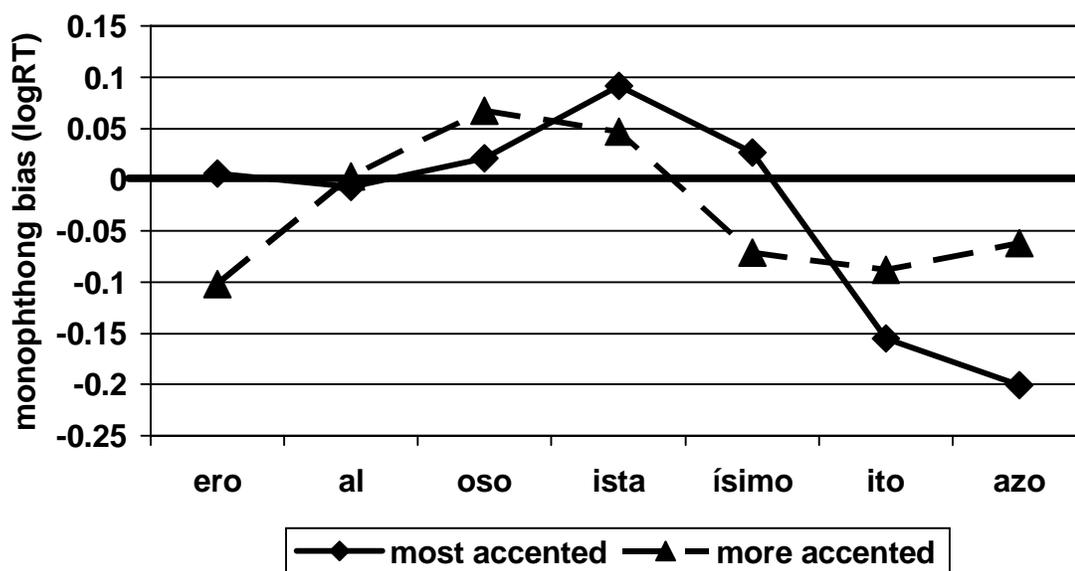


Figure 5-6: Monophthong bias (logRT) by suffix on the LDT for learners with greater accentedness in Korean word repetition

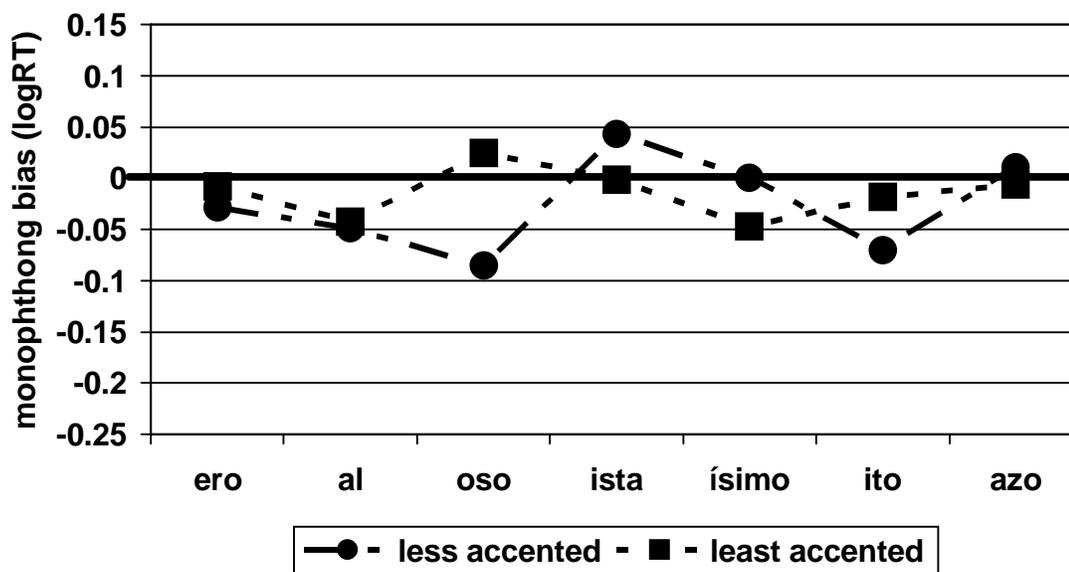


Figure 5-7: Monophthong bias (logRT) by suffix on the LDT for learners with less accentedness in Korean word repetition

In general, those learners who repeated Korean words with a stronger foreign accent (i.e. those who were less accurate in mimicry of the Korean words), as judged by Korean native speakers, more closely match the trend predicted by Eddington's corpus and behavioral results. The results for *ero* and *al* are more neutral, but apart from these, there is a relative monophthong bias for *oso*, *ista*, and sometimes *ísimo*, and a relative diphthong bias for *ísimo*, *ito*, and *azo*. Note that a difference in logRTs of .02 in the present data corresponds to a difference of nearly 100ms.

This trend of decreasing monophthong bias can be seen for the suffixes *ista*, *ísimo*, and *azo*, albeit more weakly, in the less accented group (i.e. the more accurate mimics) in Korean word repetition. In contrast, their biases tend more towards diphthongs for *ero*, *al*, and *oso*. This does not constitute a strong reflection of the native

speaker results, but it should nonetheless be noted that *ista*, *ísimo*, and *azo* perform the most consistently across tasks with native speakers, and across learner variables. This is perhaps due to the relatively greater productivity of these suffixes.

The interaction between suffix, diphthong status, and memory span is shown in Figure 5-8. The contrast between the two groups is subtle, consisting of greater diphthong bias in the cases of *ero* and *ito* for the high span group. Again excluding *ero*, the contrast between suffixes with a lexical bias for monophthongs and those with a lexical bias for diphthongs is slightly stronger for the high span group. Nonetheless, note that neologisms ending in *ero*, *al*, and *oso* deviate markedly from the predictions.

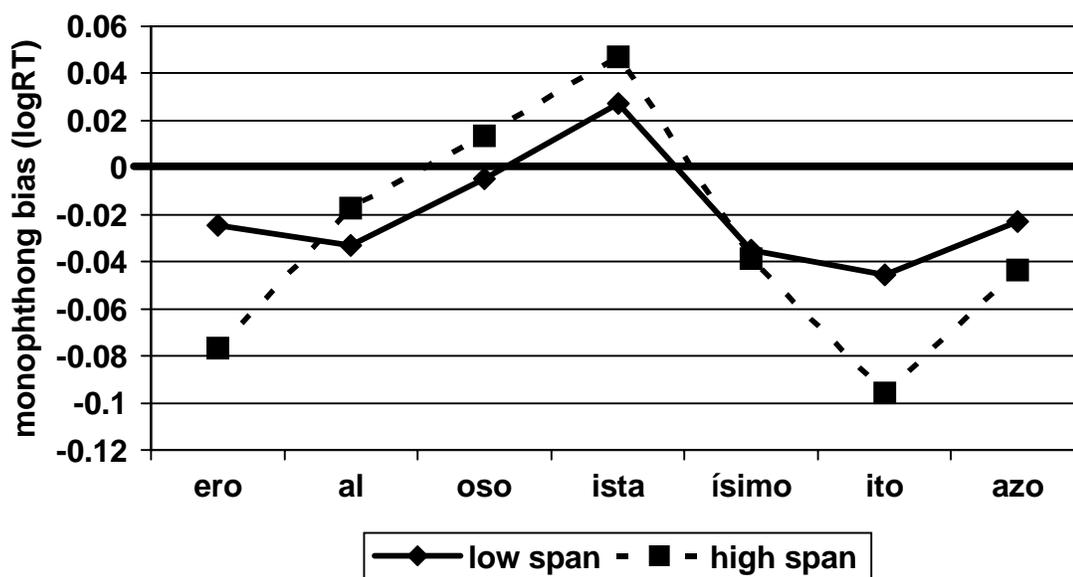


Figure 5-8: Monophthong bias (logRT) by suffix on the LDT for low and high memory span groups

We now examine the results from the language production experiment before turning to a general discussion of the results.

5.3 Results for adult learners: Conditional naming task

5.3.1 Diphthongizing stems in neologisms with derivational suffixes

As for the perception experiment above, this section begins with a presentation of the interaction between suffix and diphthong status in general. Effects of proficiency are presented next, and finally, the effects of individual cognitive differences are given.

The false alarm rates in the production experiment in general were slightly higher than for the perception experiment (between 45 and 60% for each suffix). It was also possible to detect errors in which the test stem or suffix was not reproduced correctly, suggesting that the stimulus had not been perceived correctly to begin with. Since the experimental manipulation hinged on the specific properties of real stems and affixes, trials on which such errors occurred were removed from analysis (see Section 4.3.1).

The combined effects of these two outcomes served to exacerbate the problem of empty cells in the RT matrix (although it does not create problems for the false alarm analysis). While the analysis might be attempted based on the RTs for incorrect trials as opposed to correct trials, the accuracy rate is such that the situation is not improved. Therefore, the three suffixes with the most empty cells, *ista*, *ísimo*, and *azo* were excluded from the RT analysis, leaving 15 participants with valid RTs for neologisms ending in *ero*, *al*, *oso*, and *ito*. The first three of these suffixes were predicted to have a monophthong bias, and the latter a diphthong bias, and so it is still possible to test the hypothesis. However, the scope of RT results in this experiment is much more limited, especially given the behavior of the first three suffixes in the perception data.

Based on this reduced set of suffixes, there was a trend toward shorter response latencies with neologisms containing unstressed diphthongs for *al* and *oso*, and an opposite trend for *ito*. This interaction was not significant, but there were some three way interactions that will be discussed in the following sections. There was also a significant main effect for diphthong status, $F(1, 4)=8.662, p<.05$. Neologisms with stem monophthongs were responded to significantly slower than those with diphthongs.

If this is taken to reflect the general Spanish bias against unstressed diphthongs, this result appears to contradict the facilitation effect for greater wordlikeness observed in the perception experiment for both native speakers and learners. However, there was some indication of a simultaneous inhibitory effect for wordlikeness in conditional naming for the native speakers (see Section 4.4.1). In the present task, learners also made significantly more false alarm errors for neologisms with a stem monophthong (see below). The mean log RT and false alarm rates by diphthong status are plotted in Figure 5-9. These facts support the interpretation that the higher wordlikeness of unstressed monophthongs in alternating stems leads to longer RTs for neologisms in conditional naming, based on the general, but not suffix-specific, pattern in Spanish. That is, wordlikeness has an inhibitory effect. By this interpretation, the nonsignificant trends in the suffix by diphthong status interaction in this task are also in the expected direction.

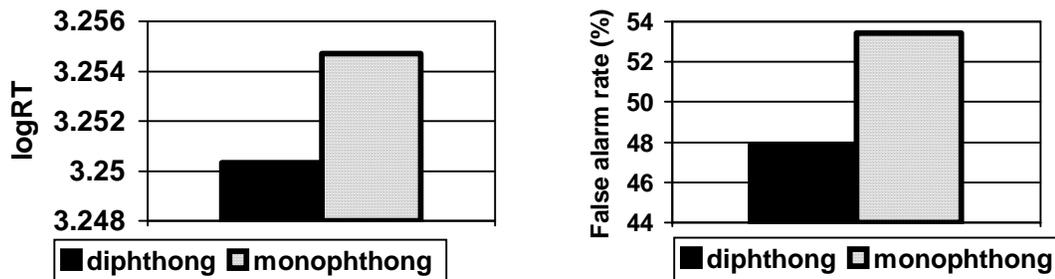


Figure 5-9: Mean logRT and false alarm rates for neologisms in conditional naming by adult learners, by diphthong status

The analysis of false alarms in the production experiment with learners revealed significant main effects for suffix, $F(6, 138)=2.512, p<.05$, as well as diphthong status, $F(6, 23)=5.563, p<.05$. The former finding reflected relatively slower response latencies for neologisms ending in *ero*, *ista*, and *ísimo*, and slightly faster latencies for *ito* and *azo*. The effect for diphthong status, as stated above, showed that there were more false alarms when neologisms contained stem monophthongs than when they contained unstressed diphthongs. Given the expected overall bias in Spanish this is not unexpected.

The interaction between suffix and diphthong status was marginally significant, $F(6, 138)=2.102, p=.057$. However, this appears to stem from the slightly higher false alarm rates for neologisms with stem monophthongs with the suffixes *ista* and *ísimo*. There is also a slight diphthong bias for *ito*. If the learners in this study are sensitive to the gradient bias towards diphthongs or monophthongs depending on derivational suffixes, this sensitivity does not appear to affect their accuracy in naming neologisms. Although this result is only marginally significant, it is illustrated in Figure 5-10.

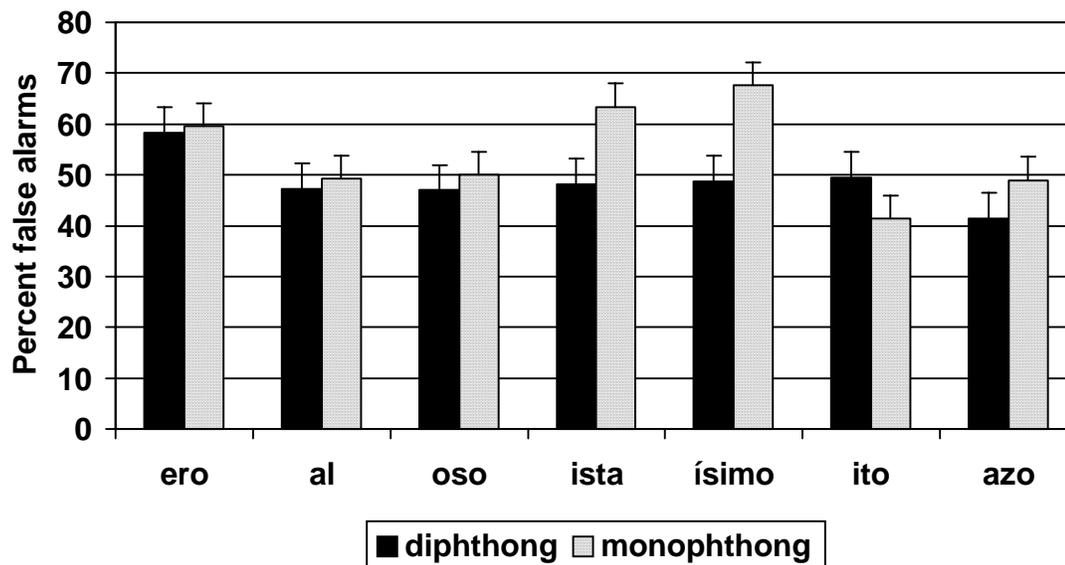


Figure 5-10: False alarm rates for neologisms in conditional naming by adult learners

The analysis of false alarms was run a second time, as for the LDT, using only those participants with valid RTs for *ero*, *al*, *oso*, and *ito*. That is, based on the same subject group as the RT analysis. This was to test that the comparison between the results on these two dependent measures was valid. The results from the false alarm analysis were already somewhat weaker than those from the RT analysis, indicating that the diphthongization pattern tested here more strongly affects the timecourse of processing than a behavioral judgment about lexical status. The marginally significant interaction between suffix and diphthong was again nonsignificant in the reduced group of subjects, and the interaction with lexical fluency (see below) was no longer significant in the reduced subject pool. However, the effects for diphthong and associated interactions with individual difference measures were shown to be stable in the reduced group (below).

Finally, there were several trials on which participants reproduced diphthongized neologisms as though they contained monophthongs and vice versa. While the number of

such errors as a proportion to the total number of trials is extremely small, it bears mentioning that there were 72 reversals in which diphthongs were reproduced as monophthongs, and one reversal in which a monophthong was reproduced as a diphthong. This supports the conclusion that neologisms with unstressed diphthongs are in general less wordlike than those with stem monophthongs as measured on the conditional naming task. The number of reversals by suffix is given in Table 5-1.

Table 5-1: Reversals of diphthong status in conditional naming by adult learners

Suffix	Reversal direction	Number of reversals
<i>ero</i>	D(ipthong) → M(onophthong)	6
<i>al</i>	D → M	8
<i>oso</i>	D → M	9
<i>ista</i>	D → M	9
	M → D	1
<i>ísimo</i>	D → M	10
<i>ito</i>	M → D	12
<i>azo</i>	D → M	18

5.3.2 Individual differences between learners

As for the LDT, three-way interactions between suffix, diphthong status, and a between-subjects measure will be reported in terms of monophthong vs. diphthong bias. However, because of the overall inhibitory effect of wordlikeness in conditional naming, the calculation as made above would yield opposite results. For example, RTs for variants with monophthongs are expected to be higher than for those with diphthongs, for suffixes with a monophthong bias. Calculating the bias by subtracting the monophthong RT from the diphthong RT would then yield a negative number reflecting that

monophthong bias, where in the LDT a monophthong bias would have been reflected with a positive number. This way of calculating bias was not changed in the following figures, in order to reflect opposing predictions of facilitation and inhibition for the LDT and conditional naming tasks, respectively. Instead, the y-axis labels on the relevant plots have been changed to read “diphthong bias”. Note, however, that this reversal does not in any way affect the data for false alarms. As before with the false alarm data from the LDT, positive numbers indicate monophthong bias, and negative numbers diphthong bias. The figures based on false alarm data are labeled to indicate this.

5.3.2.1 Effects of Spanish proficiency and experience

Spanish proficiency as indicated by the picture naming measures played a comparatively larger role in the conditional naming task than it did in the LDT. In contrast, there were no significant effects for study abroad in conditional naming.

With regard to the analysis of logRTs from conditional naming, there was a significant interaction between suffix, diphthong status, and both picture naming measures. For accuracy in Spanish picture naming, $F(1, 4)=17.095, p<.05$, and for efficiency in Spanish lexical access (controlling for speed in English picture naming), $F(1, 4)=32.397, p<.05$. The pattern of results in each case is shown in Figure **5-11**

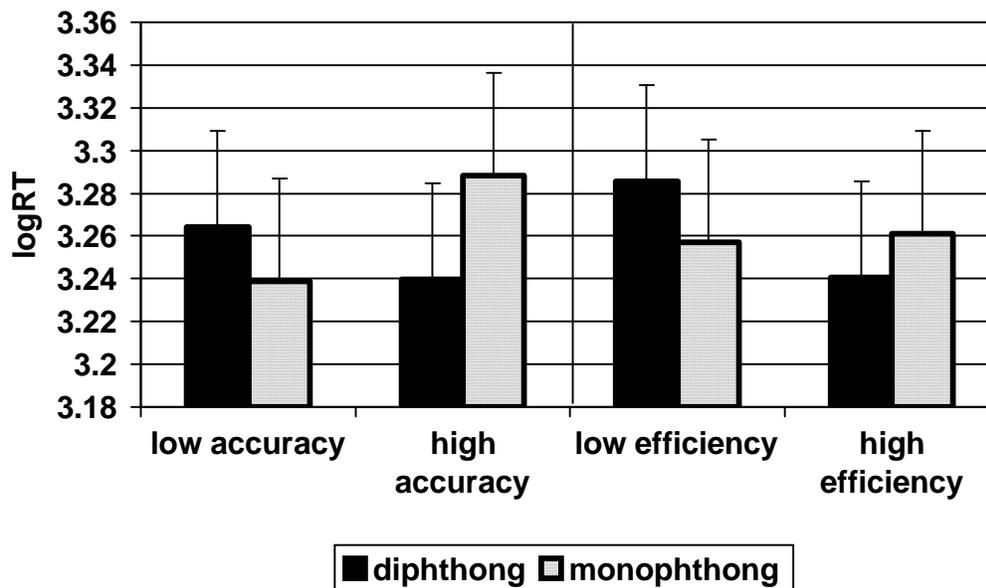


Figure 5-11: logRTs by diphthong status for neologisms in conditional naming by adult learners, by Spanish picture naming accuracy and efficiency

At first glance, this seems like an anomalous pattern of results. Earlier analyses indicated an overall inhibitory effect for wordlikeness in language production measured on this task, although there is likely to be facilitation for the lexical decision component of the task (see Section 4.4.1). The results shown here suggest that the less proficient learners are either ambivalent or prefer diphthongs overall. A diphthong bias for low proficiency learners would be inconsistent with the LDT results (see Figure 5-4), although relative ambivalence (note the error bars) would not be wholly unexpected. However, the solution may have more to do with the reduced conditions under which the logRT analysis is carried out, namely, that only four suffixes could be included: three with a predicted monophthong bias, and the other with a predicted diphthong bias.

The trends in monophthong vs. diphthong bias in the conditional naming RTs, while not significant, help to clarify this issue. The relevant data are plotted in Figures 5-

12 and 5-13. Note that the low accuracy group shows relatively neutral bias for neologisms ending in *ero*, *al*, and *oso*, and a diphthong bias for those ending in *ito*. This is in keeping with the hypothesis that *ito* is more likely to occur with an unstressed stem diphthong than are the other suffixes. More relevant is the observation that this pattern of biases implies a higher mean RT for diphthongs than monophthongs. In other words, the low accuracy learners prefer diphthongs where we expect them to, but their ambivalence with the remaining suffixes results in the bias for *ito* contributing to a slight diphthong bias overall for this group. The results for the lower efficiency group in Spanish picture naming show a varying diphthong bias for all suffixes, which is stronger for *ito*, but fairly weak for *oso*. Recalling however that the results for *ero* and *al* are relatively unstable in other analyses, including those for native speakers, it seems that the apparent diphthong bias for this group is not a reflection of learners' knowledge of Spanish phonology, but rather of the particular suffixes included in the analysis.

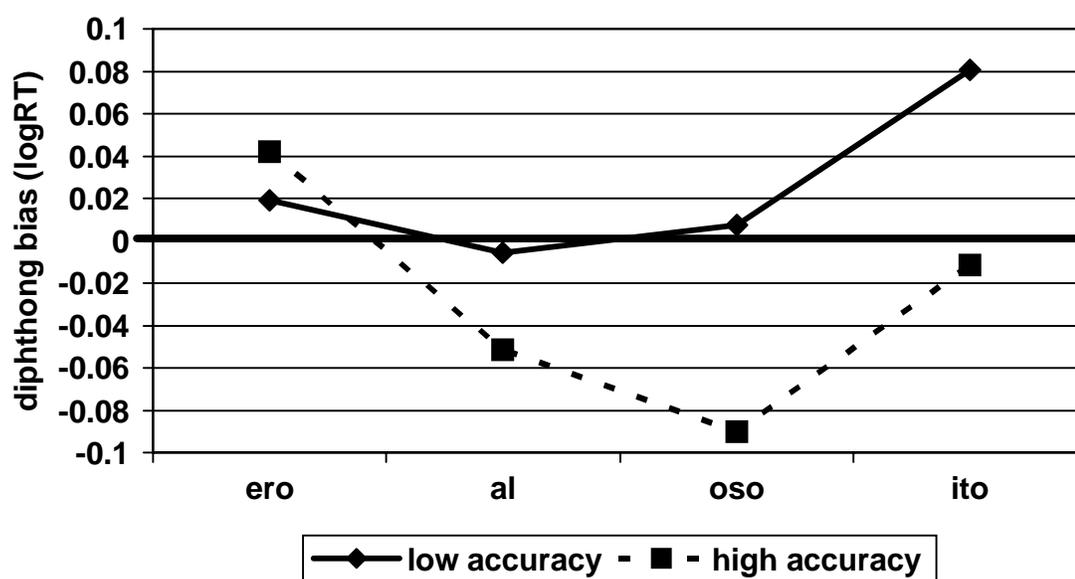


Figure 5-12: Diphthong bias (logRT) by suffix for learners on the conditional naming task with low and high accuracy in Spanish picture naming.

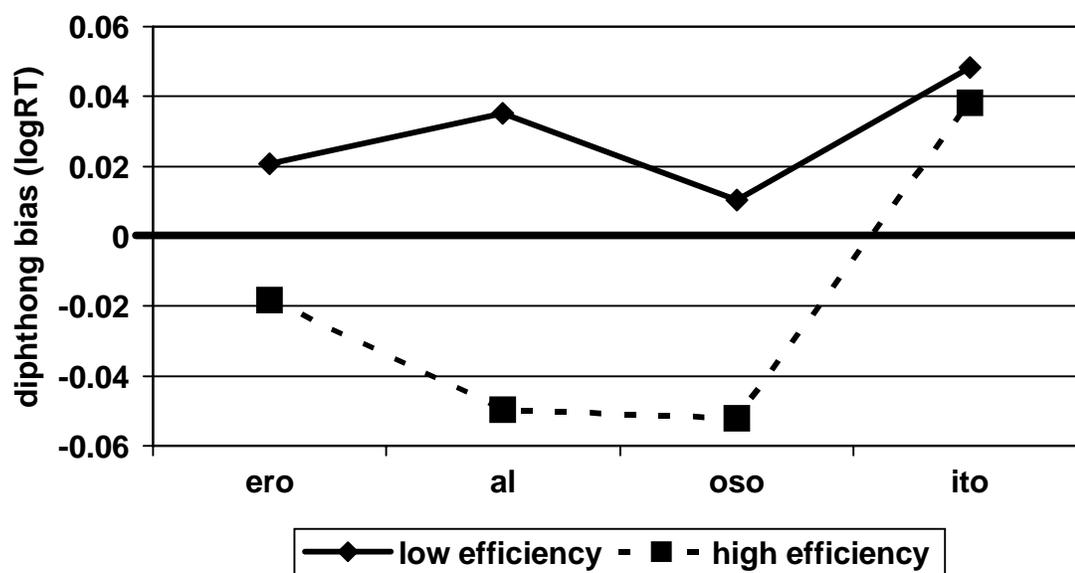


Figure 5-13: Diphthong bias (logRT) by suffix for learners with low and high efficiency in Spanish picture naming

Turning our attention to the higher proficiency group in both of these cases, learner performance actually appears to reflect the original hypotheses based on earlier corpus and behavioral data, except in the case of *ero*, which has consistently behaved contrary to those predictions. Moreover, the high monophthong bias for two suffixes is likely to outweigh the neutral to diphthong bias for *ito* and *ero*, resulting in overall longer response latencies for neologisms with monophthongs.

There may be another factor that also serves to increase RTs for neologisms with stem diphthongs for the less proficient learners. Both picture naming measures reflect lexical fluency to a certain extent, especially in the case of reaction times. This efficiency

may affect learners' success in identifying stems during this task, prior to hearing the suffix. Further, some stems may be easier to identify than others. In particular, a stem in diphthongized form contains more information relating to the identity of the stem and is, in effect, a stronger lexical cue. On the other hand, stems in monophthongized form may be more difficult to identify, particularly for learners with less efficient lexical access in Spanish. This difference in cue strength may not have any effect on native speakers' or more proficient learners' processing, which is already highly fluent, but it may affect less proficient learners to a greater extent. For these learners, the diphthongized stem may more strongly prime a *palabra* response. When it is concluded that the neologism is not a real Spanish word, it may take longer to overcome this priming and articulate a repetition of the item instead. For learners with greater lexical fluency, either form of the stem may be just as likely to prime a *palabra* response. The result of this difference would be that reaction times to neologisms with stem diphthongs would be inflated overall for lower proficiency learners, on the measures used here.

We turn now to the other significant interaction involving proficiency in the conditional naming task. The three-way interaction between suffix, diphthong status, and efficiency in Spanish picture naming was significant in the analysis based on false alarms, $F(6, 138)=3.337, p<.05$. The trends are shown in Figure 5-14. The most salient differences are in the biases for neologisms ending in *al* and *ista*. The high efficiency group shows a more natively like pattern of behavior for these suffixes, with the low efficiency group showing a strong diphthong bias for *al* and a strong monophthong bias for *ista*.

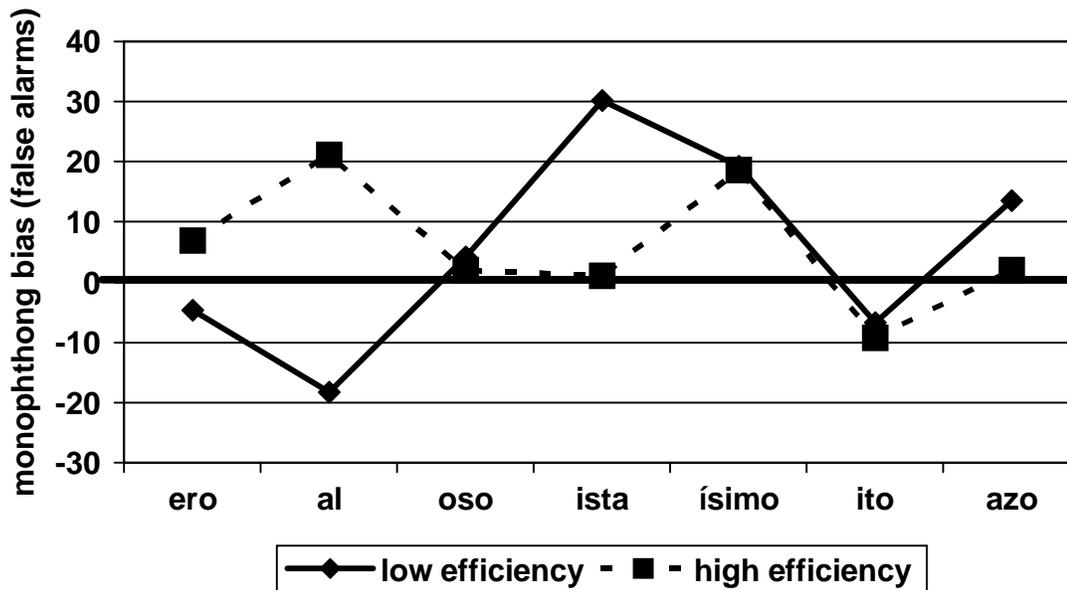


Figure 5-14: Monophthong bias (false alarms) for neologisms in conditional naming by adult learners with more or less efficient Spanish picture naming scores.

This is in keeping with the notion that *al* may be difficult to identify due to its length. Learners with more fluent lexical access in Spanish would be more likely to correctly identify it. Having identified it, its inherent bias in the Spanish lexicon is more likely to affect their processing than that of the less fluent group. Learners with less efficient lexical access will not be as likely to recognize *al*, unless perhaps they are already expecting a real word. This is more likely to be the case with a stem diphthong, because the diphthong would provide a stronger cue to stem identity. This combination of factors would then lead to higher false alarm rates for neologisms with stem diphthongs.

The picture is less clear in the case of *ista*, but note that the behavior of *ista* in both Eddington's results and the current study with native speakers has suggested either a neutral bias or a monophthong bias for this suffix. Furthermore, *ista* is a highly productive suffix that is likely to constitute a strong cue to its identity as a morpheme for

both proficiency groups. The strong monophthong bias for the less proficient group may then be related to the more general preference for unstressed monophthongs in cases where alternation is possible.

5.3.2.2 Individual differences in cognitive resources

There were significant interactions for three of the four cognitive measures in the conditional naming experiment: Korean word repetition accuracy (proportion of words correctly repeated), memory span, and attentional control. The results for these will be presented in this order.

There was a significant three-way interaction between suffix, diphthong status, and accuracy in Korean word repetition for the logRT analysis, $F(3, 12)=5.073, p<.05$, but not for the false alarm analysis. The data are plotted in Figure 5-15. Both groups show the expected trend towards a relatively higher diphthong bias for neologisms ending in *ito* than for the other suffixes. Apart from this, the group with higher scores in Korean repetition showed a relatively stronger monophthong bias for *ero* and *oso*, also in line with the original predictions, but those with lower scores showed a neutral or relative diphthong bias for these suffixes. It would appear from these data that higher phonological memory as measured through accuracy in foreign word repetition is associated with more nativelike performance in conditional naming of the test neologisms.

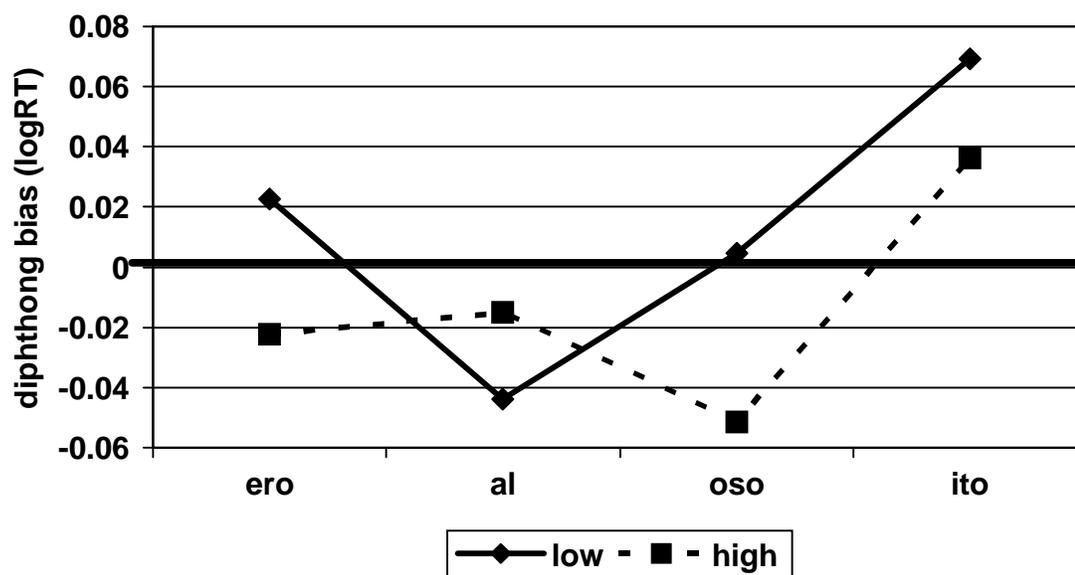


Figure 5-15: Diphthong preference (logRT) in conditional naming by adult learners as a function of accuracy in Korean word repetition

In the logRT analysis, there was also a significant interaction of diphthong status with memory span, $F(1, 4)=20.979$, $p<.05$, but there was no three-way interaction. This interaction was expressed as a very slight increase in monophthong bias for the high span group over the low span group, for which RTs were only very slightly higher. The reason for this is unclear, and it is surprising that such a small effect would be reliable, especially since, despite a much larger difference in RTs between the low and high memory span groups, the main effect of memory span was not significant ($p=.2$). Further work with a greater focus on the role of memory span may help to determine if this effect is indeed replicable, and if so, what it signifies.

There were two significant interactions involving attentional control as measured on the Simon task. First, in the logRT analyses, there was a significant three-way interaction between suffix, diphthong status, and the Simon effect, $F(3, 12)=6.762$,

$p < .05$. As shown in Figure 5-16, this consisted primarily of a high diphthong bias for *al* for the high attentional control group vs. a high monophthong bias for the low attentional control group. Based on the prior analyses and on the minimal phonemic content of this suffix, *al* has often appeared to be unique among the suffixes used in the current experiment. In this case, attentional control may have an effect similar to that of lexical fluency as measured on the picture naming task. If the brevity of *al* provides only a weak cue that it is indeed a real morpheme, then learners with higher attentional or inhibitory control may more effectively suppress any activation of candidate forms due to the hearing of this suffix. The cue strength for lexical status of the stem may then have a greater effect on RTs than that of the neologism as a whole, and in this case, stems with diphthongs contain more information regarding their identity than stems with less phonetic material, namely monophthongs.

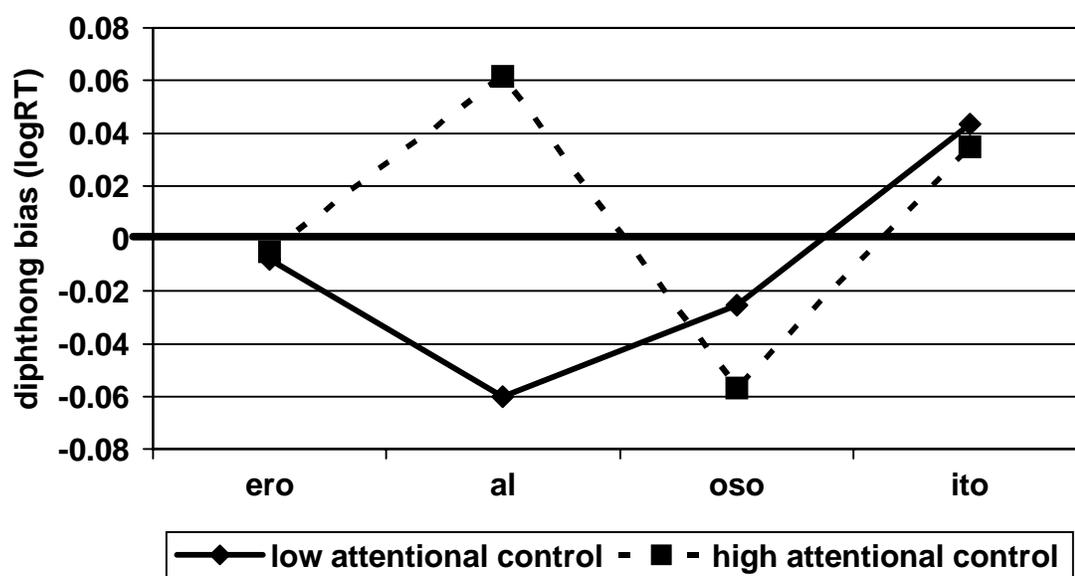


Figure 5-16: Diphthong bias (logRT) for neologisms in conditional naming for adult learners varying in attentional control ability

There was also a significant interaction between diphthong status and attentional control for the false alarm data, $F(1, 23)=6.115, p<.05$. Learners with lower Simon effect scores for attentional control made more false alarm errors on neologisms with stem monophthongs than on those with stem diphthongs, whereas learners with higher scores showed no difference in false alarms based on diphthong status. Based on this result, it is possible that learners with lower scores are more likely to follow the general pattern, that is, to regularize the diphthong alternation across suffixes following the most frequent rule that the removal of stress favors the surfacing of a monophthong.

These cognitive measures paint a complex picture in which, in some cases, less is more, and in others, more is more. We now turn to a more general discussion of these and other effects across the present experimental tasks.

5.4 Discussion

5.4.1 The acquisition of gradience in diphthongization

The data for adult learners offer a complex picture of their sensitivity to the gradient well-formedness of unstressed diphthongs in neologisms with varying suffixes, given stems with alternating diphthongs. In general, there was a significant main effect for the interaction of suffix and diphthong status in response latencies in lexical decision, but this effect on its own was not significant in conditional naming RTs or false alarm rates on either task, (the effect was marginally significant for false alarms in conditional naming). This effect was modulated by several individual differences, including study

abroad, Spanish picture naming performance, foreign word repetition, memory span, and attentional control. However, the various measures affected performance to different degrees and in different ways depending on both the experimental task (LDT or conditional naming) and the dependent measure (logRT or false alarms).

Of the two experimental tasks, the LDT was the simpler, because the conditional naming involved the same lexical decision, but added a speech production component. It is therefore not entirely surprising that the results should be somewhat clearer in the LDT, at least in response latencies. The fact that, within this task, the RT analysis yielded a significant effect and the false alarms did not is interesting. It suggests that the differing probability of an unstressed diphthong with the various suffixes has an impact on learners' processing of the test neologisms, but that their sensitivity to this effect cannot yet be detected on a behavioral measure such as decisions regarding lexical status. This differs from the pattern obtained for native speakers, where the stronger indications were for observed in the false alarm data.

The suggestion that second language acquisition may affect processing before behavior is not without support. It is intuitively appealing that changes in processing will occur leading up to and beyond the point where lexical items or grammatical forms are represented strongly enough to be accessed reliably. Furthermore, McLaughlin et al (2004) present neuroimaging evidence from beginning learners of French showing brain responses to French words before the words could be reliably recognized by the learners behaviorally.

It is also interesting to note that reaction times reflect sensitivity to the subpattern in Spanish diphthongization for the whole group on the LDT, but that the effect is only

reliable in conditional naming for subgroups of the participants based on various cognitive and proficiency measures. Some of these individual differences also affect processing on the LDT, but the interaction between suffix and diphthong status is still significant without taking them into account for this task, as opposed to conditional naming.

This observation may be explained by the increased processing load involved in producing speech on the conditional naming task. Some learners are better at dealing with these increased processing demands than others, as reflected in the significant interactions with, in particular, Korean word repetition and attentional control, as well as Spanish proficiency. While these differences still affect processing in lexical decision in which the response is made by button press, without any speech articulation, the lower demands placed on learners for this task results in more learners processing the neologisms according to more nativelike lexical biases in Spanish.

False alarm rates, on the other hand, do not reliably reflect Spanish lexical biases regarding diphthongization in derived words in the learners' performance. The effect is marginal in the conditional naming task, and the patterns for some suffixes show trends in the expected direction. It may be that the neuromotor programs underlying speech production respond to the lexical biases for some suffixes, but the effect is weak. Decisions about the lexical status of an item, however, are behavioral measures even in speeded tasks such as those used here. As such, they may rely on many factors apart from lexical biases, especially for second language learners.

A further observation that may help to explain why false alarm data were less consistent is that learners made a high proportion of errors, in some cases performing at

chance, whereas the error rates for native speakers were much lower. Nonetheless, both learners and native speakers were less accurate on the test neologisms than on either real word or nonce word distractors. The relatively high error rates on the neologisms even for native speakers suggests that they too were sensitive to the strong lexical cues provided by the use of real stems and suffixes in the neologisms, and in many cases concluded that they were real words. Since the neologisms themselves were semantically plausible, and since many of them were phonologically well-formed, many of the neologisms were actually very good potential words and might be created at any time.

The comparatively high false alarm rates for native speakers on these items versus the distractors suggests that tasks using neologisms may help to clarify the notion of a word. One could equally make the case that many of the neologisms, though not attested in large corpora of Spanish and perhaps never before encountered by a speaker, did in fact merit classification as words, being perfectly acceptable in spite of not activating a specific lexical representation. This is highly consistent with a probabilistic and usage-based notion of grammar. Bybee in particular (2001b) notes that memory traces decay if they are not frequently accessed. While in an exemplar-based model such as that described by Bybee (see also Pierrehumbert, 2001a) words are stored whole, they are also linked to related words based on semantic, phonological, and other kinds of similarity. The intact representation of infrequent words may decay over time, but the links across the lexicon are more durable because they are relied upon by other lexical items as well. The native speakers may thus have treated some of the neologisms in this experiment as infrequent words based on their activation of similar material in the lexicon, even while on the whole their judgments on neologisms were relatively accurate.

For second language learners, words in the L2 are much less frequent to begin with, and learners may more often rely on connections with similar stored material in their developing L2 (as well as L1) lexicon. In addition, learners are far more likely to encounter words they have never heard before, and, based on normal communicative situations, they are generally forced to conclude that these are, in fact, real Spanish words (it would be very odd for a learner in any L2 interaction, at least with an instructor or more proficient speaker, to assume that their interlocutor's utterances were sprinkled liberally with invented words). For this reason, they are likely to make use of guessing and other strategies that could have unpredictable effects on the experimental manipulation, in addition to increasing their overall false alarm rates.

In other words, the definition of what learners consider to be a real word may have far less to do with whether or not they can come up with a matching memory trace, and more to do with its phonological shape, interpretability, and other factors. It is thus not surprising that their performance on a behavioral measure such as lexical decision accuracy should be less consistent than might be indicated in a processing time measure.

Returning to those effects found to be significant, learner performance, especially in response latencies on the LDT, is strikingly similar to the predictions made for native speakers based on the results of Eddington (1996; 1998) and from the native speaker controls in the present experiment. This is all the more impressive given the complex morphological, phonological, and paradigmatic patterns that underlie diphthongization in Spanish derived words.

There were nonetheless, as is to be expected for learners, some discrepancies. For example, while the diphthong bias for *azo* was fairly stable in the native speaker data

(although its behavior in native speaker reaction times for conditional naming is puzzling), it was relatively weak compared to the strong bias indicated in Eddington's behavioral results (recall that the dictionary data for productive suffixes such as this are less reliable). Further and more detailed corpus-based research may help to bring the precise pattern of lexical bias in Spanish into greater focus, as well as to determine the exact level of detail to which those biases are productive.

Moreover, as discussed earlier, there are major differences in the experimental tasks used here and in Eddington's study. Numerous factors may contribute to variation in the results of written, forced-choice tasks, as he used, and online auditory tasks involving lexical decision and speech production as were used here. Implications of these differences will be discussed further in the next chapter.

In addition to the variability in the strength of biases shown in the present study, there were several suffixes that contrasted more markedly, and sometimes inconsistently, with the predictions. Before exploring individual cases, however, it should be noted that diphthong patterns in Spanish derived words do not constitute a unitary grammatical feature. Derivational suffixes are acquired at different times, and particularly given the reduced language experience of university students of L2 Spanish (even on study abroad), the probability that learners have encountered enough instances of alternating stems with each suffix for the pattern to emerge is relatively low. Furthermore, even if a word such as *herrero* 'blacksmith' were to be encountered by a learner, there is no guarantee that the word would initially be processed as containing two morphemes, *hierro* 'iron' and *ero*.

Because of the lack of cohesiveness of gradient subpatterns in diphthongization, the overall pattern is likely to be acquired in a gradual and piecemeal manner over time (Ellis, 2002a). It is therefore unsurprising that learners should deviate from the native speaker predictions in some instances while conforming in others.

This piecemeal learning of grammatical structure is central to an emergentist or usage-based view of language acquisition. Grammatical structure depends on an individual's experience with language, even for highly regular structures (Tomasello, 2003), and language experience is a probabilistic phenomenon. It is only as the statistical robustness of patterns increases (Pierrehumbert, 2003a, 2003b) that those patterns may begin to function to constrain both the perception of new experiences and the production of novel utterances. This fragmentary and piecemeal acquisition of linguistic structure in L1 is also likely to be the case for L2 acquisition, based on the emergentist hypothesis that linguistic knowledge is the same as knowledge in other domains. As experience continues to accumulate it is structured in memory, and that structure is then applied to the perception and production of new L2 experiences as well.

It is, however, not enough to say that we expect some suffixes to deviate from the predicted pattern, as shown in the present results, based on piecemeal acquisition. It is also necessary to explain why some suffixes deviate and others do not. This is possible in the present study to a certain extent, but detailed analysis of appropriate corpora of learner experience will be necessary to make firm predictions about the behavior of learners at different stages in the acquisition of Spanish.

Beginning with the most consistent data for the learners, the RT data from the LDT, we note that the suffixes *ero* and *al*, and to a lesser extent *oso* and *azo* differ

somewhat from the results in Eddington (1998). We have already noted that *azo*, which shows a somewhat weaker diphthong bias in these results, also shows a weaker bias in the present experiment with native speakers. While learner behavior on this suffix differs in degree (but not in polarity) from Eddington's results, it is entirely consistent with the behavior of native speakers in auditory lexical decision. The contrast with Eddington's data is not entirely unexpected, because the forced-choice paradigm may exaggerate what are in fact small but reliable probabilistic effects. Since we have no corpus data to go on in this case, a clearer answer in this case must await future work.

It is also worth noting that neologisms ending in *ero* and *al* behaved somewhat erratically in the native speaker data as well. In the LDT, *al* showed an anomalous diphthong bias in the RT data (only when excluding neologisms with monosyllabic stems; any bias was nonsignificant across the entire stimulus set), but a monophthong bias in false alarm rates. It showed no bias on either measure in the conditional naming task. Neologisms ending in *ero* showed similar behavior, except that the monophthong bias was shown in the LDT response latencies, and the diphthong bias appeared in the false alarms. It likewise showed no bias on the conditional naming task.

Both of these suffixes are somewhat less productive than either *ista* or the diminutive and augmentative suffixes that show more consistent behavior across tasks and measures. In addition, *ero* occurs at the end of a number of words, such as *carnero* 'ram', in which it either does not represent a distinct morpheme, or its morphology has become opaque through language change. It thus constitutes a weaker cue that the neologism is intended to be multimorphemic.

The lesser productivity of both of these suffixes may also play a role in their acquisition by adult L2 learners. Highly productive suffixes such as diminutives, augmentatives, and *ista* are likely to be more salient to learners because they are highly transparent in meaning and function. In addition to this, more productive suffixes are generally taught explicitly in lower-level language courses, but less productive suffixes may not receive as much attention. For these reasons, learners may be much more aware of the morphological behavior of words containing these suffixes, and as a result acquire their lexical biases more accurately. Furthermore, there is a potential confound in Eddington's corpus data based on the productivity of the suffixes. The most productive suffixes are also those for which the strongest diphthong bias was shown. In effect, it is easier to attach a highly productive and frequent suffix to any root, including phonologically illicit borrowings, acronyms, and other forms (see Pinker, 1998, on the compatibility of regular morphology with phonologically anomalous forms). Since a stem diphthong is a stronger cue to the identity of a stem, neologisms with productive suffixes and unstressed diphthongs may be easier to verify in lexical decision. Alternately, in conditional naming, they may provide a stronger cue for the response *palabra*, initiating the neuromotor programs involved in articulating the response more strongly. Following a decision that the neologism is not real, it would require more time to revise the articulatory program. Unfortunately, there were not enough valid RTs remaining in the conditional naming task, due to the high false alarm rate, to test this latter hypothesis, except to note the somewhat increased RTs for neologisms ending in *ito* when they contained stem diphthongs.

Nonetheless, the behavior of *ista*, *ísimo*, *ito*, and *azo* is relatively consistent across tasks, at least for learners with certain proficiency or cognitive characteristics. The three less productive suffixes may be less familiar to learners, or morphologically opaque, such that learners do not pick up on their lexical biases as quickly or as accurately. Again, we must look forward to more detailed corpus-based research, as well as to a broader sample of learners' Spanish proficiency, especially higher proficiency, to test this hypothesis.

The unique characteristics of *al*, however, bear further comment. As already mentioned, it is the shortest of the suffixes, and therefore contains the fewest cues to its own identity. Furthermore, neologisms formed from *al* and monosyllabic stems constitute disyllabic words, which are more likely to have larger neighborhoods, and thus to be interpreted as monomorphemic. In addition, one of the meanings of *al*, creating an adjective meaning 'pertaining to *x*', while very productive, may also be relatively opaque in morphological structure, especially to learners. For example, words like *natural* 'natural', *normal* 'normal', and *final* 'final' are highly frequent, and are furthermore English cognates, but they are unlikely to strongly activate a *stem+al* representation than less frequent words such as *ocupacional* 'occupational'. In the example of *final* especially, the English word is morphologically opaque (there is no related word *fin* in English, although there is in Spanish), but even given the knowledge that *natural* and *normal* are related to *naturaleza* 'nature' and *norma* 'norm' (the former being particularly transparent, not to mention frequent, in English), their derivations ending in *al* are frequent enough to be accessed whole, without significant activation of either their stems or related derivations (see Meunier & Segui, 1999).

Thus neologisms in *al* were the least likely of the test items to be interpreted as consisting of a real stem plus a real suffix. This is supported by the particular behavior of learners with certain proficiency and cognitive characteristics with neologisms in *al*, as discussed above.

The evidence described here points to an emergentist theory of L2 learning. While some of the dependent measures revealed a categorical preference for unstressed monophthongs over unstressed diphthongs, there was also strong support that this preference varied in strength and polarity depending on specific derivational suffixes. This morphophonological alternation thus cannot be captured in terms of categorical rules based on features such as stress placement, even with the designation of some suffixes as cyclic and others not (Halle et al., 1991; Harris, 1977). In the case of learners, of course, it has long been argued that whatever innate language structure underlies the acquisition of categorical rules is no longer accessible to adults (Bley-Vroman, 1989; Krashen, 1981), although this point of view is not universal (Schwartz, 1986, 1990; White, 2003).

What is striking about the present study is that not only do the processing and behavior of the learners suggest gradient well-formedness of diphthongs in Spanish derivations, but the results for the native speaker group suggest the same result. Furthermore, learners' behavior on neologisms with certain suffixes changes to more closely approximate that of native speakers as they gain in experience and proficiency, but the reverse was not observed. This supports the conclusion that native speakers and adult learners are sensitive to the same information regarding the statistical characteristics of morphophonological structure.

The present results for native speakers line up with a significant and growing tradition of research linking the knowledge of linguistic structure to the statistical information inherent in language use (Bod et al., 2003; Bybee, 2001b; Bybee & Hopper, 2001b; Eddington, 1999, 2004b). This link is crucial to the emergentist hypothesis that the source of probabilistic patterns in the cognitive representation of grammar lies in the continual interaction of general cognitive processes with the constant accumulation of language experience and use. While there are discrepancies between the present results and the corpus patterns described by Eddington (1996), and despite certain weaknesses in the use of dictionary data, the striking parallels between the corpus patterns, metalinguistic judgments, and online processing data argue that knowledge of Spanish diphthongization does indeed reflect lexical statistics.

That lexical statistics are the source of gradience in cognitive representations of grammar, and not *vice versa*, is supported by research showing the item-based nature of child grammar acquisition. Children first acquire grammatical patterns through specific lexical items (including phrases, in the case of syntax), and only later apply those patterns to the production of novel items (Bates & Goodman, 1999; Tomasello, 2000, 2003). It is reasonable to assume that Spanish speakers learn patterns of diphthongization in the same way, and given this assumption, the present data from native speakers therefore argue for an emergentist account of their knowledge of the subpattern in diphthongization examined in this study.

The central hypothesis of this dissertation was that adult L2 learners of Spanish learn this subpattern in the same way as native speakers. If we take the native speaker data as an indication of the trajectory of the statistical patterns in the developing L2

learner lexicon, then the present data support this hypothesis. Learner performance on the experimental tasks was remarkably similar in many ways to the native speaker baseline, and slightly more so on some indicators for learners with greater proficiency and for those who had studied abroad. There were some indications that this gradient pattern affected different kinds of processing for the learners than for natives, but this is not inconsistent with the emergentist hypothesis.

Learner performance was also shown to respond to the general, stress-based rule for diphthongization. This general, categorical pattern indicates a sensitivity among adult learners for large-scale patterns as well as the more subtle patterns described here, but the simultaneous influence of gradient patterns argues for a more complex account of L2 learning. Specifically, emergentist theories allow for the simultaneous representation and operation of multiple levels of grammatical structure (Langacker, 2000), that interact in different ways depending on a variety of cognitive and contextual factors. The research presented here thus reveals a more complex view of adult L2 learning and its relationship to L1 acquisition than might be suggested based on the obvious differences in the process and results of language acquisition at different ages.

5.4.2 The role of proficiency and language learning experiences

Proficiency was assessed via a Spanish picture naming task, yielding both an accuracy measure and a fluency measure based on the RTs controlling for speed of lexical access in English. Study abroad was measured simply as a binary, yes/no variable, because there was little variability in the time spent abroad for those with such

experience. Both of these measures were expected to modulate learners' acquisition of gradience in Spanish diphthongization. Study abroad was hypothesized to facilitate acquisition by greatly increasing learners' exposure to the L2, both quantitatively and qualitatively, in terms of the range of different structures experienced. Picture naming accuracy was thought to provide a relevant index of proficiency because of the relationship of vocabulary size to the statistical robustness of the pattern in question. Finally, lexical fluency was postulated to have an effect by modulating the size of the candidate set during processing of the stimuli, and possibly by playing a role in the establishment of connections between lexical items during the learning process itself.

It is therefore striking that these three measures figured relatively little in the current results. In the case of study abroad, it affected learners' processing of the test neologisms as measured in RTs, but it did not participate in any other interactions across the experiment. Learners who had studied abroad did resemble the native speaker predictions somewhat more accurately (with the exception of the suffix *ero*). However, apart from showing the opposite bias for *al* from the group with only classroom experience, the effect was mainly to exaggerate the decrease in monophthong bias for *ísimo* and *ito*. Apart from these effects, the pattern of results was fairly similar between the two groups. If both groups were relatively similar on the most sensitive measure, the LDT, it is possible that study abroad was not enough to significantly alter learner processing or behavior in more complex tasks or in their false alarm rates.

Learners with study abroad experience had generally spent about one semester abroad in a country where Spanish was the primary medium of communication. In the majority of cases, this also involved studying in an environment with other English

speaking students. Participants were tested at least one month, but in most cases considerably more, after returning to their home university in the US. It is first possible that some of the benefit of study abroad had diminished (Linck, 2005). Second, it is likely that larger-scale changes in processing or behavior with respect to such a fine-grained morphophonological pattern would only occur after longer and possibly more intense immersion in the L2. If the effects of study abroad for participants in the current study are comparatively weak, then it is not surprising that the effects are only detectable in the reaction time measure on the LDT.

Lexical accuracy and fluency as measured in Spanish picture naming played a larger role in the present results, but this was restricted to the other three measures, i.e. false alarm rates on both tasks, and conditional naming response latencies. In the LDT, learners with lower accuracy scores in picture naming made more errors on neologisms with monophthongs, but those with higher scores made equally many errors on both kinds of neologism. This may reflect less dependence on general lexical rules such as the stress-based diphthong alternation for more proficient learners, consistent with the predictions, but this conclusion cannot be confidently made based on the current results.

In conditional naming, it appears that lexical fluency and accuracy did contribute to learner performance, but that this effect was based on lexical processing, rather than on any correlation between vocabulary size and the robustness of the diphthongization pattern. As discussed in Section 5.3.2.1, there was evidence that reduced lexical fluency may have led to differences in the way neologisms with diphthongs were processed, and in the way the suffixes *al* and *ista* were processed. In particular, those structures that constituted stronger cues to the identity of a stem or suffix (e.g. diphthongs, which

increase the amount of phonetic information available for lexical identification) may have increased response latencies or false alarm rates, and those with weaker cues (e.g. *al*, which is phonologically minimal and possibly less salient to learners) either had contrasting effects or served to accentuate cues based on the structure of the stem.

While the effects on lexical processing are interesting, the picture naming measures did not provide convincing evidence that some participants had acquired the subtleties of Spanish diphthongization to a greater extent than others. In this respect it appears that the picture naming task either did not provide an adequate measure of the type of proficiency that might modulate knowledge of the lexical biases of suffixes, or it was not sufficiently sensitive. There is some evidence to support at least the latter possibility. The pictures were selected to represent words from two frequency levels in Spanish. However, learner accuracy for the less frequent items was extremely low. Accuracy for the more frequent items was more variable, and clearly sufficient for the observation of some effects in these experiments, but the inclusion of less difficult items would likely have made the test more sensitive.

5.4.3 The role of cognitive resources

Of the three types of cognitive resources measured in the present research, phonological memory as indexed by accuracy in both reproducing the correct sequence of phonemes and in mimicking a nativelike accent in an unfamiliar language (Korean) played a larger role. The Korean word repetition measures interacted with suffix and diphthong status, modulating response latencies in both tasks, but there was no

interaction between these variables affecting the false alarm rates. The comparatively greater involvement of phonological memory in this experiment is consistent with both the acoustic nature of phonological grammar (although much learner experience is likely to be in the written medium, at least in classroom environments), and with the acoustic nature of the experimental tasks.

However, the results from the perception and production experiments contrasted to a certain degree. In the LDT, learners with lower accuracy in repeating Korean words showed a more pronounced and consistent pattern as compared with the predictions, but the more accurate learners were comparatively more consistent in the conditional naming latency results, although this difference pertained primarily to the suffix *ero*, and the RT analysis for this task included only four suffixes. In effect, there is an indication that less is more with regard to lexical decision, but more is more with regard to conditional naming, in terms of the involvement of phonological memory. Some of the reasons for this are considered below, but given the limitations of the RT data for conditional naming in particular, this finding must be explored further.

The Korean word repetition paradigm was particularly similar to the conditional naming task, except that no lexical decision was involved. It is thus plausible that the production aspect of the conditional naming task was somewhat easier for participants with higher accuracy scores in foreign word repetition. If this is the case, then the overall processing demands of this task would have a reduced effect on those learners' performance, allowing a more accurate response pattern to emerge.

Foreign word repetition would not have had exactly the same effect on the LDT results because although both tasks measure accuracy in perception, the LDT did not

involve a production component. Furthermore, it was Korean accentedness ratings that interacted significantly with suffix and diphthong status, although the trends for accuracy in the repetition of words were similar. Learners who were rated as less accented (i.e. more accurate in reproducing the Korean words exactly as they were perceived) were relatively ambivalent regarding diphthong status across suffixes, with some variation. Learners rated as more accented, however, showed a pattern of results much closer to the predictions for native speakers, except in the case of *ero*. This result is somewhat puzzling given the prediction that learners who can hear and perceive foreign words more accurately are likely to form more accurate representations of those words in memory.

There is one intriguing possibility that should be researched further. It is possible that learners with greater phonological memory ability are able to rely comparatively more on individual lexical representations, and less on lexical networks between similar items. If this is the case, those lexical networks may be comparatively weaker. In a task involving neologisms, where there are no lexical representations of words, these learners may be at a disadvantage. Along the same lines, they may be more willing to accept anomalous lexical items, based on their greater ability to remember them by brute force, as it were, rather than by linking novel words to preexisting representations.

The use of foreign words in this test of phonological memory supports this account, because short term phonological memory plays a greater role in the recall of unfamiliar phonological material. Conversely, prior knowledge as instantiated in lexical networks has been shown to play a greater role in the recall of more familiar, but still novel material (Gathercole, 1995). Therefore, comparison of learners with high accuracy in repeating unfamiliar foreign words with learners having higher accuracy in the recall

of nonwords with more familiar phonetic and phonotactic characteristics may reveal a divergent effect for the two kinds of phonological memory. In this case, the latter may actually provide greater support for the learning of complex morphophonological patterns such as Spanish diphthongization in derived words, whereas the former may be shown with more certainty to hinder such learning.

Phonological memory thus had effects on learner performance in the present experiment, although those effects were not necessarily as predicted. Nonetheless, there has been very little research on phonological memory using phonetically and phonotactically foreign words. The present results are promising that further investigation into this facet of phonological memory will prove fruitful for second language acquisition research.

The other two individual difference measures yielded less clear results. In the case of working memory span as measured on the reading span task, there was a significant effect in LDT response times, but the direction of that effect was somewhat ambiguous. In general, the biases for each suffix increased slightly in the predicted direction for learners with high memory span, but this difference was slight. In the case of *ito* the change was more dramatic, resulting in a distribution more closely resembling native speaker performance and predictions. On the other hand, there was also a strong diphthong bias for neologisms ending in *ero* for this group. Nonetheless, it has been argued above that *ero* may be more difficult to acquire based on its productivity and morphological opacity in some lexical items (e.g. *compañero* ‘companion’), as well as other factors affecting the availability of its lexical biases for acquisition in most US

Spanish language classrooms. If this is the case, then memory span will have relatively little effect on its learning until its patterns are experienced with sufficient frequency.

Over the remaining suffixes, increased memory span appears to enhance performance on the diphthongization pattern. This may be related to more efficient performance of the task, but the LDT was the simplest of the tasks, and therefore the least likely to be significantly affected by cognitive differences among learners. On the other hand, diphthongization in derived words depends on the tracking and integration of a number of different kinds of information, often in communicative situations where meaning, intent, and other aspects of interaction must naturally also be processed. Since memory span measures the ability to keep track of certain information while simultaneously processing other information (by eliciting plausibility judgments while requiring that particular words be held in memory for later recall), it is likely that repeated experiences of Spanish lexical items may be represented in more detail in memory, and integrated into more robust and finer-grained grammatical knowledge based on lexical networks.

Finally, there were significant results for the Simon task only for the conditional naming task. In the false alarm data, learners for whom the incongruent trials caused a greater disturbance (increasing RTs as compared to congruent trials) showed an overall bias for monophthongs. Those with a smaller Simon effect, i.e. with greater inhibitory or attentional control, showed no bias. On the other hand, in the RT data (again, based on only four suffixes), the group with greater attentional control showed a strong diphthong bias for neologisms ending in *al*, contrary to the corpus and native speaker patterns

(although the pattern was less stable in the current experiment than was apparent in the forced choice data). The group with less control showed the predicted monophthong bias.

These results are difficult to interpret. In the case of the false alarm data, examination of the bias patterns by suffix showed that the group with less attentional control showed more consistent monophthong bias for those suffixes where this was predicted, but also for *ísimo* and *azo*. Thus, the overall bias was consistent with overregularization of the diphthong alternation based on stress. On the other hand, the group with greater attentional control showed more variability in bias, but if anything their behavior deviated more from the predictions. In any case, the Simon interaction was restricted to diphthong status, and did not modulate biases by suffix for the false alarm data, but only for the RT data. In both sets of results, however, the diphthong bias for neologisms ending in *ito* was stable and as predicted. It is possible that the remaining suffixes are simply in the process of being acquired, and that the current cross-sectional data are not detailed enough to distinguish an effect.

5.5 Conclusions

The results for the adult learners in the current experiment are without a doubt complex. Nonetheless, there is support for the hypothesis that adults are able to acquire fine-grained probabilistic structures in L2 grammar such as the diphthong/mid-vowel alternation in Spanish derived words. Interestingly, there was greater evidence that this morphophonological pattern impacts the lexical processing of intermediate to advanced learners at the university level than that it impacts their behavior. This is an intriguing

result. It indicates that learners' lexical processing is remarkably attuned to small variation in the probability of certain lexical patterns, but that these patterns do not affect the rate at which learners accept neologisms as real words. The finding of such fine-grained sensitivity, even if mainly in lexical processing, is striking for the further reason that the patterns are based on abstract categories such as stems, paradigms (in which stems are seen in various forms), and derivational morphology, which bear little resemblance to patterns in the L1.

This supports a theory of second language acquisition in which grammatical knowledge is emergent at a variety of levels of processing, modulated to some extent by proficiency and experience with the L2. It also suggests that grammatical knowledge is dependent on statistical patterns across abstract linguistic entities that may themselves be emergent.

In addition, the present study provided evidence that the course and outcome acquisition of such grammatical patterns in an L2 is dependent on learners' individual cognitive characteristics, independently of proficiency. However, different aspects of memory, attention, and processing efficiency appear to affect different levels of processing differently. In some instances, the effects were attributable to task-based effects in which certain types of learners were particularly impacted by certain experimental task conditions. In other instances, it appears that some cognitive resources, such as memory span, may facilitate the learning of fine-grained probabilistic grammar, and others, such as foreign word repetition ability, may inhibit certain aspects of that acquisition.

The incorporation of measures of cognitive resources into the present study provides an important early step in exploring the effects of adult cognition on the emergence of grammar. The present experiments provide evidence that emergentist theories of grammar acquisition are demonstrably applicable to adult L2 acquisition, and they also suggest avenues for future research into how variability in the constellation of adult cognitive capabilities may interact with language experience during L2 acquisition.

Chapter 6

Conclusions: Emergence, SLA, and Multilingualism

6.1 Overview

The research reported in this dissertation substantially reframes many questions concerning the nature, mechanisms, and results of second language acquisition in adults. It does so by taking an emergentist view of second language acquisition that makes strong predictions about the types of information available through experience of a language, first or second, and about the detailed sensitivity to statistical patterns that arises through the interaction of general cognitive mechanisms with that experience. This dissertation argues that, just as too strong a focus on the effortlessness, universality, and inexorability of child language acquisition obscures important facts about the acquisition, representation, structure and use of language (see Section 1.1.1), so also does a focus on the effortfulness, incompleteness, and variable success of adult SLA obscure important observations about how adults learn additional languages.

The finding that the adult learners of Spanish in the present study had developed keen sensitivity to such a subtle probabilistic pattern as that examined here provides striking evidence that the statistical learning mechanisms that have been observed in first language acquisition and in artificial language studies are also operative in adult second language acquisition. This reveals a basic continuity in the process of language acquisition in childhood and adulthood. However, the sensitivity exhibited by learners

differed in important ways from that of the native speakers. There were some discrepancies in the exact pattern of diphthong biases, and there was some evidence that these may be attributable to differences in L2 experience and to cognitive differences between learners. More importantly, the behavior of diphthongizing stems in existing Spanish derivations primarily affected processing time in learners, but error rates in native speakers. This suggests that there are differences as well as similarities in how lexical statistics impinge on language processing and acquisition in adulthood vs. in childhood. It may be that still more advanced learners would more closely resemble native speakers in their pattern of results, but the precise reasons for the discrepancies observed here merit further investigation.

Together, the findings summarized here serve to recast and refocus questions about adult SLA. They point the way forward to evaluating the deep continuity in the mechanisms of language acquisition throughout life, particularly in the kind of implicit learning that is grounded in the statistical characteristics of language in use. They suggest that the obvious differences in the course and outcome of adult SLA and child language acquisition are due to the potentially far-reaching consequences of subtle changes in language experience and the development of certain cognitive resources.

The chapter is organized as follows. The following sections summarize the results and discuss their larger-scale implications for our understanding of adult L2 learning, and for the ways we ask questions about research on age-related differences in the nature, course, and outcomes of SLA. Subsequently, future directions in corpus linguistics and work on memory and attentional control are discussed in light of their contribution to

usage-based research on SLA. The following section considers the impact of emergentist research on SLA for our understanding of multilingualism in general.

6.2 The present findings

The experiments reported here tested the prediction that adult second language learners are sensitive to fine-grained probabilistic patterns in L2 grammar. This was expected based on emergentist theory because the general learning processes that track the development of frequency distributions across similar features in the lexicon in L1 acquisition are likely to persist into adulthood. Given, however, the obvious and well-documented differences between L1 acquisition and adult SLA, a second hypothesis was tested in order to determine how a set of individual differences in adults' L2 learning experiences, working memory, phonological memory, and attentional or inhibitory control capacities might modulate their ability to track the frequency of specific lexical patterns.

6.2.1 Diphthongs, monophthongs, and the processing of neologisms

These hypotheses were tested based on adults' learning of a gradient subpattern in the Spanish diphthong/mid-vowel alternation. In this subpattern, alternating stems predictably contain either a monophthong or a diphthong, but the predictability of the form of a stem is modulated by both a very general, stress-based pattern across the lexicon and by the behavior of alternating stems in derived words with particular suffixes.

The predictions of these cues, stress and suffixes, often conflict, and native speakers resolve these conflicts in systematic, but probabilistic ways. The complexity and depth of lexical sensitivity on which knowledge of this subpattern rests make it an ideal structure for testing fine-grained statistical sensitivity to L2 grammar in adult learners.

The experiments in this dissertation revealed a complex pattern of sensitivity to lexical statistics that affected different levels of processing and behavior in different ways, for both native speakers and learners. To evaluate adult learners' performance, the results were compared to predictions for nativelike behavior based on three sets of data. First, predictions were made regarding the lexical bias of certain suffixes towards stem monophthongs or diphthongs based on the forms of existing derivations with alternating stems in a large, dictionary-based corpus (Eddington, 1996). This corpus data revealed a gradient proportion of derivations containing diphthongs vs. monophthongs according to particular suffixes. However, these data were limited because highly productive suffixes are generally underrepresented in dictionaries, and so corpus-based predictions were only made for some of the suffixes tested. A diphthong bias was predicted for the more productive suffixes based on the existence of a number of frequent derivations with these suffixes that contain diphthongs (e.g. *pueblito* 'little village'), and on Eddington's prior behavioral data as outlined below. Further issues and future directions concerning corpus-based predictions are discussed in Section **6.4.1**.

The second basis for measuring the nativelike target pattern of gradience came from metalinguistic judgment tasks conducted by Eddington (1996; 1998), which revealed that native speaker preferences parallel the corpus (dictionary) data for the less productive suffixes and favor unstressed diphthongs in novel derivations with the more

productive suffixes. Thirdly, the present experimental tasks were tested with a group of native Spanish speakers in Spain. This study used lexical decision and conditional naming tasks to measure rapid, low-level lexical processing as well as judgments of lexical status, reflected in reaction times and error rates, respectively.

There was general consistency among the predictions based on these three sources, but there was also some variability. In some cases, the bias of a given suffix was more neutral than expected, and at times even reversed. Some of this variability was likely due to the limitations of the dictionary corpus, but there were contrasts between the three native speaker studies that cannot be attributed to inaccuracies in the corpus-based predictions alone. Instead, it is likely that the various tasks used to evaluate native speaker behavior were sensitive to different levels of processing, and involved different types of decisions, which are impacted by lexical statistics to varying degrees.

First, the forced-choice tasks employed by Eddington may have exaggerated slight but reliable biases in the lexicon, based on the fact that his participants were given both possible forms of each neologism. The direction of the bias may have been clear and stable, but this methodology may have obscured cases where the less preferred alternative would nonetheless be acceptable.

Interestingly, while the rankings of suffixes according to their lexical bias for diphthongs or monophthongs was very similar in the results of Eddington (1996) and Eddington (1998), they were not identical. No suffixes were found to reverse their bias, but there were shifts in the strength of the biases for individual suffixes, such that, for example, *oso* showed a stronger monophthong bias than did *ero* in the earlier study, but not in the later study. This may have been due to more subtle methodological differences

between the two experiments. In his first study, Eddington supplied definitions for the neologisms, followed by the forced choice. In the second study, the questionnaire consisted of sentences illustrating the meaning of the neologisms, and participants chose the correct form of the neologism to complete the sentence. Crucially, the sentences contained other, attested forms of the test stems (see the example in Section 2.2.2), which was not the case in the 1996 study. This may have altered the activation of lexical information during the task, creating the variability in findings.

The present study employed a substantially different methodology. Items were presented in the auditory instead of the visual modality, and participants listened to isolated words and nonwords instead of sentences. They thus necessarily only encountered one possible form of each neologism, and care was taken that they heard each test stem only once. There was no semantic context. Moreover, the task was not to determine the “correct” form of a neologism (recall that Eddington told his participants that the neologisms were real, but infrequent words and asked them to choose the correct form), but rather to decide whether an item was a real word or not. Reaction times were recorded as a measure of online lexical processing, and error rates were recorded as a further indication of speakers’ biases. Finally, the conditional naming task introduced the additional demands of combining lexical decision with the repetition of neologisms, introducing a speech production component to the experiment.

These task-based differences between the native speaker studies are not trivial. These differences were generally associated with a slightly different ranking of suffixes, and total reversals of the bias associated with a given suffix were relatively rare. It is possible that some of these differences may have been due to the inherent variability

associated with probabilistic biases. More significantly, they may be attributed to differences in the amount and type of information activated during the task, and to the types of decision that must be made based on that information.

For example, the pattern of native speaker biases was observed to be somewhat weaker in the RT data, and substantially weaker in the conditional naming task. As discussed in Section 4.3, however, evidence for simultaneous effects of inhibition and facilitation was observed in different aspects of the naming task, which may have leveled the differences between suffixes to a certain extent. This was based on the unfolding of acoustic stimuli over time, with a consequent progression in the amount of information available for processing, and on possible revisions in the neuromuscular response based on the later perception of further morphological information, i.e. the suffix.

Intermediate to advanced learners' processing of the test neologisms reflected nativelike patterns of morphophonological sensitivity, but this was manifested in a slightly different way on the various dependent measures. The strongest pattern of lexical biases was observed in the RT data from the LDT. Overall, the pattern of results shown on this measure was strikingly similar to the suffix biases observed in the native speaker data, but this pattern failed to emerge significantly in the analysis of error rates on the same task, where the native speaker results had been more robust. This is an interesting finding and suggests that the gradient pattern examined does impinge on adult learners' lexical processing, but that it does not yet affect their behavior as reflected in judgments about lexical status. On the other hand, error rates for the more complex conditional naming task did reflect the lexical biases of certain suffixes, although the results differed from the precise pattern predicted on the basis of native speaker results. The suffixes that

contributed to the significant results in conditional naming were generally more productive, and it is possible that suffix productivity played a stronger role for learners than for native controls in this task. It is possible that still more advanced learners would reflect the same distribution of results across RTs and error rates as the native controls, but it is also possible that lexical statistics impact different areas of processing in adult learners.

The most important differences between the different tasks and measures used to evaluate this structure lie in the lexical information that comes into play during processing. In the metalinguistic tasks used by Eddington, speakers could exploit the lack of a time limit to explicitly consider connections with similar items, semantics, metalinguistic knowledge of rules, and other types of information. Further, the semantic context also left no doubt as to the identity of the stems or the intended meaning of the neologisms. In the LDT and conditional naming tasks, on the other hand, the speeded nature of the tasks drastically reduced the availability of metalinguistic knowledge and conscious reflection. In these tasks, the online activation of phonological networks, morphemes, and paradigms play a decisive role in verifying the lexical status of stimuli. In some cases, especially with shorter stems and suffixes, there was evidence that the intended stem and/or affix may not have been activated to the same extent as, for example, a longer stem with few neighbors, because shorter morphemes contain fewer cues to their identity than longer morphemes. In these instances, the lexical networks that instantiate the probabilistic biases regarding diphthongization may not have been activated, such that this information did not come into play.

A further possibility is that, for some suffixes, or possibly also for particular stem-suffix combinations, the dispreferred form was not significantly more difficult to process than the preferred form, even though, given the option, speakers might exhibit a stable preference for one over the other. This may also tap into differences between what speakers will accept as valid in speech perception and what they might either produce themselves, or consider to be correct in a metalinguistic task.

These findings point the way forward into exploring what lexical information comes into play at what times, and for what purposes, in L2 learning and processing. Further research along these lines holds significant promise for evaluating the context-sensitivity of language processing and use, and for further elucidating the structure of the lexicon in both monolinguals and multilinguals.

Notwithstanding the variability in the nature and results of different experimental tasks, the present research revealed a clear gradient of lexical bias in which native speakers as well as learners favor stem monophthongs in derivations with some suffixes, diphthongs with other suffixes, and are more ambivalent with regard to stem form with still other suffixes. Most strikingly, the results for native speakers and learners alike show that the suffix-based gradient pattern does not simply reflect the processing of existing derivations, but the probabilistic pattern of diphthongization observed in the lexicon also affects the processing of novel derivations.

It is this aspect of the present study that provides crucial support for the emergentist hypothesis. We saw in Section 2.2.1 that generative analyses of Spanish diphthongization require a diacritic marking on exceptional forms, marking either those forms that contain an unstressed diphthong, or possibly those instances in which stress

rules are applied cyclically (Halle et al., 1991; Harris, 1977; Roca, 1988, 1991).

However, even accepting that the diacritic marking of irregulars is necessary in some instances, these accounts cannot explain the impact of this probabilistic pattern on neologisms, which have by definition never before been encountered. At most, this account would predict that the test items in this study that contained unstressed diphthongs would be somewhat more difficult to process, because native speakers and learners alike would have to determine that a diacritic was needed in the underlying representation of the neologism.

This prediction would suggest that unstressed diphthongs would have the same impact on processing and error rates, regardless of suffix, unless there were some way of encoding the observation that some suffixes tend to favor unstressed diphthongs. This might be done by labeling some suffixes as cyclic, and others as non-cyclic, but this would suggest a binary distinction between the two types of affix. In this case, we would expect slower processing and more error rates with unstressed diphthongs for some suffixes, and with unstressed monophthongs for the others, but no ambivalence. Likewise, the strength of the effect for each class of suffix would be expected to be the same, and not gradient. However, the pattern of results observed in Eddington (1996; 1998) and in the present study shows that this effect is larger for some suffixes than for others, and furthermore, that the results for still other suffixes are ambivalent.

Even keeping the notion of diacritic marking for the moment, these studies suggest that speakers are able to predict fine-grained distinctions in the probability that a neologism will be diacritically marked to contain an unstressed diphthong. Such an analysis is certainly possible, and highlights the fact that usage-based theory does not

necessarily preclude the existence of cognitive entities that function at an abstract level, in other words, rules (Hay & Baayen, 2005). The difference between such an account and the earlier generative analyses is that whatever rules are taken to be (see, e.g. Bybee, 2001b; Goldstone, 2000; Langacker, 2000; Tomasello, 2003), they reflect probabilistic knowledge of language structure.

The most likely source of this probabilistic knowledge is the lexical evidence provided by prior experience, as hypothesized in emergentist and usage-based theory. The finding that speakers are sensitive to the fine-grained probabilistic pattern of diphthongization seen in Spanish derivations, and that this sensitivity extends to novel lexical items, therefore provides strong evidence that their productive knowledge of this pattern depends on structured memories of accumulated experience of phonologically and morphologically related words.

One further note is in order regarding the scope of the present findings. The structure examined is a probabilistic subpattern of a relatively consistent, stress-based alternation between diphthongs and mid-vowels in certain stems. Traditionally, it would be termed an irregular pattern, although a stochastic analysis shows that it is in fact predictable. Nonetheless some caution is in order regarding the extension of the present conclusions to cover regular patterns as well.

A strictly emergentist approach to grammar holds that regular structures and irregular structures alike emerge through the continual interaction of general cognitive processes with the accumulation of new and repeated exemplars of language in memory. While there is strong computational and behavioral evidence that this is the case (e.g. Baayen, 2003), many researchers also advocate Dual-Process models, in which irregular

structures are represented in an associative network compatible with emergentist accounts (although generally stated in terms of more strictly connectionist modeling), but regular structures are represented using combinatorial rules referring to abstract variables, including possibly variables such as stress and diphthongs (e.g. Pinker, 1998; Pinker & Ullman, 2002; Ullman, 1999, 2004).

The clearly probabilistic structure examined in this dissertation provides evidence for the statistical learning processes to which usage-based theory attributes the emergence of regular and irregular structures alike, and for the emergent impact of lexical statistics on lexical representation and processing (Langacker, 2000). It does not, however, allow for strong conclusions to be drawn about whether either native speakers or adult learners in fact do acquire and process regular, or default structures in this way. Notwithstanding, this dissertation argues that research on bilingualism, and on adult L2 learners in particular, has an important contribution to make to the debate concerning regular and irregular structures alike.

6.2.2 Experiential and cognitive factors

The evidence for involvement of particular experiential and cognitive factors was more complex, but nonetheless promising, opening the door to exploration of the role of specific cognitive resources in particular. Proficiency as measured in Spanish picture naming played a relatively weak role in modulating the extent to which learners' performance reflected the specific biases of particular suffixes, although it did modulate general preferences for monophthongs and learners' response to certain suffixes.

However, there were relatively few learners at the lower end of the proficiency range sampled, and the particular picture naming task used may not have been sensitive enough to detect more meaningful differences between learners. Study abroad had a stronger effect on LDT results, indicating that the more intensive exposure to lexical patterns available in an immersion environment, as well as probable exposure to a more representative corpus of language experience with regard to the pattern in question, may have facilitated sensitivity to Spanish diphthongization patterns.

Working memory, phonological memory, and attentional or inhibitory control also played a role in different aspects of the results, with the strongest result coming from phonological memory. This supports the hypothesis that adults' more developed and varied cognitive capacities affect their sensitivity to statistical patterns in grammar. However, these measures affected learners' performance in different ways on different tasks, suggesting that their effects in this experiment may pertain more to the way learners use prior knowledge to process and interact with novel linguistic experience in specific tasks than to the way they construct lexical networks based on frequency patterns in the input during acquisition.

In particular, performance on the foreign word repetition task showed that some learners more accurately matched the lexical biases of certain Spanish suffixes on some tasks. Interestingly, learners with less accurate foreign word repetition performance more closely matched the nativelike target pattern of diphthong biases on the LDT, but learners who repeated foreign words more accurately more closely matched the predictions on the conditional naming task. Both of these results were subtle, yet significant, and therefore suggest that foreign word repetition affects performance on different tasks in different

ways, particularly given the similarity between foreign word repetition and the repetition of neologisms in the conditional naming task.

6.3 Emergence, age, and language acquisition

To summarize broadly, the emergentist point of view taken here holds that language structure evolves through the continual interaction between language input and cognition. As repeated use of language accumulates, new memories are incorporated into networks based on similarity with old memories, and through this process the frequency of past experience of grammatical structures comes to influence the perception, processing, and production of new instances of language use. As we saw in Chapter 1, this approach is compatible with a wide range of evidence from synchronic and diachronic linguistics as well as psycholinguistics. Frequency effects have been shown in language change, in native speakers' metalinguistic intuitions, and in language processing, including reaction time studies as well as neuroimaging studies.

At its base, emergentist theory is quite simple. General cognitive processes of similarity identification, tracking of frequency, and categorization interact with repeated exposure to language, and this interaction leads to the emergence of linguistic structure (Tomasello, 2003). The crucial implication for questions of adult second language acquisition is that these cognitive mechanisms are not dedicated language structures, but rather they are used for learning in a variety of modalities (e.g. Creel et al., 2004; Saffran et al., 1999).

This brings vital changes to the way we ask questions about age effects in SLA. First, it points out that there is no single, monolithic cognitive structure responsible for language acquisition, which might otherwise be thought to turn off after L1 acquisition, fundamentally changing the nature of the acquisition of additional languages (Bley-Vroman, 1989). Second, the drawing of analogies, tracking of frequencies, and other processes associated with emergent grammar (Langacker, 2000; Tomasello, 2003) are observed throughout the lifespan, suggesting that language acquisition is qualitatively similar across age groups. This is not to say that there are not differences, but it suggests that the source of those differences lies in changes in various aspects of the system due to factors related to developmental or aging processes, and not in a radical change from one system to another.

Emergentist SLA research opens the door to a much closer look at the basic processes underlying the interaction between input and cognition during language acquisition, and makes it possible to begin evaluating the consequences of subtle changes in a complex set of environmental and cognitive factors for language acquisition in different age groups. Research such as that reported here is therefore poised to make significant contributions to the discussion of age effects and a possible critical period for language acquisition. This is broadly in line with the view put forward by Birdsong (2005). He concludes by pointing out that when we speak of a decline in language acquisition ability in adults, this tends to imply “a single monolithic learning faculty”, but he goes on to say that “it is likely that L2 learning involves distinct cognitive and neural components with differential susceptibilities to the effects of age” (p. 125). The complex set of cognitive processes thought to underlie language acquisition in an emergentist view

offers ways of evaluating this hypothesis because, for example, it leaves open the possibility that different processes are subject to different age-related declines in plasticity, offering one explanation for the continued decline in the asymptotic results of SLA observed long after maturation. Furthermore, the piecemeal nature of the decline noted by Birdsong, in which some areas of language structure are affected more than others, may be attributed to differential changes in specific cognitive processes as well as to changes in the quality and nature of the input.

The present research makes an initial move in this direction by evaluating individual differences in certain cognitive resources. There has been much research on individual differences relating to the acquisition by adults of larger-scale regularities in L2 grammar, but the main innovation of this dissertation is to link these cognitive resources with the kind of subtle sensitivity to very small-scale probabilistic regularities that is expected under an emergentist hypothesis. By linking resources such as phonological and working memory to learners' sensitivity to lexical statistics, the present research opens a crucial window into the precise mechanisms by which these abilities affect the low-level, recursive processes that form the roots of language learning.

Evidence for a role of different types of memory and cognitive control, such as is reported in this dissertation, lends further support for the notion that certain advantages in adult cognitive capacities alter the language acquisition process (e.g. Newport, 1990, 1991). This observation is not incompatible with the earlier view of Krashen that certain factors (e.g. the Affective Filter) obstruct at least some of the mechanisms operative in early language acquisition, leaving adult L2 learners to rely on other, more well-developed reasoning and problem solving abilities (Krashen, 1981, 1982). This might be

reformulated to say that either maturational or aging processes (or, more likely, both) effect declines in certain processes underlying language acquisition, while other, higher-level cognitive processes become better developed and thus play a larger role.

This account may be seen as a more precise statement of Krashen's distinction that language acquisition in children is implicit, but language learning in adults is explicit in nature. However, this dissertation argues that this type of binary distinction is too strong, and it obscures important aspects of language acquisition in both children and adults. For instance, an updated formulation of the implicit/explicit distinction between child and adult language acquisition is expressed by DeKeyser and Larson-Hall:

Children necessarily learn implicitly; adults necessarily learn largely explicitly. As a result, adults show an initial advantage because of the shortcuts provided by the explicit learning of structure, but falter in those areas in which explicit learning is ineffective, that is, where rules are too complex or probabilistic in nature to be apprehended fully with explicit rules. (DeKeyser & Larson-Hall, 2005, p. 103)

This generalization is supported by ample evidence, and it has led to a substantial literature on SLA and L2 pedagogy. However, it ignores ways in which the implicit learning of language structure persists into adulthood. This was shown in the statistical learning literature reviewed in Section 1.4, which used artificial languages to show that adults are indeed capable of inducing word boundaries and other aspects of language structure from statistical patterns in the input. Furthermore, the finding that children may be more prone than adults to regularize inconsistent structures in artificial language input (Hudson Kam & Newport, 2005; Wonnacott & Newport, 2005) suggests that adults possess an advantage for certain kinds of implicit learning.

The present results make an important contribution to this argument by showing that adult L2 learners are able to acquire sensitivity to very subtle probabilistic patterns in L2 grammar. Information about the statistical distribution of a morphophonological pattern involving alternating stems in a variety of derivational contexts was found to impinge probabilistically on learners' processing of novel derivations. This was true even for learners whose only exposure to L2 Spanish was in a pedagogical environment where, arguably, explicit learning processes are at their strongest. Despite the presence of an explicitly storable rule against the appearance of unstressed diphthongs in alternating stems, the learners in this study responded to systematic variability in the well-formedness of unstressed diphthongs in specific morphological contexts.

There is little doubt that explicit learning processes give adults an initial advantage in learning relatively consistent structures, and it is also reasonable to say that explicit learning is unlikely to be of much use in learning the kinds of subtle probabilistic variation exemplified in the Spanish diphthongization subpattern examined here. The finding that adult L2 learners are indeed sensitive to this subpattern suggests, therefore, that implicit, meaning in this case statistical, learning processes are still available to adults.

Put simply, adults are capable of developing a sense of what sounds right in their L2 based on the same kinds of statistical information exploited by children learning their L1. Coupled with what is known about explicit learning in adult SLA, this changes the questions that we may ask about adult language learning in several ways. First, these findings argue that explicit learning is an addition to, and not a replacement for the language acquisition processes operative in childhood. The conglomeration of cognitive

mechanisms underlying L1 acquisition persists, at least in large part, throughout life. Going further, explicit learning, and whatever other cognitive resources are added through maturation and experience, likewise do not represent a unitary ability, but rather a set of abilities and ways of thinking that themselves emerge piecemeal as part of the cognitive evolution of an individual.

Second, age-related declines in language learning ability are no longer seen as fundamental changes to a whole system, but as the result of both declines in specific processes or abilities that play a role in early language acquisition and of further development in higher-level cognitive abilities. Some of the declines may indeed be maturational, eventually reaching a minimal floor level after a critical period. Others may proceed gradually throughout the lifespan as a result of aging, rather than maturational processes (Birdsong, 2005). In any case, however, the abilities underlying first language acquisition form a complex set of interacting cognitive mechanisms, each of which may follow its own course of development. The nature of language acquisition as a whole appears to evolve in response to changes, declines, or additions among a number of mechanisms that interact synergistically in response to language exposure. Accordingly, the task of research on language acquisition at different ages is to understand how the evolution of multiple cognitive processes leads to changes in the course and outcome of language acquisition.

Thirdly, language structure emerges through the interaction of this set of cognitive processes with the specific characteristics of an individual's language experience. While progressively larger corpora of experience (e.g. the experience of two individuals) of the same language can be expected to converge on similar grammatical patterns through the

principle of statistical robustness (Pierrehumbert, 2003b), there are also necessarily significant differences between the accumulated L1 experience of a native speaker and the experience available to an adult L2 learner of the same language. For example, adult learners may experience the L2 in only limited genres or registers. The present research showed that, for the particular pattern examined, adults were sensitive to the probabilistic behavior of diphthongizing stems in derived words even when exposed to Spanish in classrooms only. Nonetheless, learners who had studied abroad showed a more nativelike pattern of sensitivity.

The picture of adult L2 learning that becomes apparent involves a complex set of cognitive processes that interacts recursively with accumulating language experience. This picture reframes the questions we ask about language acquisition across the lifespan in important ways. Just as in L1 acquisition (see Chapter 1), an emergentist perspective asks what information about L2 structure is available through the statistical characteristics of L2 exposure, how those characteristics impinge on learners' processing, and by what mechanisms adult learners develop knowledge of statistical patterns in their L2 input as well as other information accessible to them based on their more developed metalinguistic knowledge. The question of age-related differences is reframed to consider differences in language exposure at different ages as well as subtle variation in the experiences of learners in different contexts. Crucially, it must also ask how specific cognitive changes throughout the lifespan affect the accessibility of particular patterns and structures in L2 input. This involves understanding how L1 knowledge and adults' cognitive characteristics alter their perception of and memory for particular aspects of L2 structure, and how new experiences are fit into lexical networks. The present research

strongly suggests that framing these questions in terms of emergentist theory will allow for a much more detailed understanding of both similarities and differences between language acquisition in individuals of different ages, as well as of the reasons for which specific structures are more easily acquired than others.

6.4 Emergent grammar in adult SLA

The way forward from here, as already reflected in this dissertation, must be vitally interdisciplinary. Much research has already been done on areas such as individual differences, input processing, and lexical processing in bilinguals and learners, and recent advances in corpus linguistics reveal the fine detail available in language experience. The remainder of this chapter briefly considers areas of corpus research, L2 processing of language experience, and multilingualism that could be used to provide links between the present findings and other areas of research and that may be used to show how an emergentist and usage-based perspective might integrate research from a variety of domains into a more unified theory of the acquisition, representation, and use of human language (MacWhinney, 2001, 2005b, 2007).

6.4.1 Estimating the grammatical information available for acquisition: Corpus research

The hypotheses in this study relied centrally on a link between lexical patterns of diphthongization in Spanish derived words and native speaker performance on novel derivations based on these patterns. However, it has been noted above and in Eddington

(1996) that the dictionary data used to make these predictions were limited in several ways. In particular, Eddington noted marked contrasts between his dictionary data for highly productive suffixes (augmentatives and diminutives) and both intuitions regarding the behavior of these suffixes and earlier analyses of the idiosyncrasies of the diphthong/mid-vowel alternation. This weakness can be remedied by using corpora of actual Spanish in use in various contexts, such as the LexEsp corpus (Sebastián-Gallés et al., 2000), because productivity is much more likely to be accurately reflected in large samples of Spanish usage.

The use of more detailed corpus data has other benefits as well for statistical and usage-based research, as reflected for example in the research on neighborhood density and probabilistic phonotactics reviewed in Section **1.3.3**. This research relied on calculations of the positional probability of segments and larger phonological and phonotactic sequences, as well as counts of the number of related items based on certain parameters such as shared phonemic content. This can also be done (and frequently has been done) using dictionary data, but the use of naturally occurring samples of language has the added benefit of yielding token frequency statistics, that is, the frequency of individual lexical items. This was exploited in, among other things, the research on neighborhood density in which neighborhood size was weighted for word frequency (e.g. Vitevitch & Luce, 1998).

Frequency distributions across other levels of linguistic structure related to Spanish diphthongization have also been shown to affect processing, including the frequencies of morphologically complex words and their associated stem and affix frequencies as well as morphological family size (Baayen & Schreuder, 2003). Eddington

reduced his dictionary-based sample of alternating stems in derived words through a measure of the familiarity of derivations, approximately removing derivations with low token frequency. However, the other measures mentioned above may also affect the processing of novel word forms, complicating analysis of patterns such as diphthongization. It may be that a closer look at these multiple factors in more accurate samples of language in use, using techniques such as multilevel statistical modeling, will be able to make more sophisticated predictions about speaker performance on tasks such as those employed here. With regard to the present data, this may help to explain, for example, why the diphthongization effect was stronger in the error rates for native speakers, but relatively weak and susceptible to other factors such as stem length in the RT analyses.

Corpus research such as this is currently being done in many ways, some of which have been reviewed in this dissertation. However, less has been done along these lines in relation to L2 acquisition. This is partially due to the much higher level of variability in learner experience, especially in classrooms. Vocabulary is not generally taught according to word frequency (but see Davies, 2006), and lower-frequency sets of vocabulary may receive disproportionate attention at certain stages but be virtually ignored at other times. Similarly, grammatical structures are generally presented in units, with the focus on certain aspects of structure at some times, and other aspects at other times. This type of pedagogical practice, while possibly justified for other reasons, may distort the statistical characteristics of the native speaker lexicon, affecting the probabilistic generalizations that may emerge through statistical learning processes.

It is clear that accurate, extensive, and contextually-appropriate (e.g. language classroom, literature seminars, study abroad) corpora of learner experience and use of language are likely to yield more sophisticated and accurate predictions about learner performance on specific grammatical patterns. The present research opens a significant door to emergentist investigations of L2 acquisition, but corpora of this kind will be a valuable tool for continuing this research. They will allow more precise observation of the evolution of learners' L2 processing and behavior as a function of their language experience, and more precise evaluation of the factors that modulate their response to frequency.

6.4.2 Learner interaction with L2 experience

Corpus analyses such as those suggested above offer significant promise for extending the work begun in this dissertation. However, they must be coupled with specific work investigating what L2 features are tracked by learners and why. There are two primary lines along which this inquiry might proceed from here, external factors that either highlight or obscure certain aspects of structure, and individual differences between learners that might affect the availability of structures for acquisition. Both areas have significant research traditions already, but the present research suggests that coupling this research with an emergentist view of language structure, acquisition, and representation holds significant promise for understanding language acquisition across contexts of learning and use and throughout the lifespan.

There are many ways to examine the factors modulating the salience of certain features, and thereby their availability for incorporation into lexical networks and for frequency tracking. We have already discussed research suggesting that some features may become highlighted and others suppressed as emergent structure begins to constrain the perception of phonetic categories (Kuhl, 2000; Pierrehumbert, 2003a). This “filtering” or “warping” of perceptual space may also affect higher-level structures such as Spanish diphthongization.

There are of course lexical factors in the L2 that may serve to constrain the perception and integration of new experiences into memory, but L1 structure may also play a significant role. Parallel structures may be somewhat less likely at some levels of structure, and indeed the Spanish diphthongization structure examined here has little similarity with the L1 English structures known to the learners who participated in these experiments. English is replete with diphthongs, but they do not necessarily alternate in ways similar to Spanish, and exact parallels between some of the derivational suffixes (especially the augmentatives and diminutives) are difficult to find. On the other hand, there are cases in English where diphthongs are reduced as a result of morphological changes, as in *div[ɑ']ne* and *div[ɪ]nity*. It is as yet unclear how this English pattern of trisyllabic laxing might interact with Spanish diphthongization during L2 learning, but there may be other structures in which clearer parallels may be made, and which may serve to alter the perception or integration of L2 experience into memory. Investigation of such structures may shed light on the interactions between languages in learners and multilinguals (Cook, 1995, 2002, 2003; Hall et al., 2006).

Apart from crosslinguistic comparisons such as these, there are also a variety of semantic, contextual, and communicative factors that may affect learners' sensitivity to statistical variability in L2 grammar. Derivations with infrequent stems or suffixes may be interpreted as being monomorphemic, such that their morphological structure may not be tracked in the same way as words perceived as multimorphemic. The semantic transparency of a derivation may also affect its potential decomposition during processing. One example might be the word *mostrador* 'counter, desk', which may or may not evoke any semantic connection to the verb *mostrar* 'to show' for learners at a given stage in acquisition.

This raises the important question of how learners confront morphologically complex words. The results of the present experiment suggest that patterns involving more productive suffixes are acquired more easily than those involving less productive suffixes (see Figure 5-1). This may reflect greater salience of the more productive suffixes, leading to an increased tendency to decompose derivations with those suffixes than with other suffixes such as *ero*. While this provides a plausible explanation for some of the current results, the developmental course of morphological sensitivity and its consequences for L2 learners' sensitivity to subtle statistical patterns in grammar is far from clear.

The other category of factors modulating learner sensitivity to gradience in L2 grammar consists of cognitive differences, both between adults and children (Newport, 1990) and between individuals within age groups. The present experiments have provided evidence that adults are still sensitive to the types of fine-grained structure underlying the representation of grammar in native speakers, and that statistical learning mechanisms by

which these structures may be acquired are still operative in adults. However, there was also evidence that certain aspects of adult memory and cognitive control were associated with more or less accurate performance on the experimental tasks, in comparison with native speaker norms.

As noted above, accuracy in repeating words from an unfamiliar foreign language (Korean) yielded the strongest effects on the tests of Spanish diphthongization, but the results affected learners differently depending on the experimental task. Less accurate foreign word repetition appeared to be beneficial on the LDT, but there was a slight trend in the opposite direction for conditional naming. This result is difficult to interpret because error rates on the conditional naming task resulted in too little data for some of the suffixes, and the trends are therefore analyzed over a subset of the experimental conditions.

The possibly task-bound nature of some of these effects (cf. the effects of memory span and Simon task performance, which played an even less consistent role in the results than foreign word repetition) raises questions about the stages at which these cognitive variables affect learning or performance. For example, it is unclear whether higher foreign word repetition ability facilitates performance in conditional naming, allowing for more detailed grammatical information to affect processing, or whether it affects learners' ability to perceive and track lexical statistics. The results from the LDT, in fact, indicate that higher foreign word repetition ability may actually hinder the accurate tracking of lexical statistics by enabling learners to attend to too much information at once. Based on findings such as this, there is little doubt that the effects of particular cognitive abilities are highly complex, and that certain abilities may affect learning and

processing in different ways at different stages. Further research using a variety of experimental techniques will be necessary to pinpoint the locus of influence of specific cognitive abilities.

Despite these unanswered questions, the research reported here has shown adult learners to possess similar kinds of knowledge about Spanish diphthongization to that of native speakers. Extending the emergentist perspective to more proficient bilinguals and multilinguals, this suggests that the knowledge of additional languages is essentially similar to first language knowledge, regardless of age of acquisition. Both are dynamic, continually evolving throughout life as new experiences of language are encountered, perceived, and incorporated into emergent lexical networks. Before closing this dissertation, a brief discussion of how emergentist research can inform research on multilingualism in general is relevant.

6.5 Emergent grammar in multilingualism

Emergentist theory has much to say about the interaction between different aspects of language knowledge in human cognition, including the interaction between two or more languages. Language acquisition from this point of view, and indeed language knowledge in general, involves a continual process of tracking features of linguistic experience at a variety of levels simultaneously. In multilingualism many of these features may naturally be related to identifying and distinguishing one language from another, but this by no means precludes the possibility of rich connections between similar structures in multiple languages of the kind underlying the sensitivity to gradient

structure in a single language examined in this dissertation. We have alluded to how L1 influence on L2 learning may be understood in this way, but there is also increasing attention being paid to the influence of subsequent languages on the first (Cook, 2003).

This evidence supports two conclusions, both highly compatible with emergentist theory. First, it demonstrates that even monolingual language competence never reaches a stable end-state, but remains constantly dynamic. Second, it shows that multiple languages within the mind of one speaker are not inviolate, but they interact in complex and synergistic ways. An emergentist view of these conclusions has recently been linked to the developing concept of *multicompetence* (Hall et al., 2006), first suggested by Cook (1991) as a way of overcoming difficulties of accounting for multilingual language competence in traditional linguistic theory.

The central proposal of research on multicompetence is that speakers of more than one language cannot be seen as possessing multiple, monolingual-like language systems. Instead, multilingual language competence comprises a compound system in which knowledge of one language is inseparable from knowledge of the others (Cook, 1992, 1995, 2002). As such, L2 learners can be viewed as developing multilinguals, engaged in an additive process of rapidly accumulating new linguistic experiences, and in a synthetic process of integrating new experiences into evolving structures in both L1 and L2 grammar (Cook, 1999).

Hall et al. (2006) argue for a usage-based approach to multicompetence research based on a wide variety of research in cognitive linguistics, language acquisition, and language variation. The present research provides a valuable way of examining the

emergence of particular levels of grammatical knowledge in individuals who are rapidly developing with regard to multicompetence.

6.6 Conclusion: Adult L2 learners as a testing ground for emergentist theory

Research on adult L2 learners from the emergentist perspective taken in this dissertation provides a valuable opportunity for providing a unique window on diverse strands of research concerning the human ability for language. It supports the notion that, while cognitive abilities and functioning are constantly evolving throughout the lifespan due to both developmental and experiential factors, there are no abrupt qualitative changes in the cognitive architecture underlying language acquisition, representation, and use. Instead, general learning abilities including classification, abstraction, and the tracking of statistical patterns operate on linguistic experience within a dynamically evolving cognitive environment.

Given this continuity of experience and learning throughout the lifespan, even the knowledge of a single, native language does not ever reach a stable end-state in which linguistic competence in that language can be said to be complete. Of course, adult monolingual competence (or relatively stable multilingual competence) develops for the most part very slowly and imperceptibly. For this reason, adult L2 learners, whose L2 knowledge is in a more rapid state of flux, are a valuable population for testing the implications of emergentist theory for adult language acquisition and use.

Adult L2 learning provides an environment in which new experiences are rapidly accumulating, and in which lexical networks and connections must evolve at a much

more rapid pace than in adults whose new language experience is largely consistent with their already-existing lexical networks. It is not that adults in the latter situation have ceased to track lexical statistics, but that their new experiences mostly serve to reinforce the statistical robustness and entrenchment of existing patterns. Second language experience, even in a classroom, provides significant amounts of language input that are not so easily fitted into existing patterns.

This is not to say that adult L2 acquisition of a given language proceeds in exactly the same way as L1 acquisition of the same language. It clearly does not. Rather, there are a large number of interacting factors that promote both change and stability in language representation and use, and thereby affect the course of L2 acquisition. In some cases, statistical learning mechanisms may operate on novel L2 input to create new patterns, perhaps simultaneously influencing established and similar patterns in the L1. This is evident, for example, in Spanish-English bilinguals who produce stops with VOT characteristics between the monolingual norms for both languages (Flege & Eefting, 1987) and in other research concerning L2 effects on the L1 (Cook, 2003). In other cases, the entrenchment of phonetic categories in one language may serve to obscure distinctions in another language that fall within a single category in the first (e.g. Sebastián-Gallés et al., 2005). In both cases, adult L2 learning provides a context of rapidly changing linguistic experience and representation that can serve to highlight those factors that promote change as well as those that promote stability in language representation and use, as related to adult cognition.

The research reported here makes an important step in this direction. It shows that adult L2 learners are sensitive to an intricate, probabilistic pattern in Spanish grammar,

namely diphthongization, but that this sensitivity is reflected to varying degrees at different levels of language processing and behavior. Importantly, this sensitivity was measured with neologisms, showing that learner performance reflects productive knowledge of gradience in Spanish diphthongization in certain morphological contexts. Native speaker performance regarding Spanish diphthongization also varies between levels of processing and behavior, but to a lesser extent and in different ways. Furthermore, the pattern of results observed in the learners was modulated by certain cognitive and experiential differences between learners. In particular, foreign word repetition ability appears to affect the processing of neologisms depending on certain processing and task conditions. These findings open a wide door to new ways of examining adult L2 learning in particular, and multilingualism in general, based on the interactions between specific experience of language and a constellation of general cognitive resources that serve to structure experience in a dynamic and emergent way.

References

- Achard, M. (1997). Cognitive Grammar and SLA Investigation. *Journal of Intensive English Studies*, 11(2), 157-176.
- Achard, M., & Niemeier, S. (2004a). Introduction: Cognitive Linguistics, Language Acquisition, and Pedagogy. In M. Achard & S. Niemeier (Eds.), *Cognitive Linguistics, Second Language Acquisition, and Foreign Language Teaching* (pp. 1-11). Berlin: Mouton de Gruyter.
- Achard, M., & Niemeier, S. (Eds.). (2004b). *Cognitive Linguistics, Second Language Acquisition, and Foreign Language Teaching*. Berlin: Mouton de Gruyter.
- Adams, A.-M., & Gathercole, S. E. (2000). Limitations in Working Memory: Implications for Language Development. *International Journal of Language and Communication Disorders*, 35(1), 95-116.
- Alameda, R., & Cuetos, F. (1995). *Diccionario De Frecuencias De Las Unidades Lingüísticas Del Castellano*. Oviedo: Servicio de publicaciones de la Universidad de Oviedo.
- Albright, A. (2002). Islands of Reliability for Regular Morphology: Evidence from Italian. *Language*, 78(4), 684-709.
- Albright, A., Andrade, A., & Hayes, B. (2001). Segmental Environments of Spanish Diphthongization. *UCLA Working Papers in Linguistics*, 7, 117-151.
- Albright, A., & Hayes, B. (2003). Rules vs. Analogy in English Past Tenses: A Computational/Experimental Study. *Cognition*, 90, 119-161.
- Álvarez, C. J., Carreiras, M., & Perea, M. (2004). Are Syllables Phonological Units in Visual Word Recognition?. *Language and Cognitive Processes*, 19(3), 427-452.
- Alvord, S. M. (2003). The Psychological Unreality of Quantity Sensitivity in Spanish: Experimental Evidence. *Southwest Journal of Linguistics*, 22(2), 1-12.
- Andrews, S. (1997). The Effect of Orthographic Similarity on Lexical Retrieval: Resolving Neighborhood Conflicts. *Psychonomic Bulletin & Review*, 4(4), 439-461.
- Aske, J. (1990). Disembodied Rules Versus Patterns in the Lexicon: Testing the Psychological Reality of Spanish Stress Rules. In K. Hall, J.-P. Koenig, M. Meacham, S. Reinman & L. A. Sutton (Eds.), *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistic Society* (pp. 30-45). Berkeley: Berkeley Linguistic Society.
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science*, 9(4), 321-324.
- Baayen, R. H. (2003). Probabilistic Approaches to Morphology. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 229-287). Cambridge, MA: MIT Press.
- Baayen, R. H., & Schreuder, R. (Eds.). (2003). *Morphological Structure in Language Processing*. Berlin: Mouton de Gruyter.

- Baddeley, A. (2003). Working Memory and Language: An Overview. *Journal of Communication Disorders*, 36, 189-208.
- Baddeley, A., Gathercole, S. E., & Papagno, C. (1998). The Phonological Loop as a Language Learning Device. *Psychological Review*, 105(1), 158-173.
- Bárkányi, Z. (2002). A Fresh Look at Quantity Sensitivity in Spanish. *Linguistics*, 40(2), 375-394.
- Barlow, M., & Kemmer, S. (Eds.). (2000). *Usage-Based Models of Language*. Stanford, CA: CSLI.
- Bates, E., & Goodman, J. C. (1999). On the Emergence of Grammar from the Lexicon. In B. MacWhinney (Ed.), *The Emergence of Language* (pp. 29-79). Mahwah, NJ: Lawrence Erlbaum.
- Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney & E. Bates (Eds.), *The Crosslinguistic Study of Sentence Processing* (pp. 3-76). Cambridge: Cambridge University Press.
- Berent, I., Shimron, J., & Vaknin, V. (2001). Phonological Constraints on Reading: Evidence from the Obligatory Contour Principle. *Journal of Memory and Language*, 44, 644-665.
- Best, C. (1995). A Direct Realistic Perspective on Cross-Language Speech Perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 171-206). Timonium, MD: York Press.
- Best, C. T., McRoberts, G., & 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.
- Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, Aging, and Cognitive Control: Evidence from the Simon Task. *Psychology and Aging*, 19(2), 290-303.
- Birdsong, D. (2005). Interpreting Age Effects in Second Language Acquisition. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 109-127). New York: Oxford University Press.
- 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* (pp. 41-68). Cambridge: Cambridge University Press.
- Bod, R., Hay, J., & Jannedy, S. (Eds.). (2003). *Probabilistic Linguistics*. Cambridge, MA: MIT Press.
- Boersma, P., & Hayes, B. (2001). Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry*, 32(1), 45-86.
- Boersma, P., & Weenink, D. (2006). Praat: Doing Phonetics by Computer (Version 4.3.14).
- Boylard, J. T. (2001). Hypercorrect Pronoun Case in English? Cognitive Processes That Account for Pronoun Usage. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the Emergence of Language Structure* (pp. 383-404). Amsterdam: John Benjamins.
- Browman, C. P., & Goldstein, L. (1992). Articulatory Phonology: An Overview. *Phonetica*, 49, 155-180.

- Burani, C., & Thornton, A. M. (2003). The Interplay of Root, Suffix, and Whole-Word Frequency in Processing Derived Words. In R. H. Baayen & R. Schreuder (Eds.), *Morphological Structure in Language Processing* (pp. 157-208). Berlin: Mouton de Gruyter.
- Bybee, J. L. (1985). *Morphology: A Study of the Relation between Meaning and Form*. Philadelphia: John Benjamins.
- Bybee, J. L. (2001a). Frequency Effects on French Liaison. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 337-360). Amsterdam: John Benjamins.
- Bybee, J. L. (2001b). *Phonology and Language Use*. New York: Cambridge University press.
- Bybee, J. L. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language*, 82(4), 711-733.
- Bybee, J. L., & Hopper, P. J. (2001a). Introduction to Frequency and the Emergence of Linguistic Structure. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 1-24). Amsterdam: John Benjamins.
- Bybee, J. L., & Hopper, P. J. (Eds.). (2001b). *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.
- Bybee, J. L., & Moder, C. L. (1983). Morphological Classes as Natural Categories. *Language*, 59, 251-270.
- Bybee, J. L., & Pardo, E. (1981). On Lexical and Morphological Conditioning of Alternations: A Nonce-Probe Experiment with Spanish Verbs. *Linguistics*, 19, 937-968.
- Carlson, M. T. (2006). The Development of Fine-Grained Phonological Knowledge in Adult Second Language Learners of Spanish. *Southwest Journal of Linguistics*, 25(2).
- Carreiras, M., & Perea, M. (2002). Masked Priming Effects with Syllabic Neighbors in a Lexical Decision Task. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5), 1228-1242.
- Carreiras, M., & Perea, M. (2004). Naming Pseudowords in Spanish: Effects of Syllable Frequency. *Brain and Language*, 90, 393-400.
- Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of Orthographic Neighborhood in Visual Word Recognition: Cross-Task Comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(4), 857-871.
- Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants Learn Phonotactic Regularities from Brief Auditory Experience. *Cognition*, 87, B69-B77.
- Chee, M. W. L., Soon, C. S., Lee, H. L., & Pallier, C. (2004). Left Insula Activation: A Marker for Language Attainment in Bilinguals. *Proceeding of the National Academy of Science, USA*, 101, 15265-15270.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.
- Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger.
- Clahsen, H., Eisenbeiss, S., & Sonnenstuhl-Henning, I. (1997). Morphological Structure and the Processing of Inflected Words. *Theoretical Linguistics*, 23(201-249).

- Clahsen, H., Rothweiler, A. W., & Marcus, G. F. (1993). Regular and Irregular Inflection in the Acquisition of German Noun Plurals. *Cognition*, 45, 225-255.
- Coady, J. A., & Aslin, R. N. (2003). Phonological Neighborhoods in the Developing Lexicon. *Journal of Child Language*, 30, 441-469.
- Coleman, J., & Pierrehumbert, J. B. (1997). Stochastic Phonological Grammars and Acceptability. In *Computational Phonology: Third Meeting of the ACL Special Interest Group in Computational Phonology* (pp. 49-56). Somerset, NJ: Association for Computational Linguistics.
- Collentine, J. (2004). The Effects of Learning Contexts on Morphosyntactic and Lexical Development. *Studies in Second Language Acquisition*, 26, 227-248.
- Cook, V. J. (1991). The Poverty-of-the-Stimulus Argument and Multicompetence. *Second Language Research*, 7(2), 103-117.
- Cook, V. J. (1992). Evidence for Multicompetence. *Language Learning*, 42(4), 557-591.
- Cook, V. J. (1995). Multi-Competence and the Learning of Many Languages. *Language, Culture, and Curriculum*, 8(2), 93-98.
- Cook, V. J. (1999). Going Beyond the Native Speaker in Language Teaching. *TESOL Quarterly*, 33(2), 185-209.
- Cook, V. J. (Ed.). (2002). *Portraits of the L2 User*. Clevedon: Multilingual Matters.
- Cook, V. J. (Ed.). (2003). *Effects of the Second Language on the First*. Clevedon: Multilingual Matters.
- Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant Melodies: Statistical Learning of Nonadjacent Dependencies in Tone Sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(5), 1119-1130.
- Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Cuetos, F., Mitchell, D. C., & Corley, M. M. B. (1996). Parsing in Different Languages. In M. Carreiras, J. E. García-Albea & N. Sebastián-Gallés (Eds.), *Language Processing in Spanish* (pp. 145-187). Mahwah, NJ: Lawrence Erlbaum.
- Daelemans, W., Zavrel, J., van der Sloot, K., & van den Bosch, A. (1999). *TiMBL: Tilburg Memory Based Learner, Version 2.0, Reference Guide*. Tilburg, Netherlands: ILK Research Group, Tilburg University.
- Davies, M. (2006). *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*. New York: Routledge.
- Davis, C. J., & Perea, M. (2005). Buscapalabras: A Program for Deriving Orthographic and Phonological Neighborhood Statistics and Other Psycholinguistic Indices in Spanish. *Behavior Research Methods*, 37(4), 665-671.
- de Jong, N. H., Schreuder, R., & Baayen, R. H. (2003). Morphological Resonance in the Mental Lexicon. In R. H. Baayen & R. Schreuder (Eds.), *Morphological Structure in Language Processing* (pp. 65-88). Berlin: Mouton de Gruyter.
- Dehaene-Lambertz, G., & Gliga, T. (2004). Common Neural Basis for Phoneme Processing in Infants and Adults. *Journal of Cognitive Neuroscience*, 16(8), 1375-1387.
- DeKeyser, R., & Larson-Hall, J. (2005). What Does the Critical Period Really Mean? In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 88-108). New York: Oxford University Press.

- Dell, G. S., & Gordon, J. K. (2003). Neighbors in the Lexicon: Friends or Foes? In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production* (pp. 9-37). Berlin: Mouton de Gruyter.
- Díaz-Campos, M. (2004). Context of Learning in the Acquisition of Spanish Second Language Phonology. *Studies in Second Language Acquisition*, 26, 249-273.
- Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of Cognates and Interlingual Homographs: The Neglected Role of Phonology. *Journal of Memory and Language*, 41, 496-518.
- Dijkstra, T., & Van Heuven, W. J. B. (1998). The BIA Model and Bilingual Word Recognition. In J. Grainger & A. M. Jacobs (Eds.), *Localist Connectionist Approaches to Human Cognition* (pp. 189-225). Mahwah, NJ: Lawrence Erlbaum.
- Eddington, D. (1996). Diphthongization in Spanish Derivational Morphology: An Empirical Investigation. *Hispanic Linguistics*, 8(1), 1-35.
- Eddington, D. (1998). Spanish Diphthongization as a Non-Derivational Phenomenon. *Rivista di Linguistica*, 10(2), 335-354.
- Eddington, D. (1999). Role of Experiments in Phonological Investigation. *Journal of Quantitative Linguistics*, 6(1), 14-28.
- Eddington, D. (2000). Spanish Stress Assignment within the Analogical Modeling of Language. *Language*, 76(1), 92-109.
- Eddington, D. (2002). A Comparison of Two Analogical Models: Tilburg Memory-Based Learner versus Analogical Modeling. In R. Skousen, D. Lonsdale & D. B. Parkinson (Eds.), *Analogical Modeling: An Exemplar-Based Approach to Language* (pp. 141-155). Amsterdam: John Benjamins.
- Eddington, D. (2004a). A Computational Approach to Resolving Certain Issues in Spanish Stress Placement. In T. L. Face (Ed.), *Laboratory Approaches to Spanish Phonology* (pp. 95-116). Berlin: Mouton de Gruyter.
- Eddington, D. (2004b). *Spanish Phonology and Morphology: Experimental and Quantitative Perspectives*. Amsterdam: John Benjamins.
- Edelman, S., Solan, Z., Horn, D., & Ruppín, E. (2003). Rich Syntax from a Raw Corpus: Unsupervised Does It.
- 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(3), 356-371.
- Ellis, N. C. (2002a). Frequency Effects in Language Processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition. *Studies in Second Language Acquisition*, 24, 143-188.
- Ellis, N. C. (2002b). Reflections on Frequency Effects in Language Processing. *Studies in Second Language Acquisition*, 24, 297-339.
- Ellis, N. C. (2006). Cognitive Perspectives on SLA: The Associative-Cognitive CREED. *AILA Review*, 19, 100-121.
- Ellis, N. C., & Beaton, A. (1993). Factors Affecting the Learning of Foreign Language Vocabulary: Imagery Keyword Mediators and Phonological Short-Term Memory. *The Quarterly Journal of Experimental Psychology*, 46A(3), 533-558.

- Ellis, N. C., & Sinclair, S. G. (1996). Working Memory in the Acquisition of Vocabulary and Syntax: Putting Language in Good Order. *The Quarterly Journal of Experimental Psychology*, 49A(1), 234-250.
- Face, T. L. (2000). The Role of Syllable Weight in the Perception of Spanish Stress. In H. Campos, E. Herburger, A. Morales-Front & T. J. Walsh (Eds.), *Hispanic Linguistics at the Turn of the Millennium: Papers from the 3rd Hispanic Linguistics Symposium* (pp. 1-13). Somerville, MA: Cascadilla.
- Face, T. L. (2004). Perceiving What Isn't There: Non-Acoustic Cues for Perceived Spanish Stress. In T. L. Face (Ed.), *Laboratory Approaches to Spanish Phonology* (pp. 117-141). Berlin: Mouton de Gruyter.
- Flege, J. E. (2003). Assessing Constraints on Second-Language Segmental Production and Perception. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 319-355). Berlin: Mouton de Gruyter.
- 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., 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., & MacKay, I. R. A. (2004). Perceiving Vowels in a Second 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.
- Ford, M. A., Marslen-Wilson, W. D., & Davis, M. H. (2003). Morphology and Frequency: Contrasting Methodologies. In R. H. Baayen & R. Schreuder (Eds.), *Morphological Structure in Language Processing* (pp. 89-124). Berlin: Mouton de Gruyter.
- Friedman, N. P., & Miyake, A. (2005). Comparison of Four Scoring Methods for the Reading Span Test. *Behavior Research Methods*, 37(4), 581-590.
- Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of Wordlikeness: Effects of Segment Probability and Length on the Processing of Nonwords. *Journal of Memory and Language*, 42, 481-496.
- Frisch, S. A., Large, N. R., Zawaydeh, B., & Pisoni, D. B. (2001). Emergent Phonotactic Generalizations in English and Arabic. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 159-180). Amsterdam: John Benjamins.
- Frisch, S. A., & Zawaydeh, B. (2001). The Psychological Reality of OCP-Place in Arabic. *Language*, 77(1), 91-106.
- Gathercole, S. E. (1995). Is Nonword Repetition a Test of Phonological Memory or Long-Term Knowledge? It All Depends on the Nonwords. *Memory & Cognition*, 23(1), 83-94.
- Gathercole, S. E., Willis, C., Baddeley, A., & Emslie, H. (1994). The Children Test of Nonword Repetition: A Test of Phonological Working Memory. *Memory*, 2, 103-127.
- Gibson, E., & Pearlmutter, N. J. (1994). A Corpus-Based Analysis of Psycholinguistic Constraints on Prepositional Phrase Attachment. In C. Clifton, Jr., L. Frazier & K.

- Rayner (Eds.), *Perspectives on Sentence Processing* (pp. 181-198). Hillsdale, NJ: Lawrence Erlbaum.
- Giraud, H., & Grainger, J. (2003). On the Role of Derivational Affixes in Recognizing Complex Words: Evidence from Masked Priming. In R. H. Baayen & R. Schreuder (Eds.), *Morphological Structure in Language Processing* (pp. 209-232). Berlin: Mouton de Gruyter.
- Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldstein, L., & Fowler, C. A. (2003). Articulatory Phonology: A Phonology for Public Language Use. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production* (pp. 159-207). Berlin: Mouton de Gruyter.
- Goldstone, R. L. (2000). Unitization during Category Learning. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1), 86-112.
- Gomez, R. L., & Gerken, L. (1999). Artificial Grammar Learning by 1-Year-Olds Leads to Specific and Abstract Knowledge. *Cognition*, 70, 109-135.
- Gomez, R. L., & Gerken, L. (2000). Infant Artificial Language Learning and Language Acquisition. *Trends in Cognitive Sciences*, 4(5), 178-186.
- Green, D. W. (1998). Mental Control of the Bilingual Lexico-Semantic System. *Bilingualism: Language and Cognition*, 1, 67-81.
- Gregg, K. R. (1984). Krashen's Monitor and Occam's Razor. *Applied Linguistics*, 5(2), 79-100.
- Hall, J. K., Cheng, A., & Carlson, M. T. (2006). Reconceptualizing Multicompetence as a Theory of Language Knowledge. *Applied Linguistics*, 27(2), 220-240.
- Halle, M., Harris, J. W., & Vergnaud, J.-R. (1991). A Reexamination of the Stress Erasure Convention and Spanish Stress. *Linguistic Inquiry*, 22, 141-159.
- Hallé, 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.
- Harris, J. W. (1969). *Spanish Phonology*. Cambridge, MA: MIT Press.
- Harris, J. W. (1977). Remarks on Diphthongization in Spanish. *Lingua*, 41, 261-305.
- Harris, J. W. (1995). Projection and Edge Marking in the Computation of Stress in Spanish. In J. A. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 867-887). Cambridge, MA: Blackwell.
- Hay, J., & Baayen, R. H. (2005). Shifting Paradigms: Gradient Structure in Morphology. *Trends in Cognitive Sciences*, 9(7), 342-348.
- Hay, J., Pierrehumbert, J. B., & Beckman, M. (2004). Speech Perception, Well-Formedness, and the Statistics of the Lexicon. In R. O. J. Local, and R. Temple (Ed.), *Papers in Laboratory Phonology VI* (pp. 58-74). Cambridge, UK: Cambridge University Press.
- Hazenbergh, S., & Hulstijn, J. (1996). Defining a Minimal Receptive Second-Language Vocabulary for Non-Native University Students: An Empirical Investigation. *Applied Linguistics*, 17, 145-163.
- Hooper, J. B., & Terrell, T. (1976). Stress Assignment in Spanish: A Natural Generative Analysis. *Glossa*, 10, 64-110.

- Hopper, P. J. (1998). Emergent Grammar. In M. Tomasello (Ed.), *The New Psychology of Language* (pp. 155-175). Mahwah, NJ: Lawrence Erlbaum.
- Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development, 1*(2), 151-195.
- Hume, E., & Johnson, K. (Eds.). (2001). *The Role of Speech Perception in Phonology*. New York: Academic Press.
- Jurafsky, D. (2003). Probabilistic Modeling in Psycholinguistics: Linguistic Comprehension and Production. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 39-96). Cambridge, MA: MIT Press.
- Jurafsky, D., & Martin, J. H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, NJ: Prentice Hall.
- Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' Sensitivity to Phonotactic Patterns in the Native Language. *Journal of Memory and Language, 33*, 630-645.
- Kim, C. W., & Park, S. G. (1995). Pronunciation Problems of Australian Students. *Australian Review of Applied Linguistics, Supplement 12*, 183-202.
- Klein, D. (2005). *The Unsupervised Learning of Natural Language Structure*. Unpublished Doctoral Dissertation, Stanford University.
- Krashen, S. D. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press.
- Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press.
- Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The Development of Lexical Fluency in a Second Language. *Second Language Research, 18*(2), 137-171.
- Kuhl, P. K. (2000). A New View of Language Acquisition. *Proceedings of the National Academy of Science, 97*, 11850-11857.
- 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*, 606-608.
- Landauer, T. K., & Streeter, L. A. (1973). Structural Differences between Common and Rare Words: Failure or Equivalence Assumptions for Theories of Word Recognition. *Journal of Verbal Learning & Verbal Behavior, 12*, 119-131.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar* (Vol. 1). Stanford, CA: Stanford University Press.
- Langacker, R. W. (1991). *Foundations of Cognitive Grammar* (Vol. 2). Stanford, CA: Stanford University Press.
- Langacker, R. W. (2000). A Dynamic Usage-Based Model. In M. Barlow & S. Kemmer (Eds.), *Usage-Based Models of Language* (pp. 1-63). Stanford, CA: CSLI.
- Laufer, B., & Nation, I. S. P. (1999). A Vocabulary-Size Test of Controlled Productive Ability. *Language Testing, 16*, 33-51.
- Lee, S.-O., Fouser, R., & Baxter, D. (Eds.). (1995) *Basic Korean Dictionary: Korean-English/English-Korean* (Revised ed.). Elizabeth, NJ: Hollym International.

- Lehtonen, M., & Laine, M. (2003). How Word Frequency Affects Morphological Processing in Monolinguals and Bilinguals. *Bilingualism: Language and Cognition*, 6(3), 213-225.
- Linck, J. A. (2005). *Language and Cognitive Processing in a Study Abroad Context*. Unpublished Master's Thesis, The Pennsylvania State University, University Park
- Luce, P. A., & Large, N. R. (2001). Phonotactics, Density, and Entropy in Spoken Word Recognition. *Language and Cognitive Processes*, 16(5/6), 565-581.
- Macizo, P., & Van Petten, C. (2007). Syllable Frequency in Lexical Decision and Naming of English Words. *Reading and Writing*, 20(4), 295-331.
- MacKay, I. R. A., Flege, J. E., Piske, T., & Schirru, C. (2001). Category Restructuring During Second-Language Speech Acquisition. *Journal of the Acoustical Society of America*, 110(1), 516-528.
- MacWhinney, B. (2001). Emergentist Approaches to Language. In J. Bybee & P. J. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 449-470). Amsterdam: John Benjamins.
- MacWhinney, B. (2005a). New Directions in the Competition Model. In M. Tomasello & D. I. Slobin (Eds.), *Beyond Nature-Nurture* (pp. 81-110). Mahwah, NJ: Lawrence Erlbaum.
- MacWhinney, B. (2005b). A Unified Model of Language Acquisition. In J. F. Kroll & A. M. B. DeGroot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 49-67). New York: Oxford University Press.
- MacWhinney, B. (2007). A Unified Model. In N. C. Ellis & P. Robinson (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- MacWhinney, B. (Ed.). (1999). *The Emergence of Language*. Mahwah, NJ: Lawrence Erlbaum.
- MacWhinney, B., & Snow, C. (1985). The Child Language Exchange System. *Journal of Computational Linguistics*, 12, 271-296.
- Magnuson, J. S. (2005). Moving Hand Reveals Dynamics of Thought. *Proceeding of the National Academy of Science, USA*, 102(29), 9995-9996.
- Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The Dynamics of Lexical Competition during Spoken Word Recognition. *Cognitive Science*, 31, 1-24.
- Manning, C. D. (2003). Probabilistic Syntax. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 289-341). Cambridge, MA: MIT Press.
- Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., Woest, A., & Pinker, S. (1995). German Inflection: The Exception That Proves the Rule. *Cognitive Psychology*, 29(189-256).
- Maye, J., Weiss, D., & Aslin, R. N. (In Press). Statistical Phonetic Learning in Infants: Facilitation and Feature Generalization. *Developmental Science*.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant Sensitivity to Distributional Information Can Affect Phonetic Discrimination. *Cognition*, 82, B101-B111.
- McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural Correlates of Second-Language Word Learning: Minimal Instruction Produces Rapid Change. *Nature Neuroscience*, 7(7), 703-704.

- Meunier, F., & Segui, J. (1999). Morphological Priming Effect: The Role of Surface Frequency. *Brain and Language*, 68, 54-60.
- Michas, I. C., & Henry, L. A. (1994). The Link between Phonological Memory and Vocabulary Acquisition. *British Journal of Developmental Psychology*, 12, 147-163.
- Miyake, A., & Friedman, N. P. (1998). Individual Differences in Second Language Proficiency: Working Memory as Language Aptitude. In A. F. Healy & L. E. Bourne Jr. (Eds.), *Foreign Language Learning: Psycholinguistic Studies on Training and Retention* (pp. 339-364). Mahwah, NJ: Lawrence Erlbaum.
- Moritz, C. E. B. (1995). *Self-Assessment of Foreign Language Proficiency: A Critical Analysis of Issues and a Study of Cognitive Orientations of French Learners*. Unpublished Doctoral Dissertation, Cornell University.
- Newport, E. L. (1990). Maturation Constraints on Language Learning. *Cognitive Science*, 14, 11-28.
- Newport, E. L. (1991). Contrasting Conceptions of the Critical Period for Language. In S. Carey & R. Gelman (Eds.), *The Epigenesis of Mind* (pp. 111-130). Hillsdale, NJ: Lawrence Erlbaum.
- Newport, E. L. (1999). Reduced Input in the Acquisition of Signed Languages: Contributions to the Study of Creolization. In M. Degraff (Ed.), *Creolization, Diachrony, and Language Acquisition* (pp. 161-178). Cambridge, MA: MIT Press.
- Newport, E. L., & Aslin, R. N. (2000). Innately Constrained Learning: Blending Old and New Approaches to Language Acquisition. In S. C. Howell, S. A. Fish & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 1-21). Somerville, MA: Cascadilla Press.
- Nooteboom, S., Weerman, F., & Wijnen, F. (Eds.). (2002). *Storage and Computation in the Language Faculty*. Dordrecht: Kluwer.
- Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning Phonotactic Constraints from Brief Auditory Experience. *Cognition*, 83, B13-B23.
- Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The Influence of Native-Language Phonology on Lexical Access. *Psychological Science*, 12(6), 445-449.
- Perea, M., Rosa, E., & Gómez, C. (2005). The Frequency Effect for Pseudowords in the Lexical Decision Task. *Perception and Psychophysics*, 67(2), 301-314.
- Phillips, B. S. (2001). Lexical Diffusion, Lexical Frequency, and Lexical Analysis. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 123-136). Amsterdam: John Benjamins.
- Pierrehumbert, J. B. (1999). What People Know About Sounds of Language. *Studies in the Linguistic Sciences*, 29(2), 111-120.
- Pierrehumbert, J. B. (2000). The Phonetic Grounding of Phonology. *Bulletin de la Communication Parlee*, 5, 7-23.
- Pierrehumbert, J. B. (2001a). Exemplar Dynamics: Word Frequency, Lenition, and Contrast. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 136-157). Amsterdam: John Benjamins.
- Pierrehumbert, J. B. (2001b). Stochastic Phonology. *Glott international*, 5(6), 195-207.

- Pierrehumbert, J. B. (2003a). Phonetic Diversity, Statistical Learning, and Acquisition of Phonology. *Language and Speech*, 46(2-3), 115-154.
- Pierrehumbert, J. B. (2003b). Probabilistic Phonology: Discrimination and Robustness. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics*. Cambridge, MA: MIT Press.
- Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York: William Morrow and Company.
- Pinker, S. (1998). Words and Rules. *Lingua*, 106, 219-242.
- Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. New York: Basic Books.
- Pinker, S., & Ullman, M. T. (2002). The Past and Future of the Past Tense. *Trends in Cognitive Sciences*, 6, 456-463.
- 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.
- Rastle, K., & Davis, M. H. N., Boris. (2004). The Broth in My Brother's Brothel: Morpho-Orthographic Segmentation in Visual Word Recognition. *Psychonomic Bulletin & Review*, 11(6), 1090-1098.
- Real Academia Española. (Ed.) (1970). *Diccionario De La Lengua Española*. Madrid: Real Academia Española.
- Real Academia Española (Ed.) (1995) *Diccionario de la lengua Española*. Madrid: Espasa Calpe.
- Roca, I. (1988). Theoretical Implications of Spanish Word Stress. *Linguistic Inquiry*, 19, 393-423.
- Roca, I. (1991). Stress and Syllables in Spanish. In H. Campos & F. Martínez-Gil (Eds.), *Current Studies in Spanish Linguistics* (pp. 599-635). Washington, DC: Georgetown University Press.
- Ross, D. S., & Newport, E. L. (1996). The Development of Language from Non-Native Linguistic Input. In A. Stringfellow, D. Cahana-Amitay, E. Hughs & A. Zukowski (Eds.), *Proceedings of the 20th Annual Boston University Conference on Language Development* (pp. 623-645). Boston: Cascadilla.
- Saffran, J. R. (2003). Statistical Language Learning: Mechanisms and Constraints. *Current Directions in Psychological Science*, 12(4), 110-114.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Olds. *Science*, 274, 1926-1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical Learning of Tone Sequences by Human Infants and Adults. *Cognition*, 70, 27-52.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, 35, 606-621.
- Saffran, J. R., & Wilson, D. P. (2003). From Syllables to Syntax: Multilevel Statistical Learning by 12-Month-Old Infants. *Infancy*, 4(2), 273-284.
- Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting Nonsense: An Event-Related Potential Index of Perceived Onsets in Continuous Speech. *Nature Neuroscience* 5(7), 700-703.
- Schmitt, N. (2004). *Formulaic Sequences*. Amsterdam: John Benjamins.

- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and Exploring the Behavior of Two New Versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55-88.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools, Inc.
- Schwartz, B. D. (1986). The Epistemological Status of Second Language Acquisition. *Second Language Research, 2*(2), 120-159.
- Schwartz, B. D. (1990). Un-Motivating the Motivation for the Fundamental Difference Hypothesis. In H. Burmeister & P. L. Rounds (Eds.), *Variability in Second Language Acquisition: Proceedings from the Tenth Meeting of the Second Language Research Forum* (pp. 667-684). Eugene, OR: American English Institute.
- Sebastián-Gallés, N., & Bosch, L. (2002). Building Phonotactic Knowledge in Bilinguals: Role of Early Exposure. *Journal of Experimental Psychology: Human Perception and Performance, 28*(4), 974-989.
- Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The Influence of Initial Exposure on Lexical Representation: Comparing Early and Simultaneous Bilinguals. *Journal of Memory and Language, 52*(2), 240-255.
- Sebastián-Gallés, N., & Kroll, J. F. (2003). Phonology in Bilingual Language Processing: Acquisition, Perception, and Production. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 279-317). Berlin: Mouton de Gruyter.
- Sebastián-Gallés, N., Martí, M. A., Carreiras, M., & Cuetos, F. (2000). LexEsp, Léxico Informatizado Del Español. Barcelona: Ediciones de la Universitat de Barcelona.
- Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online Processing of Native and Non-Native Phonemic Contrasts in Early Bilinguals. *Cognition, 72*, 111-123.
- Service, E. (1992). Phonology, Working Memory and Foreign-Language Learning. *Quarterly Journal of Experimental Psychology, 45A*, 21-50.
- Service, E., & Craik, F. I. M. (1993). Differences between Young and Older Adults in Learning a Foreign Vocabulary. *Journal of Memory and Language, 32*, 608-623.
- Service, E., & Kohonen, V. (1995). Is the Relation between Phonological Memory and Foreign Language Learning Accounted for by Vocabulary Acquisition? *Applied Psycholinguistics, 16*, 155-172.
- Simon, J. R., & Ruddell, A. P. (1967). Auditory S-R Compatibility: The Effect of an Irrelevant Cue on Information Processing. *Journal of Applied Psychology, 51*, 300-304.
- Singleton, J. L., & Newport, E. L. (2004). When Learners Surpass Their Models: The Acquisition of American Sign Language from Impoverished Input. *Cognitive Psychology, 49*, 370-407.
- Skousen, R. (1989). *Analogical Modeling of Language*. Dordrecht: Kluwer.
- Skousen, R. (1992). *Analogy and Structure*. Dordrecht: Kluwer.
- Skousen, R. (2002). An Overview of Analogical Modeling. In R. Skousen, D. Lonsdale & D. B. Parkinson (Eds.), *Analogical Modeling: An Exemplar-Based Approach to Language* (pp. 11-26). Amsterdam: John Benjamins.

- Solan, Z., Horn, D., Ruppin, E., & Edelman, S. (2003). Unsupervised Efficient Learning and Representation of Language Structure. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum.
- Solan, Z., Horn, D., Ruppin, E., & Edelman, S. (2004). Unsupervised Context Sensitive Language Acquisition from a Large Corpus. In L. Saul (Ed.), *Advances in Neural Information Processing*. Cambridge, MA: MIT Press.
- Solan, Z., Ruppin, E., Horn, D., & Edelman, S. (2002). Automatic Acquisition and Efficient Representation of Syntactic Structures.
- Sonnenstuhl-Henning, I., & Huth, A. (2002). Processing and Representation of German *-n* Plurals: A Dual Process Approach. *Brain and Language, 81*, 276-290.
- Soto-Faraco, S., Sebastian-Galles, N., & Cutler, A. (2001). Segmental and Suprasegmental Mismatch in Lexical Access. *Journal of Memory and Language, 45*(3), 412-432.
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous Attraction toward Phonological Competitors. *Proceeding of the National Academy of Science, USA, 102*(29), 10393-10398.
- Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., et al. (2004). A New on-Line Resource for Psycholinguistic Studies. *Journal of Memory and Language, 51*(2), 247-250.
- Thiessen, E. D., & Saffran, J. R. (2003). When Cues Collide: Use of Stress and Statistical Cues to Word Boundaries by 7- to 9-Month-Old Infants. *Developmental Psychology, 39*(4), 706-716.
- Thompson, S. P., & Newport, E. L. (2007). Statistical Learning of Syntax: The Role of Transitional Probability. *Language Learning and Development, 3*(1), 1-42.
- Tomasello, M. (2000). The Item-Based Nature of Children's Early Syntactic Development. *Trends in Cognitive Sciences, 4*(4), 156-163.
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Ullman, M. T. (1999). Acceptability Ratings of Regular and Irregular Past-Tense Forms: Evidence for a Dual-System Model of Language from Word Frequency and Phonological Neighbourhood Effects. *Language and Cognitive Processes, 14*(1), 47-67.
- Ullman, M. T. (2004). Contributions of Memory Circuits to Language: The Declarative/Procedural Model. *Cognition, 92*, 231-270.
- Vitevitch, M. S. (1997). The Neighborhood Characteristics of Malapropisms. *Language and Speech, 40*(3), 211-228.
- Vitevitch, M. S. (2002). The Influence of Phonological Similarity Neighborhoods on Speech Production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 735-747.
- Vitevitch, M. S., & Luce, P. A. (1998). When Words Compete: Levels of Processing in Perception of Spoken Words. *Psychological Science, 9*(4), 325-329.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic Phonotactics and Neighborhood Activation in Spoken Word Recognition. *Journal of Memory and Language, 40*, 374-408.

- Vitevitch, M. S., & Luce, P. A. (2005). Increases in Phonotactic Probability Facilitate Spoken Nonword Repetition. *Journal of Memory and Language*, 52(2), 193-204.
- Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, Neighborhood Activation, and Lexical Access for Spoken Words. *Brain and Language*, 68, 306-311.
- Vitevitch, M. S., & Sommers, M. S. (2003). The Facilitative Influence of Phonological Similarity and Neighborhood Frequency in Speech Production in Younger and Older Adults. *Memory & Cognition*, 31(4), 491-504.
- Vitevitch, M. S., & Stamer, M. K. (2006). The Curious Case of Competition in Spanish Speech Production. *Language and Cognitive Processes*, 21(6), 760-770.
- Waltermire, M. (2004). The Effect of Syllable Weight on the Determination of Spoken Stress in Spanish. In T. L. Face (Ed.), *Laboratory Approaches to Spanish Phonology* (pp. 171-191). Berlin: Mouton de Gruyter.
- Waters, G. S., & Caplan, D. (1996). The Measurement of Verbal Working Memory Capacity and Its Relation to Reading Comprehension. *Quarterly Journal of Experimental Psychology*, 49(4), 51-79.
- White, L. (2003). *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.
- Williams, J. N., & Lovatt, P. (2003). Phonological Memory and Rule Learning. *Language Learning*, 53(1), 67-121.
- Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., et al. (1999). Brain Responses Reveal the Learning of Foreign Language Phonemes. *Psychophysiology*, 36, 638-642.
- Wonnacott, E., & Newport, E. L. (2005). Novelty and Regularization: The Effect of Novel Instances on Rule Formation. In A. Brugos, M. R. Clark-Cotton & S. Ha (Eds.), *Proceedings of the 29th Annual Boston University Conference on Language Development* (pp. 663-673). Somerville, MA: Cascadilla.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Zamuner, T. S., Gerken, L., & Hammond, M. (2005). The Acquisition of Phonology Based on Input: A Closer Look at the Relation of Cross-Linguistic and Child Language Data. *Lingua*, 115, 1403-1426.
- Ziegler, J. C., Ferrand, L., & Montant, M. (2004). Visual Phonology: The Effects of Orthographic Consistency on Different Auditory Word Recognition Tasks. *Memory & Cognition*, 32(5), 732-741.
- Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood Effects in Auditory Word Recognition: Phonological Competition and Orthographic Facilitation. *Journal of Memory and Language*, 48, 779-793.
- Zuraw, K. (2003). Probability in Language Change. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 139-176). Cambridge, MA: MIT Press.

Appendix A

Test items for the lexical decision and conditional naming tasks Table A-1

Table A-1: Test neologisms with possible diphthongs, by stem. Note: each participant heard only one version of each neologism, with either a monophthong or a diphthong

Stem	Neologism with Monophthong	Neologism with Diphthong	Suffix
abiert	abertista	abiertista	ista
abiert	abertito	abiertito	ito
acuest	acostito	acuestito	ito
alient	alentoso	alientoso	oso
alient	alentista	alientista	ista
aparien	aparentísimo	apariantísimo	ísimo
apriet	apretoso	aprietoso	oso
apuest	apostero	apuestero	ero
asient	asentísimo	asientísimo	ísimo
atiend	atendísimo	atiendísimo	ísimo
atravies	atravesero	atraviesero	ero
calient	calental	caliental	al
calient	calentero	calientero	ero
cieg	cegal	ciegal	al
cieg	cegazo	ciegazo	azo
cierr	cerrazo	cierrazo	azo
ciert	certísimo	ciertísimo	ísimo
comienz	comenzal	comienzal	al
comienz	comenzazo	comienzazo	azo
conciert	concertal	conciertal	al
conciert	concertito	conciertito	ito
confies	confesal	confiesal	al
confies	confesista	confiesista	ista
consuel	consolal	consuelal	al
consuel	consolero	consuelero	ero
conviert	convertista	conviertista	ista
conviert	convertito	conviertito	ito
cuent	contal	cuental	al
cuerp	corpito	cuerpito	ito
defiend	defendito	defiendito	ito
dependent	dependentito	dependentito	ito
desplieg	desplegazo	despliegazo	azo
dient	dentoso	dientoso	oso

dient	dentero	dientero	ero
diviert	divertito	diviertito	ito
diviert	divertazo	diviertazo	azo
duerm	dormísimo	duermísimo	ísimo
enciend	encendísimo	enciendísimo	ísimo
entiend	entendero	entiendero	ero
extiend	extendista	extiendista	ista
fiebr	febrísimo	fiebrísimo	ísimo
fiest	festal	fiestal	al
fiest	festísimo	fiestísimo	ísimo
fuent	fontal	fuental	al
fuert	fortista	fuertista	ista
fuert	fortazo	fuertazo	azo
grues	grosal	gruesal	al
hiel	helal	hielal	al
hiel	helazo	hielazo	azo
huert	hortazo	huertazo	azo
huesped	hospedoso	huespedoso	oso
huesped	hospedito	huespedito	ito
inmuebl	inmoblista	inmueblista	ista
inviern	invernero	inviernero	ero
inviern	invernito	inviernito	ito
lluev	llovito	lluevito	ito
lluev	llovazo	lluevazo	azo
manifiest	manifestero	manifiestero	ero
manifiest	manifestísimo	manifiestísimo	ísimo
muerd	mordero	muerdo	ero
muerd	mordísimo	muerdísimo	ísimo
muert	mortoso	muertoso	oso
muestr	mostrero	muestrero	ero
muev	movísimo	muevísimo	ísimo
niebl	neblosa	nieblosa	oso
niebl	neblazo	nieblazo	azo
nieg	negazo	niegazo	azo
niev	nevoso	nievoso	oso
niev	neval	nieval	al
pierd	perdisto	pierdisto	ista
pierd	perdísimo	pierdísimo	ísimo
plieg	pleguito	plieguito	ito
puebl	poblosa	pueblosa	oso
puebl	poblazo	pueblazo	azo
pued	podísimo	puedísimo	ísimo
puert	portito	puertito	ito
quiebr	quebroso	quiebroso	oso
quiebr	quebrero	quebrero	ero

recomiend	recomendal	recomiendal	al
recuerd	recordero	recuerdero	ero
refier	referísimo	refierísimo	ísimo
revuelt	revoltista	revueltista	ista
revuelt	revoltazo	revueltazo	azo
rued	rodal	ruedal	al
rueg	roguista	rueguista	ista
rueg	roguísimo	rueguísimo	ísimo
sentimient	sentimentero	sentimentero	ero
sentimient	sentimentazo	sentimentazo	azo
siembr	sembrista	siembrista	ista
siembr	sembrito	siembrito	ito
suen	sonoso	suenoso	oso
suen	sonal	suenal	al
sueñ	soñero	sueñero	ero
sueñ	soñista	sueñista	ista
supuest	supostista	supuestista	ista
tiembl	temblista	temblista	ista
tiembl	temblito	temblito	ito
tiern	ternoso	tiernoso	oso
tierr	terral	tierral	al
tierr	terrista	tierrista	ista
travies	travesísimo	travesísimo	ísimo
travies	travesazo	travesazo	azo
tropiec	tropezoso	tropiezoso	oso
tropiec	tropecero	tropiecero	ero
valient	valentoso	valentoso	oso
vergüenz	vergoncista	vergüencista	ista
vergüenz	vergonzazo	vergüenzazo	azo
viej	vejoso	viejoso	oso
viej	vejito	viejito	ito
vuel	voloso	vueloso	oso
vuelc	volcoso	vuelcoso	oso
vuelc	voltero	vueltero	ero

Table A-2: Distractors: nonwords, words, and words with suffixes

Item	Type
amuela	Distractor nonword
anetuesto	Distractor nonword
anuentiti3n	Distractor nonword
asajaor	Distractor nonword
asaldiensa	Distractor nonword
betienso	Distractor nonword

carufitia	Distractor nonword
casolueda	Distractor nonword
cutarir	Distractor nonword
dagatal	Distractor nonword
datinir	Distractor nonword
demijo	Distractor nonword
desbenseto	Distractor nonword
desorebio	Distractor nonword
dilumico	Distractor nonword
disnaso	Distractor nonword
elieso	Distractor nonword
emueco	Distractor nonword
estojoso	Distractor nonword
ipante	Distractor nonword
iquienta	Distractor nonword
isuenra	Distractor nonword
larcoleta	Distractor nonword
letiento	Distractor nonword
licote	Distractor nonword
mantrasada	Distractor nonword
marseleco	Distractor nonword
matraja	Distractor nonword
migalo	Distractor nonword
pefusina	Distractor nonword
pentajatio	Distractor nonword
pinlataca	Distractor nonword
piquitienta	Distractor nonword
probitajo	Distractor nonword
quebudar	Distractor nonword
quenanista	Distractor nonword
queponturo	Distractor nonword
radanso	Distractor nonword
ritotitria	Distractor nonword
rodotesal	Distractor nonword
rosisad	Distractor nonword
samiguilo	Distractor nonword
soguira	Distractor nonword
soperto	Distractor nonword
tajintana	Distractor nonword
tamodar	Distractor nonword
trapudo	Distractor nonword
uestarira	Distractor nonword
uetusa	Distractor nonword
umedarta	Distractor nonword
adecuada	Distractor word

antes	Distractor word
aparente	Distractor word
aquel	Distractor word
aquí	Distractor word
atender	Distractor word
ayer	Distractor word
básica	Distractor word
buena	Distractor word
cinco	Distractor word
confieso	Distractor word
derecho	Distractor word
destruir	Distractor word
diecisiete	Distractor word
echarse	Distractor word
enorme	Distractor word
entera	Distractor word
española	Distractor word
esperar	Distractor word
estar	Distractor word
expediente	Distractor word
frecuente	Distractor word
fresca	Distractor word
fugaz	Distractor word
honda	Distractor word
insólito	Distractor word
íntima	Distractor word
inventar	Distractor word
izquierda	Distractor word
izquierdo	Distractor word
juego	Distractor word
llega	Distractor word
luego	Distractor word
médico	Distractor word
mientras	Distractor word
mucha	Distractor word
necesario	Distractor word
ocho	Distractor word
oculta	Distractor word
orgánica	Distractor word
peor	Distractor word
poder	Distractor word
pronto	Distractor word
propietario	Distractor word
rápida	Distractor word
redondo	Distractor word

representar	Distractor word
seguir	Distractor word
siete	Distractor word
siguiente	Distractor word
simple	Distractor word
tratarse	Distractor word
veinte	Distractor word
viuda	Distractor word
volver	Distractor word
vulgar	Distractor word
acción	Distractor word
base	Distractor word
calle	Distractor word
camino	Distractor word
cliente	Distractor word
clima	Distractor word
coincidencia	Distractor word
conveniencia	Distractor word
crisis	Distractor word
cuestión	Distractor word
cultura	Distractor word
desenlace	Distractor word
dinero	Distractor word
dirigente	Distractor word
eco	Distractor word
esfuerzo	Distractor word
especie	Distractor word
europa	Distractor word
existencia	Distractor word
experiencia	Distractor word
forma	Distractor word
futuro	Distractor word
gato	Distractor word
gente	Distractor word
hierba	Distractor word
historia	Distractor word
huevo	Distractor word
iglesia	Distractor word
índole	Distractor word
intuición	Distractor word
lado	Distractor word
lanzamiento	Distractor word
liga	Distractor word
lugar	Distractor word
mayoría	Distractor word

mención	Distractor word
miedo	Distractor word
monarquía	Distractor word
movimiento	Distractor word
niña	Distractor word
nombramiento	Distractor word
número	Distractor word
país	Distractor word
pariente	Distractor word
perfil	Distractor word
pierna	Distractor word
presencia	Distractor word
presidente	Distractor word
problema	Distractor word
promesa	Distractor word
reino	Distractor word
respuesta	Distractor word
riesgo	Distractor word
rostro	Distractor word
ruina	Distractor word
saliva	Distractor word
secuencia	Distractor word
secuestro	Distractor word
serie	Distractor word
siglo	Distractor word
sopón	Distractor word
suerte	Distractor word
trance	Distractor word
ventaja	Distractor word
abuelita	Distractor word, suffixed
altísimo	Distractor word, suffixed
amante	Distractor word, suffixed
americano	Distractor word, suffixed
amorosa	Distractor word, suffixed
ansiedad	Distractor word, suffixed
artista	Distractor word, suffixed
bellísima	Distractor word, suffixed
brillante	Distractor word, suffixed
cabrito	Distractor word, suffixed
camarero	Distractor word, suffixed
capacidad	Distractor word, suffixed
castellano	Distractor word, suffixed
cercano	Distractor word, suffixed
ciclista	Distractor word, suffixed
ciudadana	Distractor word, suffixed

codazo	Distractor word, suffixed
consejero	Distractor word, suffixed
cotidiana	Distractor word, suffixed
creatividad	Distractor word, suffixed
cristiano	Distractor word, suffixed
curioso	Distractor word, suffixed
delantero	Distractor word, suffixed
despacito	Distractor word, suffixed
difícilmente	Distractor word, suffixed
dudoso	Distractor word, suffixed
elegante	Distractor word, suffixed
embajador	Distractor word, suffixed
enfermedad	Distractor word, suffixed
enteramente	Distractor word, suffixed
especialista	Distractor word, suffixed
espectador	Distractor word, suffixed
espinazo	Distractor word, suffixed
final	Distractor word, suffixed
generoso	Distractor word, suffixed
global	Distractor word, suffixed
guerrero	Distractor word, suffixed
hermosa	Distractor word, suffixed
importante	Distractor word, suffixed
ingeniero	Distractor word, suffixed
interesante	Distractor word, suffixed
investigador	Distractor word, suffixed
latigazo	Distractor word, suffixed
lejano	Distractor word, suffixed
libremente	Distractor word, suffixed
ligero	Distractor word, suffixed
luminosa	Distractor word, suffixed
manita	Distractor word, suffixed
manotazo	Distractor word, suffixed
maravilloso	Distractor word, suffixed
mediante	Distractor word, suffixed
mismísimo	Distractor word, suffixed
monjita	Distractor word, suffixed
moral	Distractor word, suffixed
nacionalista	Distractor word, suffixed
natural	Distractor word, suffixed
normalmente	Distractor word, suffixed
pasito	Distractor word, suffixed
peligroso	Distractor word, suffixed
pelotazo	Distractor word, suffixed
pinchazo	Distractor word, suffixed

poquito	Distractor word, suffixed
progresista	Distractor word, suffixed
protagonista	Distractor word, suffixed
provisional	Distractor word, suffixed
puñetazo	Distractor word, suffixed
purísima	Distractor word, suffixed
realmente	Distractor word, suffixed
santísimo	Distractor word, suffixed
seguridad	Distractor word, suffixed
semejante	Distractor word, suffixed
sendero	Distractor word, suffixed
serenidad	Distractor word, suffixed
simplemente	Distractor word, suffixed
sociedad	Distractor word, suffixed
solamente	Distractor word, suffixed
sumarísimo	Distractor word, suffixed
superficial	Distractor word, suffixed
taxista	Distractor word, suffixed
terrorista	Distractor word, suffixed
total	Distractor word, suffixed
trabajador	Distractor word, suffixed
tradicional	Distractor word, suffixed
urbano	Distractor word, suffixed
variedad	Distractor word, suffixed
varita	Distractor word, suffixed
vencedor	Distractor word, suffixed
viajero	Distractor word, suffixed
visitante	Distractor word, suffixed
vistazo	Distractor word, suffixed

Table A-3: Distractor neologisms with nonce diphthongizing stems Note: each participant heard only one version of each neologism, with either a monophthong or a diphthong

Stem	Nonce Word with Monophthong	Nonce Word with Diphthong	Suffix
alsuere	alsoral	alsueral	al
aluetio	alotiero	aluetiero	ero
biemo	bemoso	biemoso	oso
biesio	besiero	biesiero	ero
buedo	bodito	buedito	ito
cansueta	cansotal	cansuetal	al
crielo	crelista	crielista	ista
dapuenso	daponsazo	dapuensazo	azo
duento	dontito	duentito	ito
fariena	farenero	farienero	ero

fuesgo	fosgal	fuesgal	al
guieba	guebal	guiebal	al
guiela	guelísima	guielísima	ísimo
humiebe	humbista	humiebista	ista
ipliente	iplental	ipliental	al
laguieno	laguenito	laguienito	ito
lueso	losazo	luesazo	azo
mabueso	mabosazo	mabuesazo	azo
mesliento	meslentero	meslentero	ero
mieja	mejista	miejista	ista
muerjo	morjero	muerjero	ero
nietro	netrito	nietrito	ito
nuedo	nodísimo	nuedísimo	ísimo
oruelma	orolmoso	oruelmoso	oso
paquiento	paquentito	paquentito	ito
peruesta	perostita	peruestita	ito
pinieno	pinenísimo	pinienísimo	ísimo
polienso	polensal	poliensal	al
priemo	premísimo	priemísimo	ísimo
pruense	pronsoso	pruensoso	oso
pueta	potísimo	puetísimo	ísimo
quelienda	quelendoso	queliendoso	oso
quierpo	querpista	quierpista	ista
quipieno	quipenal	quipienal	al
raluero	ralorista	raluerista	ista
ramuese	ramosazo	ramuesazo	azo
reguienta	reguentísima	reguentísima	ísimo
ricuesca	ricosquita	ricuesquita	ito
ruelto	roltazo	rueltazo	azo
setruela	setrolito	setruelito	ito
suedal	sodalista	suedalista	ista
suelba	solbal	suelbal	al
suerso	sorsero	suersero	ero
tiesna	tesnazo	tiesnazo	azo
timieso	timesazo	timiesazo	azo
tosueso	tososista	tosuesista	ista
tuebo	tobísimo	tuebísimo	ísimo
tueta	totero	tuetero	ero

Appendix B

Picture Naming Items

Spanish name	English name	Task version	Frequency Group	Cognate
batidora	mixer	1	low	no
bombero	fireman	1	low	no
buzón	mailbox	1	low	no
camisa	shirt	1	high	no
caña de pescar	fishing pole	1	low	no
carriola	stroller	1	low	no
carrito	wagon	1	low	no
castor	beaver	1	low	no
clip	paperclip	1	low	yes
corazón	heart	1	high	no
corbata	tie	1	high	no
dedal	thimble	1	low	no
dentista	dentist	1	low	yes
espárrago	asparagus	1	low	yes
faro	lighthouse	1	low	no
hoja	leaf	1	high	no
hombre	man	1	high	no
hueso	bone	1	high	no
iglesia	church	1	high	no
jarrón	vase	1	low	no
lavadora	washing machine	1	low	no
libro	book	1	high	no
mapa	map	1	high	yes
niña	girl	1	high	no
perro	dog	1	high	no
pez	fish	1	high	no
pie	foot	1	high	no
pierna	leg	1	high	no
plato	plate	1	high	yes
refrigerador	refrigerator	1	low	yes
reloj	clock	1	high	no
rey	king	1	high	no
rueda	wheel	1	high	no
sacapuntas	pencil sharpener	1	low	no
sandía	watermelon	1	low	no
techo	roof	1	high	no
tostador	toaster	1	low	yes

tractor	tractor	1	low	yes
tren	train	1	high	yes
trofeo	trophy	1	low	yes
árbol	tree	2	high	no
ballena	whale	2	low	no
bisagra	hinge	2	low	no
bolsa	bag	2	high	no
bomba	bomb	2	high	yes
caja	box	2	high	no
carretilla	wheelbarrow	2	low	no
carro	car	2	high	yes
cisne	swan	2	low	no
copa	glass	2	high	no
desarmador	screwdriver	2	low	no
escoba	broom	2	low	no
espejo	mirror	2	high	no
flor	flower	2	high	yes
huevo	egg	2	high	no
lámpara	lamp	2	high	yes
mano	hand	2	high	no
mesa	table	2	high	no
micrófono	microphone	2	low	yes
mujer	woman	2	high	no
pala	shovel	2	low	no
pan	bread	2	high	no
pañal	diaper	2	low	no
pinzas	tweezers	2	low	no
raqueta	tennis racket	2	low	yes
rastrillo	rake	2	low	no
recogedor	dustpan	2	low	no
regadera	watering can	2	low	no
reina	queen	2	high	no
silbato	whistle	2	low	no
silla	chair	2	high	no
silla de ruedas	wheelchair	2	low	no
sol	sun	2	high	no
teléfono	telephone	2	high	yes
telescopio	telescope	2	low	yes
tijeras	scissors	2	low	no
trompeta	trumpet	2	low	yes
velero	sailboat	2	low	no
ventana	window	2	high	no
vestido	dress	2	high	no

Appendix C

Korean word repetition task items

Word	Romanization
유의하다	yuũihada
자취하다	chach'wihada
이리저리	irijõri
쌀쌀하다	ssalssalhada
불가능	pulganũng
결혼하다	kyõlhonhada
되풀이	toep'uri
여러가지	yõrõgaji
유쾌한	yuk'waehan
공무원	kongmuwõn
박물관	pangmulgwan
뚜렷하다	tturyõt'ada
착하다	ch'ak'ada
만들다	mandũlda
시끄럽다	sikkũrõpta
특별히	t'ũkppyõlhi
백화점	paek'wajõm
슬프다	sũlpũda
늘씬하다	nũlssinhada
어느새	õnũsae
저절로	chõjõllo
이쑤시개	issusigae
휘발유	hwiballyu
생각나다	saenggangnada

태어나다	t'aeõnada
익숙하다	ikssuk'ada
확실한	hwakssilhan
운동화	undonghwa
안타깝다	ant'akkapta
밤새우다	pamsaeuda
낚시질	nakssijil
비슷하다	pisüt'ada
장난감	changnankkam
복덕방	pokttökppang
달려들다	tallyödũlda
예쁘다	yeppũda
거스름돈	kösürũmtton
공산주의	kongsanjuũi
농작물	nongjangmul
앞사귀	ipssagwi
나아지다	naajida
우체국	uch'eguk
쓰러지다	ssürõjida
관찰다	kwaench'ant'a
영수증	yõngsujũng
의미하다	ũimihada
어떻게	õttök'e
검푸르다	kõmp'urũda
끊임없다	kkũnimoptta
찌푸리다	tchip'urida

Appendix D

Language History Questionnaire Items

This questionnaire is designed to give us a better understanding of your experience with other languages.

We ask that you be as accurate and thorough as possible when answering the following questions.

You may elect not to answer certain questions if you do not want to.

Gender:

Age (years):

Do you have any known visual or hearing problems?

If yes, please explain:

Native Language(s):

Native Country:

Language(s) spoken at home:

Please record below all languages that you are able to use at least partially. Begin with your most proficient language.

For each language, record:

1. The age you first came into contact with it
2. In what context you learned it, (for example, in your family, in a classroom...)
3. How and if you currently use the language.

If English is your Native Language, please rate yourself on the following scale

1=no proficiency – 10=very fluent

**If English is NOT your Native Language, please contact the experimenter for further instructions*

Please rate your English reading proficiency:

Please rate your English writing proficiency:

Please rate your English speaking ability:

Please rate your English comprehension ability:

The next section of the questionnaire deals with your Spanish learning experience.

Have you studied Spanish? If NO, please go to the last question

Years in High School:

Semesters in College:

Have you studied/lived abroad in a Spanish-speaking country?

If YES, where and when did you study, for how long, and what language(s) did you speak?

The next section asks you to rate your skills in Spanish, using the following scale

1=no proficiency – 10=very fluent

Please rate your Spanish reading proficiency:

Please rate your Spanish writing proficiency:

Please rate your Spanish speaking ability:

Please rate your Spanish speech comprehension ability:

In my Spanish classes I generally get the following grades:

Are you (please check all that apply):

1. Taking Spanish for a requirement but interested in being a major or minor
2. Taking Spanish for a requirement, NOT interested in a major or minor
3. Taking Spanish because will be useful in my career
4. Taking Spanish because it is spoken so much in the US
5. Learning Spanish because it is important for my education
6. Learning Spanish because it is part of my heritage
7. A Spanish major
8. A Spanish minor
9. A graduate student in Spanish
10. Other

Please use the following scale to answer the next list of questions:

1=Strongly Disagree; 6=Strongly Agree

1. I can use appropriate greetings and farewells to begin and end a short conversation in Spanish.
2. I can understand enough of a conversation between two fluent speakers to summarize what was said.
3. I can understand enough of a 2-3 page nontechnical text to answer specific questions about it.
4. I could express myself well in a personal letter to a friend, although there might be some mistakes.
5. When I write a composition, I can generally find and correct my errors if I look carefully.
6. I can communicate well in conversations about many topics, although I might make some mistakes.
7. If I pay attention, I can speak with correct grammar about a range of topics.

Please use the following scale to answer the next list of questions:

1: Never or almost never true of me – 6: Always or almost always true of me

1. I'd like to sound as native as possible when speaking Spanish.
2. Acquiring proper pronunciation in Spanish is important to me.
3. I will never be able to speak Spanish with a good accent.
4. I believe I can improve my pronunciation skills in Spanish.
5. I believe more emphasis should be given to proper pronunciation in class.
6. I try to imitate Spanish speakers as much as possible.
7. Communicating is much more important than sounding like a native speaker of Spanish.
8. Good pronunciation skills in Spanish are as important as learning vocabulary and grammar.
9. I want to improve my accent when speaking Spanish.

Is there anything else about your language history that you would like to include here?

Appendix E

Reaction Time Figures in Milliseconds

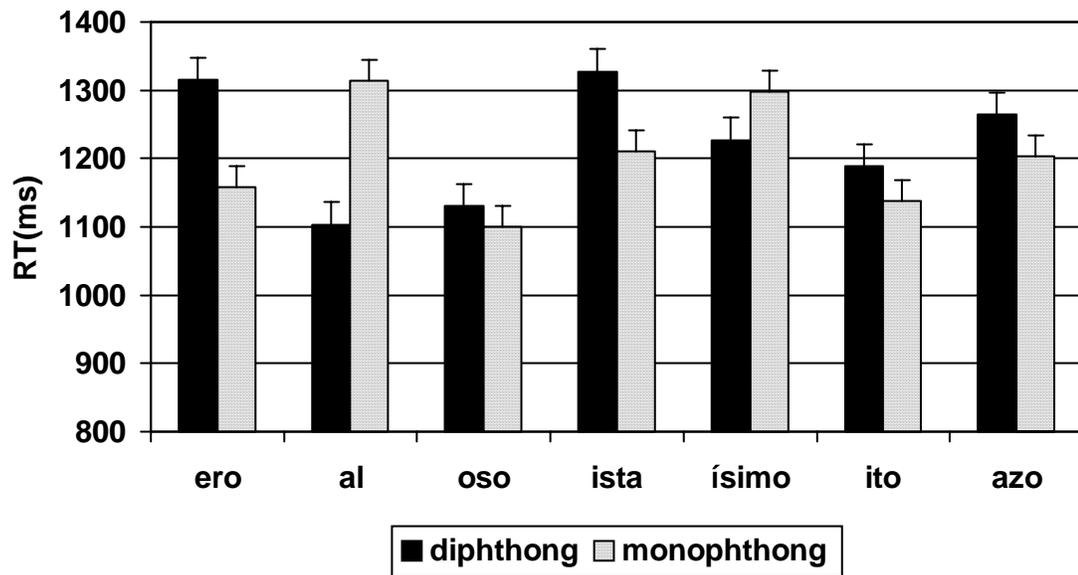


Figure E-1: RTs in milliseconds for neologisms with multisyllabic stems in LDT by native speakers, analysis by items (Figure 4-3)

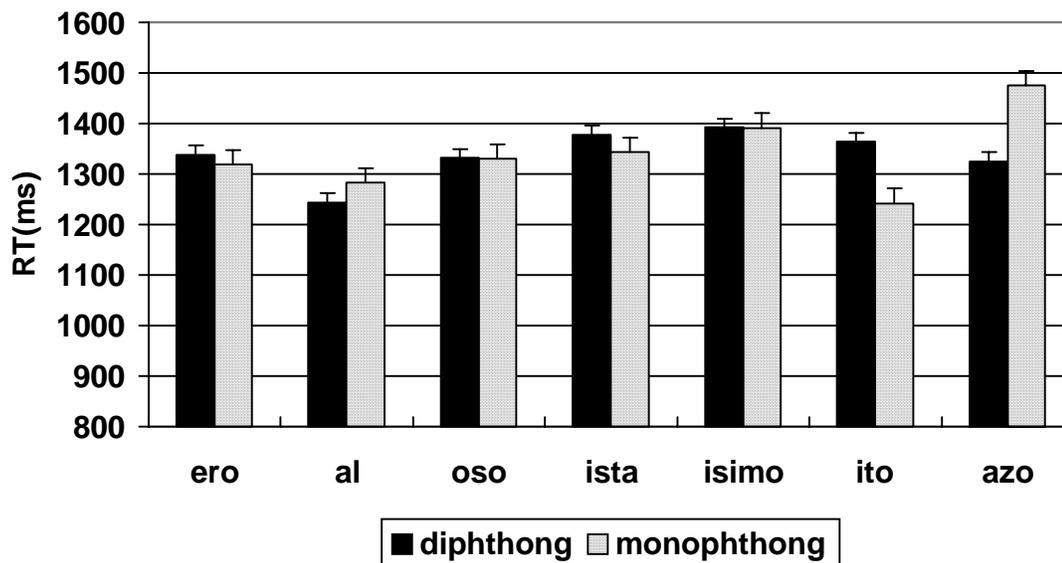


Figure E-2: RTs in milliseconds for neologisms in conditional naming by native speakers, analysis by items (Figure 4-5)

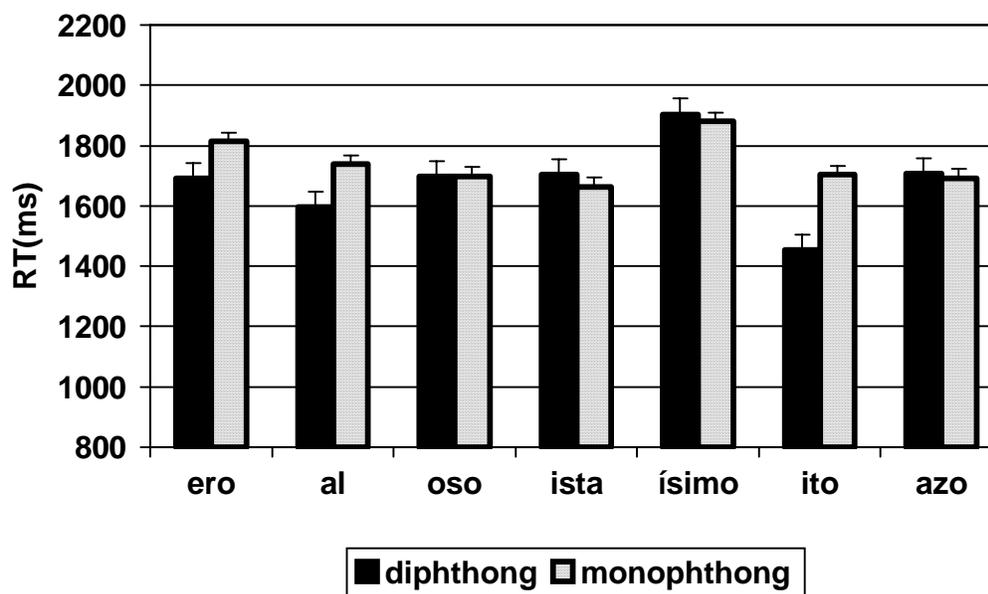


Figure E-3: RTs in milliseconds for neologisms in LDT by learners (Figure 5-1)

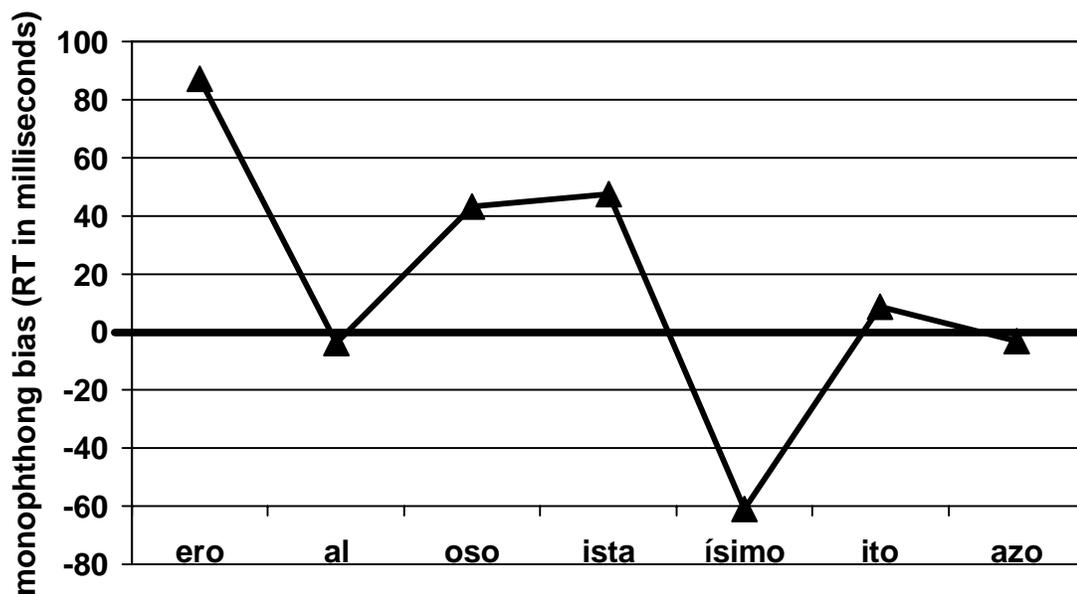


Figure E-4: Monophthong bias in (RT in milliseconds) for native speakers on the LDT, all neologisms (Figure 5-3)

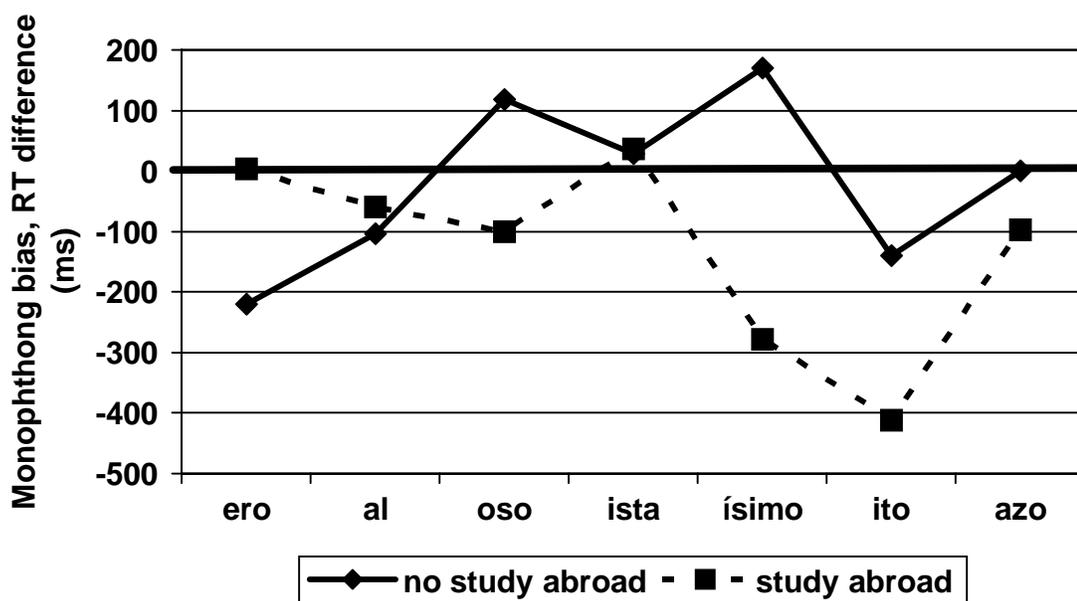


Figure E-5: Monophthong bias (RT in milliseconds) by suffix on the LDT for learners with and without study abroad experience (Figure 5-5)

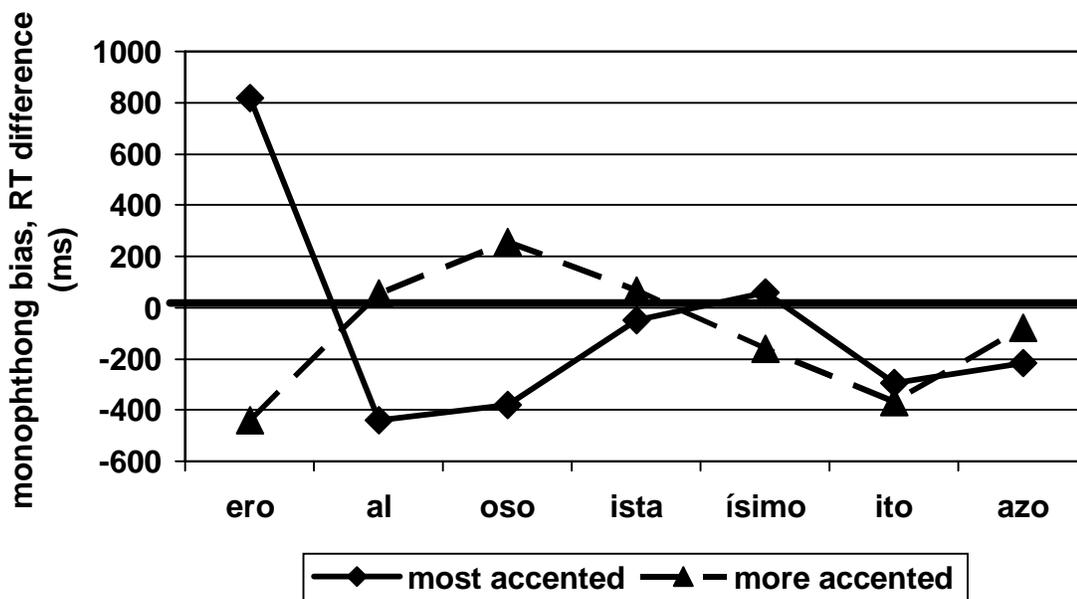


Figure E-6: Monophthong bias (RT in milliseconds) by suffix on the LDT for learners with greater accentedness in Korean word repetition (Figure 5-6)

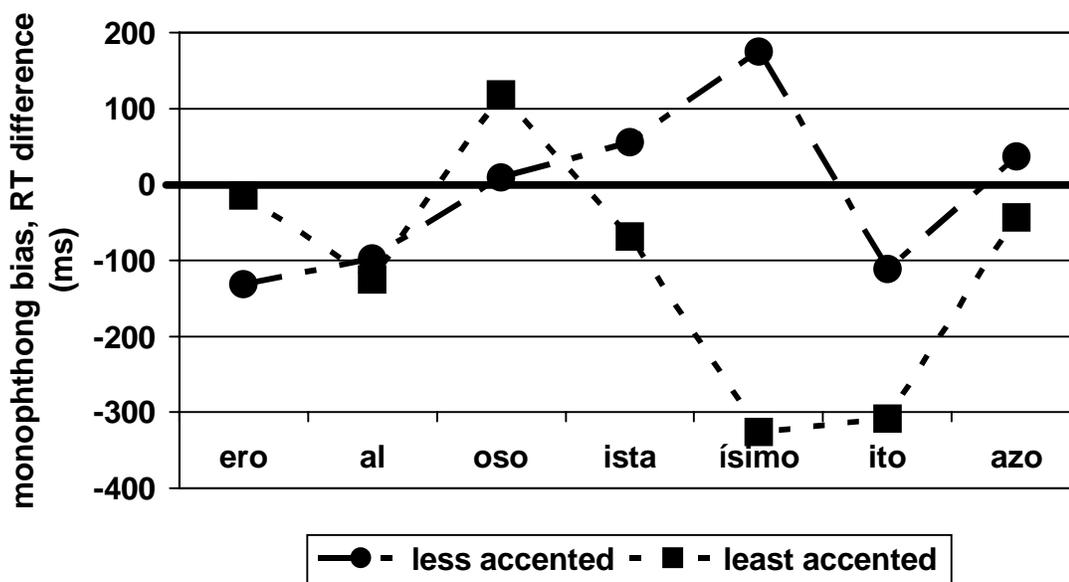


Figure E-7: Monophthong bias (RT in milliseconds) by suffix on the LDT for learners with less accentedness in Korean word repetition (Figure 5-7)

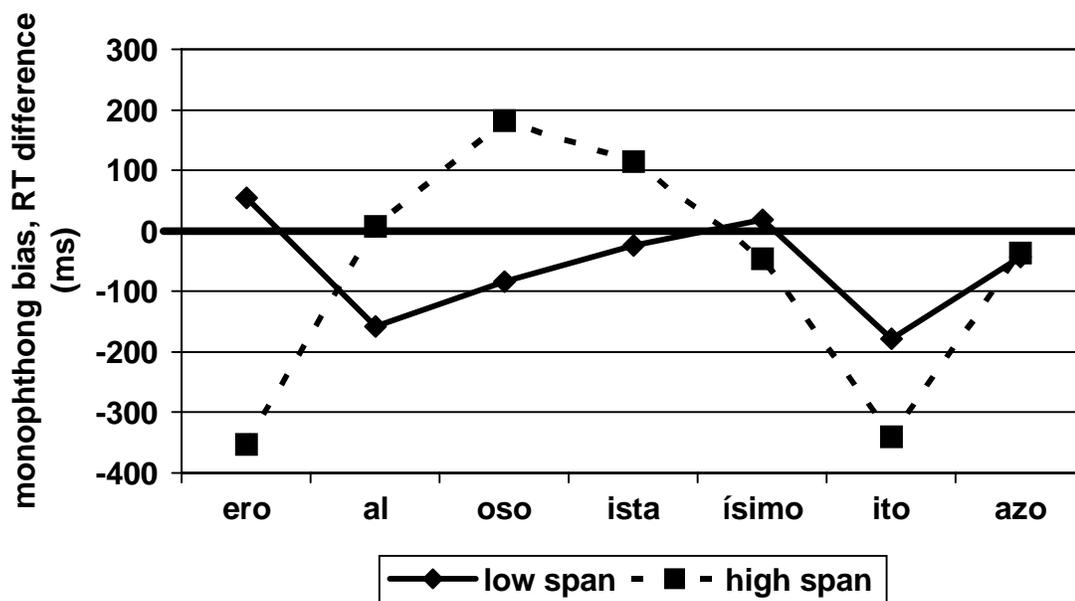


Figure E-8: Monophthong bias (RT in milliseconds) by suffix on the LDT for low and high memory span groups (Figure 5-8)

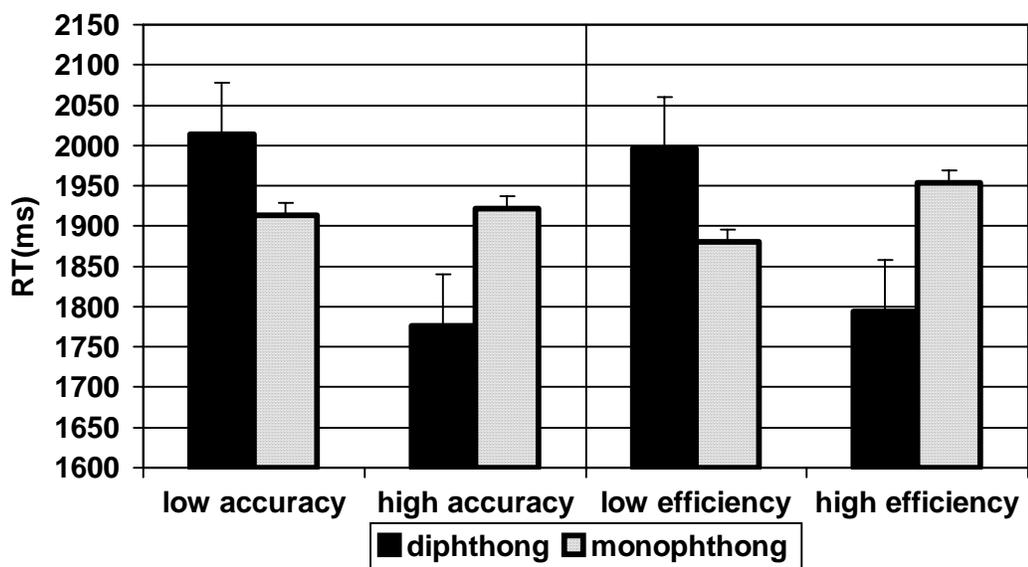


Figure E-9: RTs in milliseconds by diphthong status for neologisms in conditional naming by adult learners, by Spanish picture naming accuracy and efficiency (Figure 5-11)

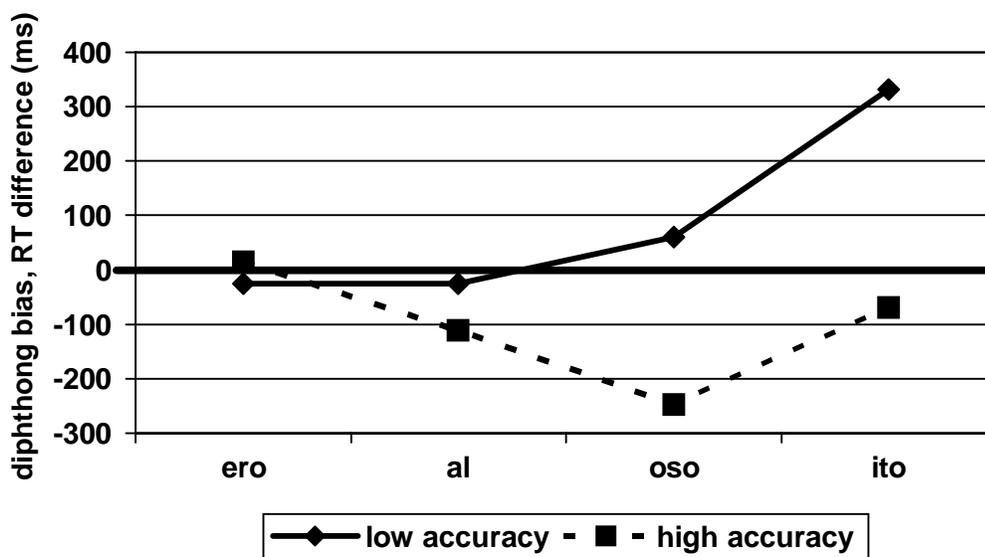


Figure E-10: Diphthong bias (RT in milliseconds) by suffix for learners on the conditional naming task with low and high accuracy in Spanish picture naming (Figure 5-12)

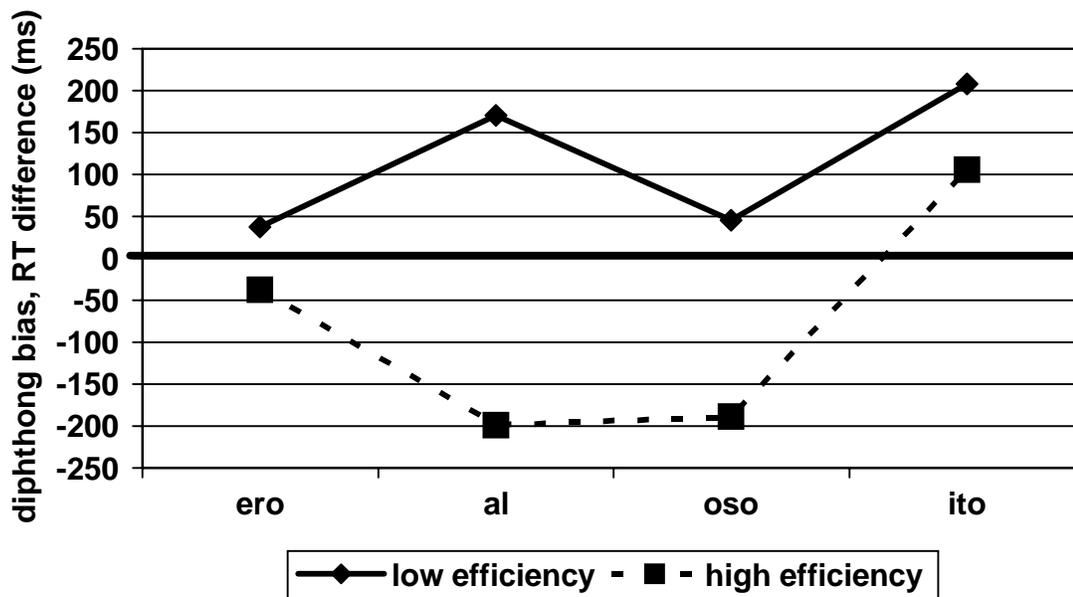


Figure E-11: Diphthong bias (RT in milliseconds) by suffix for learners with low and high efficiency in Spanish picture naming (Figure 5-13)

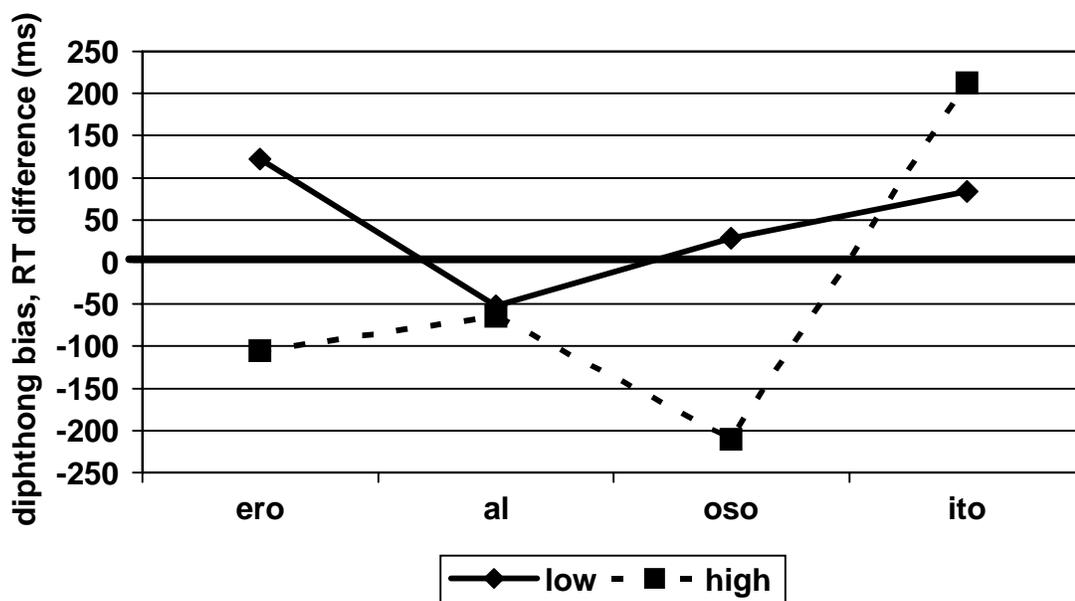


Figure E-12: Diphthong preference (RT in milliseconds) in conditional naming by adult learners as a function of accuracy in Korean word repetition (Figure 5-15)

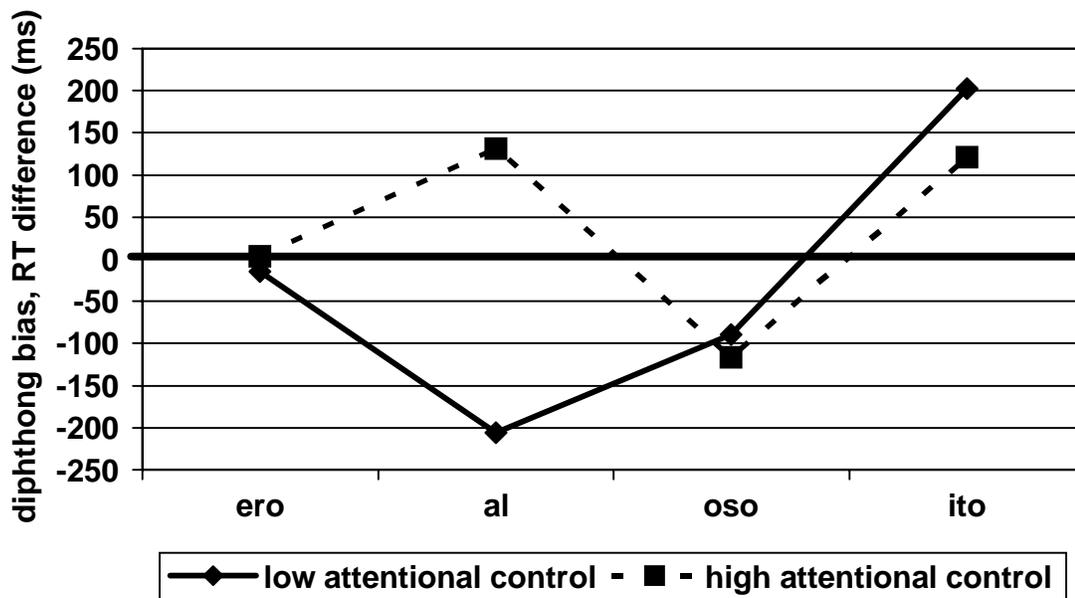


Figure E-13: Diphthong bias (RT in milliseconds) for neologisms in conditional naming for adult learners varying in attentional control ability (Figure 5-16)

VITA

Matthew Thomas Carlson

Ph.D. in Spanish, Applied Linguistics Option, Penn State University, 2007

MA in German, Middlebury College, 2005

MA in Spanish, Penn State University, 2003

NH teaching certificate, K-12 Spanish, Upper Valley Teacher Training Program, 1997

AB in Spanish and Music (high honors), Summa Cum Laude, Dartmouth College, 1996

GRANTS: National Science Foundation Dissertation Improvement Grant BCS-0544969, \$11,836: *The acquisition of probabilistic patterns in Spanish phonology by adult foreign language learners*, under the direction of Dr. Chip Gerfen, dissertation advisor

ARTICLES

Carlson, M. T. (2006). The development of fine-grained phonological knowledge in adult second language learners of Spanish. *The Southwest Journal of Linguistics*. 25(2).

Hall, J. K., Cheng, A., and Carlson, M. T. (2006). Reconceptualizing multicompetence as a theory of language knowledge. *Applied Linguistics* 27(2), 220-240.

Carlson, M. T. (2004). A critical look at the construction of power between Applied Linguistics and Critical Applied Linguistics. *International Journal of Applied Linguistics*. 14(2), 167-184.

Carlson, M. T. (2003). Preverbal elements in non-agentive Spanish predicates: Theta roles, case, and control. in P. Kempchinsky and C.-E. Piñeros (Eds.) *Theory, Practice, and Acquisition: Papers from the 6th Hispanic Linguistics Symposium and the 5th Conference on the Acquisition of Spanish and Portuguese* (pp. 121-139). Somerville, MA: Cascadilla Press.

PAPERS PRESENTED

Carlson, M.T. (2007, March). *The acquisition of probabilistic phonological patterns in adult learners of Spanish as a second language*. Poster presented at the Penn State University Graduate Exhibition, State College, PA.

Carlson, M.T. (2006, November). *El aprendizaje de patrones probabilísticos en la fonología española por anglohablantes adultos*. Invited talk at the University of Granada, Spain.

Carlson, M. T. (2005, March). *Sensitivity to probabilistic phonological patterns in the early stages of adult English Spanish bilingualism*. Paper presented at the Fifth International Symposium on Bilingualism, Barcelona, Spain.

Carlson, M. T. (2004, October). *The acquisition of fine-grained phonological knowledge in adult second language learners of Spanish*. Paper presented at the Second Language Research Forum, Penn State University.

Carlson, M. T., and Lantolf, J. P. (2003, March). *Mind making in the private writing of L2 learners*. Paper presented at the AAAL annual conference, Arlington, VA.

Carlson, M. T., and Lantolf, J. P. (2002, October-November). *The Mind on Paper: Private Writing in the Notebooks of Second Language Learners*. Paper presented at the 9th annual meeting of the Sociocultural and Second Language Learning Research Working Group, Florida State University.

Carlson, M. T. (2002, October). *Preverbal Elements in Non-agentive Spanish Predicates: Theta Roles, Case, and Control*. Paper presented at the 6th Hispanic Linguistics Symposium, University of Iowa.

AWARDS AND FELLOWSHIPS

2006 Jesús Díaz Prize for Spanish, Department of Spanish, Penn State

2005 Spector Award for outstanding graduate student in Applied Linguistics, Department of Linguistics and Applied Language Studies, Penn State

2005 Sparks Fellow, Department of Spanish, Italian, and Portuguese, Penn State

2001-3 University Graduate Fellow, Penn State

2001 Dartmouth General Fellowship

1996 Phi Beta Kappa