UNIVERSITY OF CALIFORNIA
Los Angeles

Grammar and Analogy in Phonotactic Well-formedness Judgments

A dissertation submitted
in partial satisfaction of the requirements
for the degree Doctor of Philosophy in Linguistics

By

Shabnam Shademan

2007

The dissertation of Shabnam Shademan is approved.

Donka Minkova Stockwell

Carson T. Schütze

Colin C. Wilson

Susan R. Curtiss, Committee Co-chair

Bruce P. Hayes, Committee Co-chair

University of California, Los Angeles

2007

TABLE OF CONTENTS

Table of Figures

Table of Tables

ACKNOWLEDGEMENTS

I have been dreaming of arriving at the point of writing the acknowledgments for this document because I knew that it would mean that the most difficult parts were behind me. I have arrived at this point with the realization that this project could not have started, nor could it have finished, without the guidance and encouragement of a number of people. This work has been truly collaborative and I am humbled that I am allowed to put my name on it. All errors and flaws are, of course, mine.

I have had the fortune of having a group of gifted and generous professors on my doctoral committee. I am grateful to Donka Minkova for joining this project with a big smile and many encouraging words. Carson Schütze, who was my very first linguistics teacher at UCLA, has been a mentor in critical thinking and hard work. Colin Wilson has been the strong pillar on which I have leaned every time I ran out of breath. He could spot every hole in any theory before I had time to completely absorb the theory itself. I could not imagine embarking on this project without his guidance.

Susie Curtiss has been a role model in courage and kindness. She has given me tremendous guidance in stating my thoughts and assumptions clearly. She has also taught me lessons that I would have never learned in any school.

It is hard to describe what Bruce Hayes has meant to me. Not only does he have a gift for constructing your argument better than you ever could, but he also has the generosity of making you feel as if you knew the answers the whole time. And he can effortlessly do all that, even when you appear at his office door with a shaved head. I could not have asked for a better supporter, and will never be able to thank him enough.

Donca Steriade introduced me to phonology, and her passion was contagious. My educational experience has significantly changed for the better because I have had educators and advisors such as Nina Hyams, Pat Keating, Sun-Ah Jun, Anoop Mahajan, Pam Munro, Tim Stowell, and Daniel Buring. I also owe many thanks to Mike Hammond (University of Arizona) for generously sharing a large corpus (used in this project) with me, to Adam Albright (MIT) for always helping me with various technical problems with a smile and lots of encouragement, and to Kie Zuraw for her endless willingness to hear my half-baked ideas.

I was introduced to Generative Linguistics by Bruce Rowe at Los Angeles Pierce College (Anthropology), where I received a magnificent education by having instructors such as Scarlett Gani (Modern Languages) and Gail Hobbs (Geography). It is unfortunate that I do not remember the name of my English teacher at LAPCC, who had the patience to teach me, for the first time in my life, how to write a book review. Her lessons have been invaluable.

The development and finalization of this project has also been helped by every other graduate student in the Linguistics Department at UCLA who has helped this place to be one in which new ideas are nourished, work ethics and patience are cherished, and friendships flourish. In particular, I would like to single out (in no particular order) Kuniko Y. Nielsen for calling me her "soul sister" and staying true to her words, Brook D. Lillehaughen for having the touch for making me feel better every single time, Jeffrey Heinz for injecting linguistic enthusiasm into everyone with whom he comes in contact, Julia Berger for making me smile by keeping me guessing, Sarah

VanWagenen for never making me feel stupid when I asked the stupidest questions about experimental design, Greg M. Kobele for being the drinking buddy who is willing to discuss linguistics any time and anywhere, and Sahyang Kim for her open mind, deep heart, and humble soul.

I have made lifelong friends who happen to be great linguists who have, emotionally and/or intellectually, supported me at some point of this journey– (in no particular order) Ivano Caponigro, Manola Salustri, Gianluca Storto, Harold Torrance, Leston Buell (for good reasons aka bulbul), Katya Pertsova, Jason Riggle, Jason Kandybowicz, Jie Zhang, Heriberto Avelino, Rebecca Scarborough, Molly Shilman, Christina Kim, Asia Furmanska, Christina Esposito, Alex MacBride, Heidi MacBride, Jill Gilkerson, Kristine Yu, and Victoria Thatte.

The presence of a few non-linguists has been crucial in completing this project. I am grateful to Haleh and Shahram for making sure that I got enough warm homemade Persian food and play time, to Karin for tooting my horn every chance she got, and to Nancy for her relentless support. Mama Kitty and Frizbee kept me company on long writing days by lounging all over my papers, books, and keyboard – which reminds me: all typos are their fault.

My parents did more than instill the love of reading in me. My father taught me how to own a free heart, and my mother taught me about beauty. I hope their lessons are occasionally reflected in who I am today. My sister, Naz, will never know how much her belief in my character and in my ability to do anything has propelled me forward.

While everyone who is acknowledged in this thesis deserves it, no one's contribution would have been even possible if I did not have David Ross as a partner. He has read more drafts of linguistics squibs, abstracts, posters, PowerPoint presentations, papers, and theses than any non-linguist should be subjected to; he has judged the well-formedness of enough novel words to be able to write a book in nonce; he has made sure that during crunch time I had a clean house to come to, warm food on the table to sit at, clean sheets to sleep in, and loving arms to rest in. His presence has changed my life.

This research was partially funded by the UCLA Department of Linguistics (2000-07), a Dissertation Year Fellowship from UCLA (2005-06), a Research Mentorship Fellowship from UCLA (2002), and Regent Graduate Quality Education (2003). In addition, I would like to thank my neighbors on the peninsula for participating in the experiments, and Mrs. Portia Rose for sharing her phone book.

VITA

Born	Tehran, Iran
1986-89	Daneshgah-e-Azad, Tehran, Iran Department of Electrical Engineering
2000	B.A., Linguistics, *summa cum laude*, *departmental highest honors* University of California, Los Angeles
2001-02	Research Mentorship Fellowship University of California, Los Angeles
2002	M.A., Linguistics University of California, Los Angeles
2003	Visiting Student, Department of Linguistics and Philosophy Massachusetts Institute of Technology, Boston
2004-07	Teaching Associate, Fellow University of California, Los Angeles
2005-06	Dissertation Year Fellowship, UCLA

PUBLICATIONS AND PRESENTATIONS

Shademan, Shabnam. 2003. Epenthetic Vowel Harmony in Farsi. *Talk presented at the 22nd West Cost Conference on Formal Linguistics*, University of California, San Diego, May 21-23.

___. 2004. Glottal Deletion and Compensatory Lengthening in Farsi – a Phonetic Study. *Poster presented at the 147th meeting of the Acoustical Society of America*, New York, NY, May 24-28.

___. 2005. Glottal Deletion and Compensatory Lengthening in Farsi – a Phonetic Study. *UCLA Working Papers in Phonetics*, 104: 61-81.

___. 2006. Analogy vs. Grammar in English Phonotactics. *Talk presented at the 80th meeting of Linguistic Society of America*, 5-8 January.

___. 2006. Is Phonotactic Knowledge Grammatical Knowledge?" *Proceedings of the 25th West Coast Conference on Formal Linguistics*, 371-379.

ABSTRACT OF THE DISSERTATION

Grammar and Analogy in Phonotactic Well-formedness Judgments

By

Shabnam Shademan

Doctor of Philosophy in Linguistics
University of California, Los Angeles, 2007
Professor Bruce Hayes, Co-chair
Professor Susan Curtiss, Co-chair

Speakers of a language are able to make judgments of *phonotactic well-formedness* in their language. For instance, English speakers typically judge that a form like *blick* is phonologically acceptable, whereas a form like *bnick* is phonotactically deviant. This dissertation reports an experimental study of phonotactic well-formedness judgments, intended to shed light on their origin. The central issue is whether such judgments arise from a formal *grammar*, of the kind constructed by linguists, or from *analogy*, that is, resemblance to existing lexical items. The key research method was to use formal (implemented) models intended to embody in a precise way the concepts of grammar and analogy. Such models made it possible to select experimental stimuli that varied independently along the two dimensions of grammatical probability and lexical analogy.

The results indicate that acceptability ratings are the result of the combination of both mechanisms, grammatical and analogical, acting together. However, the grammatical mechanism appears to be stronger than the analogical one. Further examination of the results indicates that the two mechanisms do not interact additively; rather, analogy has the most influence on forms for which the grammatical mechanism's predictions are not extreme.

Two further experiments suggest that under varying circumstances, the balance between rules and analogy can change. An experiment on older adults (over 65) demonstrates that an increase in age results in an increase in the effect of lexical analogy, an effect attributed here to their greater life experience, and hence acquaintance with their lexicons. Another experiment tested the effect of including real words in the stimulus set, under the view that the inclusion of real words would reinforce the analogical mechanism. The results of this experiment were suggestive but inconclusive. The dissertation concludes by suggesting a number of possibilities for further investigation.

Chapter 1: Introduction

Our linguistic knowledge comprises two distinct components: 1) a mental lexicon for storing words along with their various (e.g., phonological, morphological) information[1], and 2) a mental grammar that contains properties and rules that underlie the composition and combination of lexical items, and hence the generation of an infinite number of larger units (e.g., Chomsky, 1965, 1995; Pinker, 1994). There is considerable evidence that these two components are distinct in their computational (e.g., Fodor, 1983; Marslen-Wilson and Tyler, 1998) and neural bases (e.g., Kempler, Curtiss and Jackson, 1987; Damasio and Damasio, 1992; Ullman et at., 1997; Ullman, 2001), and work in parallel in the construction and calculation of form-meaning relations. Nevertheless, the issue of whether the lexical and grammatical components of language are independent from each other is a controversial one. Generative theorists assume the independence of these components (e.g., Chomsky, 1965, 1995; Pinker, 1994, 1999, among others), while other researchers do not necessarily hold this view, and suggest general cognitive mechanisms at work (e.g., MacDonald et al., 1994; Elman et al., 1996; Bates and Goodman, 1997, among others).

[1] Current linguistic theories place syntactic information into every lexical entry. In this thesis, when I refer to lexicon or lexical entries, I am referring to only phonological information stored within lexical entries and not other kinds of grammatical information (e.g., semantic or syntactic).

This dissertation project tests the independence of a grammatical mechanism and an analogical mechanism by investigating the mechanisms that potentially underlie our phonotactic well-formedness judgments. In doing so, I will provide data to bear on the extent to which analogy and grammar contribute and interact in such judgments, as well as the manner by which they may change over the course of aging. This goal will be accomplished by using phonotactic well-formedness judgment tasks. In these tasks, a participant is asked to rate the acceptability of a novel form as a word of their language.

Phonotactic well-formedness judgments obtained from speakers have two well-established characteristics: they are consistent across subjects and they are gradient (Greenberg and Jenkins, 1964; Scholes, 1966; Ohala and Ohala, 1986; Schütze, 1996; Coleman and Pierrehumbert, 1997; Vitevitch et al., 1997; Frisch, Large and Pisoni, 2000; Treiman et al., 2000; Bailey and Hahn, 2001; Albright and Hayes, 2003; Hay, Pierrehumbert and Beckman, 2003). For example, English speakers consistently reject *bnick* as a potential word, while they accept *blick* as one (Chomsky and Halle, 1965). Moreover, speakers often judge *bwick* to be not as acceptable as *blick*, but more acceptable than *bnick* (Albright, 2006). This indicates that speakers make gradient distinctions within a set of possible/impossible words.

What remains unclear is the source of these well-formedness judgments. One view is that speakers calculate the acceptability ratings for a novel form based on how similar to real words it sounds. In this view, acceptability judgments are the result of comparing a form to real words in the lexicon, i.e., "analogy." For example, *blick* is judged as an acceptable English word by virtue of its similarity to real words, such as *black* or *lick*. Therefore, acceptability of a novel form is measured based on existing lexical items because there is an analogical mechanism that compares a novel form to the real words in the lexicon in order to determine its well-formedness.

Another account of how novel forms are accepted or rejected is based on whether they are legal or illegal according to the grammatical rules of the language. This means that speakers render their acceptability judgments using a rule-based system. Therefore, well-formedness judgment tasks engage the rules of the phonotactic grammar, and *bnick* is judged as unacceptable because /bn/ is an ungrammatical onset.

Finally, there is the possibility that the two mechanisms both participate in the process of determining well-formedness of novel forms. In this alternative view, grammar and analogy work together to determine acceptability judgments. In order to distinguish among these views, we must be explicit about each mechanism's function and its expected effects.

In this thesis, the term "grammar" refers specifically to a probabilistic phonotactic grammar. Such a phonotactic grammar is defined as a formal object that assigns a real number to phonological strings, claimed to correspond to native speaker judgments of well-formedness. This dissertation will use two phonotactic grammars, primarily the one proposed by Coleman and Pierrehumbert (1997), and supplemented by one based on the maxent model of Hayes and Wilson (forthcoming). These models are discussed in section 2.2.2.

A grammar is learned by combing through the data of the lexicon, but in applying the rules of the grammar, the lexical data used in learning the grammar are not assessed.

Thus, it is logically possible that a grammar could remain stable through adult life, despite the learning of novel lexical items.

The stochastic phonotactic grammars of Hayes and Wilson are straightforward instances of their type, made up of constraints that are based on distinctive features and natural classes. However, the claim that the Coleman and Pierrehumbert's system is a grammar is more controversial, since (as we will see in Chapters 2 and 3) these "grammars" are formed simply by counting syllable onsets and rimes in a variety of contexts. My defense for the use of Coleman and Pierrehumbert's model as a "grammatical" model is that it yields results quite similar to those obtained from the uncontroversially grammatical Hayes and Wilson model.[2]

In sum, the requirements for a formal grammatical model to be used in this thesis were limited to the following features: 1) it embodied in some intuitive sense the idea of a phonological grammar with distributional knowledge, 2) it allowed for gradient predictions, since phonotactic judgments are gradient, and 3) it was explicitly stated, to the point of being machine implemented. As a result, while the model of Coleman and Pierrehumbert (1997) lacks elaborate grammatical knowledge and limits its application of phonological structure to onsets and rimes, it was employed for the present study because: 1) it uses, at least, some aspects of grammatical structure, 2) its properties are explicitly stated, and 3) it can be implemented.

The alternative Hayes and Wilson model, discussed in section 2.2.2, uses more elaborate phonological apparatus. However, this model was not available when the

---

[2] The two model's scores are highly correlated (r^2=.66)

studies in this thesis were designed; therefore, it was not used in selecting the stimuli. However, its performance for the current data will be reported in Chapters 4, 5, and 6.

In addition, I use the term "analogy" to refer to a theory that says that words should be phonotactically well-formed to the extent that they resemble the other words in the lexicon. Unlike a grammar, an analogical model is dependent on the actual presence of the lexicon for purposes of computing these resemblances. In principle, the predictions of an analogical model would be expected to change as the speaker's vocabulary evolves – a prospect to be examined in detail below.

While the resemblance of words to each other is determined by comparing a word to every other word in the lexicon to determine the degree of similarities between them, different analogical models make different comparisons in their computation of similarity. I will describe a quantified measure of this comparison based on a formal model (section 3.2) which allowed me to compare the analogical strength of different novel forms in the study.

This dissertation will investigate the presence of each mechanism (grammatical and analogical) by using stimuli that are designed in a particular way: they are either of the type for which the two mechanisms' predictions are in the same direction (e.g., they both reject the form as unacceptable), or of the type for which the two mechanisms' predictions are in the opposite direction (e.g., the analogical mechanism predicts the form to be not highly acceptable, while the grammatical mechanism predicts the form to be highly acceptable).

By constructing the stimuli such that the two dimensions (analogy and grammar) vary independently from each other, I will monitor the independent effects of the two mechanisms. The stimuli with opposite influences from grammar and analogy will be the most informative, in that they will provide an opportunity to examine the strength of each factor. For example, if analogy is, indeed, an influential factor in the process of determining the phonotactic well-formedness of a form, we expect that when the grammatical mechanism predicts forms to be not highly acceptable, the effect of analogy will be more robust for the acceptability ratings.

In addition, this thesis will investigate whether the two mechanisms could be affected independently by aging. This investigation will start by examining the acceptability ratings given by younger adults. Next, I will compare the ratings from younger adults and older adults in order to determine whether the strengths of the two effects (grammar and analogy) remain invariant in healthy aging. If either mechanism's effect is altered, I will determine whether the strength of that effect has decreased or increased with age. Since older adults have larger lexicons and more experience with the words they know, my prediction is that the effect of analogy increases with age, while the effect of grammatical mechanism is not altered by aging and will not show any change in the size of its effect.

The experiments reported in this dissertation are phonotactic well-formedness judgment tasks, and will test three hypotheses. Hypothesis 1: I predict that phonotactic well-formedness judgments are a reflection of two mechanisms: grammar and analogy. Hypothesis 2: In healthy aging, the effect of grammatical probability will remain invariant, while the effect of lexical analogy will be strengthened. Hypothesis 3: The inclusion of real words in the stimulus set will result in an increased effect of the analogical mechanism.

Three experiments will be reported. Experiment 1 is designed to test hypothesis 1; that is, to show that there are, indeed, two mechanisms at work in predicting acceptability ratings. Experiment 2 is designed to test hypothesis 2; that is, to show that the effect of lexical analogy is strengthened in older adults. Experiment 3 is designed to test hypothesis 3; that is, to test whether the strength of lexical analogy is strengthened when real words are included in the list of stimuli.

This thesis is organized as follows.

Chapter 2 provides the theoretical background and motivation for this research. It focuses on the scientific questions raised and the motivations for the three hypotheses.

Chapter 3 describes the selection of the stimuli and experimental methods that will address the proposed hypotheses. It starts by defining the formal models of grammatical probability and lexical analogy that were used in the present study, followed by the method of stimulus selection. It provides an outline of the stimulus conditions used in the three phonotactic well-formedness judgment experiments.

Chapter 4 describes Experiment 1, in which the effects of grammar and analogy in well-formedness judgments were investigated. It reports on the results of the experiment and concludes that both the grammatical mechanism and the analogical one influence well-formedness judgments.

Chapter 5 describes Experiment 2, in which older adults participated. This chapter reports on the results of the experiment, which tested the second hypothesis. The results show that only the strength of one mechanism, analogy, is affected in this paradigm in healthy aging.

Chapter 6 describes Experiment 3, which was similar to Experiment 1, except that real words were removed from the stimulus set. It reports the results of the experiment and suggests that the results are inconclusive and require further investigation.

Chapter 7 reports on a post-hoc analysis of the interaction between the grammatical and analogical mechanisms. It proposes that the effect of analogy is least influential for stimuli that are at the extreme edges of grammatical probability (i.e., highly probable or highly improbable), and that analogy is most influential in the mid-range of grammatical probability.

Chapter 8 summarizes the results presented in this thesis. It also provides suggestions for the direction of future work in the area of phonotactic knowledge.

Chapter 2: Research Plan and Literature Review

Knowledge of phonotactic restrictions refers to speakers' knowledge of the possibility for a given sequence of sounds to constitute a word in their language. There are, at least, two intriguing aspects of this knowledge that any theory or model of phonotactics has to be able to explain. First, acceptability ratings of novel forms are consistent across subjects (e.g., Greenberg and Jenkins, 1964; Ohala and Ohala, 1986; Vitevitch, Luce, Charles-Luce and Kemmerer, 1997; Coetzee and Pater, 2005). For example, any native speaker of English is able to distinguish between a possible English word, such as *blick*, and an impossible word, such as *bnick* (Chomsky and Halle, 1965). Second, the well-formedness judgments obtained from speakers have gradient patterns (e.g., Coleman and Pierrehumbert, 1997; Vitevitch et al., 1997; Frisch, Large and Pisoni, 2000; Treiman et al., 2000; Bailey and Hahn, 2001; Albright and Hayes, 2003; Hay, Pierrehumbert and Beckman, 2003; Hammond, 2004). This means that speakers distinguish within a set of possible/impossible words. For example, the novel form *bwick* is rated as more acceptable than *bnick*, even though it is judged as less well-formed than *blick* (Albright, 2006). While the characterizations of well-formedness judgments are clear and well-established, the mechanisms by which speakers arrive at their judgments are not well understood.

2.1 Research Plan

This dissertation investigated the mechanisms that underlie phonotactic well-formedness judgments. In my first hypothesis, I predicted that acceptability ratings would be a reflection of the combination of two mechanisms: a grammatical one and an analogical one. A careful examination of this issue had two fundamental requirements. First, I had to make explicit my assumptions about each mechanism's function and its expected effects. The second requirement of the study was to use stimuli that allowed for independent monitoring of the effects of the two mechanisms. To this end, I selected a set of stimuli in which items could be distinguished according to the predictions made by the analogical and grammatical mechanisms based on the two formal models. Previous studies on phonotactic well-formedness had been inconclusive because of their methods of stimulus selection (section 2.2), where both mechanisms (analogy and grammar) made similar predictions about the stimuli's well-formedness. Therefore, one could not disassociate the effects of each mechanism.

Hypothesis 1 predicted that both mechanisms (grammar and analogy) would be shown to be effective in phonotactic well-formedness judgments. In addition, I predicted that by using stimuli that covered the full range of variation with respect to these mechanisms, we would observe that grammatical probability had a greater influence on well-formedness judgments. In order to test these hypotheses, Experiment 1 was carried out. The results of this experiment are reported in Chapter 4.

Another goal of this thesis was to investigate whether aging affected these mechanisms differently. My prediction was that the analogical mechanism would be affected, but not the grammatical one. There have been claims that language changes observed in older adults are not particular to language processing. Instead, these changes are the result of a general slow-down in these populations (section 2.4). If changes observed in older adults are the result of a general cognitive slowing, then there would be no reason to expect aging to affect analogy any more than grammar, or vice versa. However, one could argue that the continued exposure to and use of words that comes with age, and perhaps a continually growing lexicon, would impact analogy more than grammar.

Hypothesis 2, therefore, predicted that the effect of the analogical mechanism in well-formedness judgments would increase with age, because, as a result of increased language use, there would be an advantage for the analogical mechanism over the grammatical mechanism in making well-formedness judgments. The expectation was that in comparing the acceptability ratings from younger and older adults, one would find that analogical strength had an increased predictive power for the older population's acceptability ratings. In order to test this hypothesis, Experiment 2 was carried out. Only healthy adults over the age of 65 participated in Experiment 2, which was otherwise the same as Experiment 1. The details and results of this experiment are presented in Chapter 5.

In addition to aiming to distinguish the two active mechanisms in well-formedness judgments, I also wanted to determine whether the linguistic processing involved in well-formedness judgment tasks would be sensitive to the inclusion of real words in the stimulus set. It has been reported that the experimental material (e.g., words and non-words mixed in the same block vs. presented in separate blocks) and/or the experimental tasks (e.g., lexical decision tasks vs. repetition tasks) may influence the degree to which analogy and grammar affect linguistic processing (sections 2.3.1.1 and 2.3.2.2). Prior studies have varied with respect to the reported influential mechanisms, as well as whether or not they included real words as part of their stimulus set. If language processing were sensitive to the experimental environment (including the stimuli and the task), it seemed reasonable to ask whether the inclusion of real words in well-formedness judgment tasks had had an impact on the observed patterns reported in prior studies.

Thus, in hypothesis 3, I predicted that with the inclusion of real words in the stimuli, participants were simultaneously involved in two tasks: the original task of judging acceptability of stimuli, as well as an inadvertently added task, that of discriminating words from non words (i.e., a lexical decision task). Thus, a third goal of this thesis was to determine whether the removal of the lexical decision task from the well-formedness judgment task would result in a decrease in the effect of the analogical mechanism. To test this hypothesis, Experiment 3 was carried out, and the ratings were compared to those collected in Experiment 1. The crucial manipulation was that in Experiment 3 the stimulus set consisted of novel forms only, while the stimulus set in Experiment 1 consisted of the same novel forms together with real forms. Experiment 3 and its results are discussed in Chapter 6.

2.2 Experimental Literature on Phonotactic Well-formedness Judgments

The main goal of this thesis is to contribute to a better understanding of the mechanisms engaged in speakers' phonotactic well-formedness judgments. There are three fundamental approaches to the issue. One view is that acceptability ratings of a novel form are the result of its degree of similarity to existing words in the lexicon (i.e., "analogy"). A second view is that novel forms are accepted or rejected based on the rules of the phonotactic grammar. Finally, a third view is that the two mechanisms co-contribute to phonotactic well-formedness judgments. In what follows, I review selected experimental reports on acceptability ratings from speakers.

2.2.1 Analogy with Existing Words

There is evidence that some aspects of well-formedness judgments may be associated with the degree to which novel forms are similar to words in the lexicon (e.g., Greenberg and Jenkins, 1964; Ohala and Ohala, 1986; Frisch, Large and Pisoni, 2000; Bailey and Hahn, 2001). This means that a given non-word is compared to existing words in the lexicon, and the acceptability judgment for that non-word is based on how close to a real word the non-word is, as well as its phonological neighborhood density (i.e., how many real words it is closely related to phonologically). For example, a non-word like *blick* has a high rating of acceptability by English speakers, not only because it is similar to actual words (e.g., it differs in only one phoneme from *black* and *lick*), but also because there are many such words in its phonological neighborhood, such as *black*, *bleak*, *block*, *bloke*, *blink*, *blip*, *bliss*, *brick*, *lick*, *click*, *flick*, and *slick*.

On this view, the mechanism by which a word's acceptability is assessed is by directly analogizing the novel form to existing forms in the lexicon. "Analogical strength" in this context depends on the relation between the strings being compared. However, the exact nature of this comparison is not well defined and varies greatly across studies. The reason for this uncertainty is that the factors that influence similarity between two forms can be one or more of the following: 1) the number of words that are phonologically similar to the target (i.e., the target's neighbors), 2) the degree of phonetic similarity between the target and its neighbors, and 3) the frequencies of occurrence (word frequency) of the neighbors.

In computing analogical strength, researchers have generally used the number of neighbors as the basis of their computation, with varying degrees of influence allowed for the other two factors. For example, Greenberg and Jenkins (1964), who conducted one of the early studies that attempted to answer the question of what contributes to speakers' phonotactic intuitions, defined two forms as phonological neighbors if one representation could be turned into another by means of substituting one or more phonemes. This means that the number of neighbors was calculated based on how many phoneme substitutions in various positions would result in an existing word.

Greenberg & Jenkins reported experiments in which they used CCVC stimuli that consisted of real words and non-words. The stimuli were assigned an analogy score based on how distant a novel form was from a real word. This distance was computed based on substitution (i.e., no deletion or addition) of phonemes in the following way.

Any number of phoneme substitutions in any position was allowed. They calculated how many of the substitutions resulted in an existing word. For example, substituting phonemes in a form like /**klæb**/ could result in *slab*, *crab*, *club*, *clam*, *stab*, *crib*, *grub*, and so on. Therefore, /**klæb**/ would have 15 various phoneme substitutions that would result in an existing word. Since all of the study's stimuli were four phonemes long, substituting any phoneme in any position, in order to determine whether each substitution would result in a real word, allowed for the possibility of 16 phoneme substitutions (from substituting no phoneme to substituting all phonemes).

The analogical score for each stimulus was derived by deducting the number of possible substitutions from 17, which avoided zero for real words that had neighbors at every substitution distance (i.e., words that would have 16 substitutions). Therefore, a derived low score meant that the novel form was close to many existing forms. For example, /**klæb**/ would be given a similarity score of 2, because there are 15 various phoneme substitutions for that form that would result in an existing word.

Participants were told that the stimuli were from other languages and that their task was to provide a rating of "how far" the target seemed "from being an English word." The results showed a high correlation between the ratings given by subjects and the distance of a novel form from real words, i.e., its analogy score. In an additional study, participants were asked to produce a list of words that each stimulus made them think of. Greenberg and Jenkins reported that most responses were phonological neighbors of the stimuli. For example, for the stimulus /**swɪt**/, accumulated responses from all participants resulted in the following list: *wit*, *quit*, *spit*, *sit*, *sweat*, *sweet*, *Swiss*, *slit*, *swat*, *swim*, *swig*, *switch*, and *swing*. They claimed that this kind of result was clear evidence that speakers used substitution as a strategy to judge novel forms.

2.2.1.1 Factors in Measuring Analogical Strength

There are several concerns with respect to conclusions made about the effects of the analogical mechanism in the study by Greenberg and Jenkins (1964). The first set of concerns is about the particular properties of analogical strength used in the study. For example, Greenberg and Jenkins did not include phoneme deletion or insertion as an operation that could result in a neighboring word. This seemed like an arbitrary decision, and one not consistent with the results obtained from speakers. As can be seen from the examples given above for neighbors of /swɪt/, participants used deletion as an operation to look up words in their lexicon that were similar to the stimuli.

Second, the measure of analogy used by Greenberg and Jenkins assumed that phonemes in any given position were equivalent. While it is generally assumed that similarity between two forms depends on comparing phonemes that occur in corresponding positions, Bailey and Hahn (2004) have claimed that different prosodic positions have unequal roles in computing similarity. For example, they have suggested that when comparing two forms, an inserted phoneme in coda position makes a smaller difference than a substituted phoneme (e.g., /stɪp/ and /stɪm/ are judged to be more similar to /stɪmp/, than to each other).

In addition, Greenberg and Jenkins considered the number of phonological neighbors to be a sufficient measure of lexical analogy, regardless of the degree of similarity between the substituted phonemes. This meant that they were assuming that all phonemes are equally similar to each other. We know this assumption to be incorrect because not only do the underpinnings of feature theory predict that phonetic similarity would depend on shared phonetic features, but a number of psycholinguistic studies have indicated that speakers are aware of different degrees of similarity among phonemes (e.g., Derwing and Nearey, 1984; Frisch, 1996; Bailey and Hahn, 2005).

In fact, Greenberg and Jenkins report that the participants in the study made more substitutions that were "phonetically similar" to the target. For example, *skit* was not listed as a similar word to /swɪt/ by any of the participants, while *slit* was. This enforces that phonetic similarity should be included in computing a target's distance from real words in the lexicon. While there is evidence that speakers are sensitive to degrees of similarities among phonemes, disregarding this sensitivity in computing analogy is common practice (e.g., Goldinger, Luce and Pisoni, 1989; Newman, Sawusch, and Luce 1997; Luce and Pisoni, 1998; Vitevitch and Luce, 1998, 1999; Pylkkänen, Stringfellow and Marantz, 2002; Lipinski and Gupta, 2005). Finally, word frequency of neighbors has also been shown to be an influential factor in measuring analogical strength (Luce, Pisoni and Goldinger, 1990; Vitevitch and Luce, 1998, 1999; Bailey and Hahn, 2001), which is a measure that was absent from the analysis given by Greenberg and Jenkins and various other studies.

In this thesis, I have used a measure of lexical analogy based on the Generalized Neighborhood Model (Bailey and Hahn, 2001). This model considers the following factors: 1) the degree of phonetic similarity between target and existing words; 2) the frequency of those existing words; and 3) no sharp distinction between neighbors and non-neighbors (i.e., neighbors are considered to exist on a continuous scale). This model is discussed in more detail in section 3.1.

2.2.1.2 Does the Influence of Analogy Exclude a Grammatical Influence?

A second concern, one crucial for understanding phonotactic intuitions, is that while the high correlation between analogical strength of novel forms and acceptability ratings from subjects suggests that an analogical mechanism contributes to speakers' judgments, this does not exclude grammar as a contributor to the ratings. In fact, in a study by Ohala and Ohala (1986), not all the observed patterns could be explained within an analogical model. The aim of Ohala and Ohala's study was to test the explanatory difference between the analogical model of Greenberg and Jenkins and the generative grammatical model of Chomsky and Halle (1968).

Their stimuli were designed such that the two models predicted different acceptability ratings. For example, the forms /θlɛz/ and /θlɛd/ would both be rejected by the grammatical model, while the analogical model would predict /θlɛz/ to be rated better than /θlɛd/. However, the results showed that speakers rated some forms in a fashion that neither model would predict. For example, subjects judged /ʃrɪd/ as more acceptable than /ʃriz/, even though neither their analogical model (it predicted /ʃriz/ as more acceptable), nor the grammatical model (it made no distinction) predicted the observed data.

In a follow up experiment, Ohala and Ohala (1986) presented the stimuli in pairs and did, in fact, find that the results of this kind of presentation were more in line with Greenberg and Jenkins' model. They claimed that their results suggested that speakers referred to the lexicon, and not to lexicon-independent rules of the grammar. Their argument was based mainly on forms such as /mløf/ and /spøf/, in which the existence of a non-English vowel would render both unacceptable in a grammatical model, but speakers showed a preference for /spøf/. This argument neglected the fact that a form like /mløf/ has an additional phonotactic violation, namely it contains the illegal onset /ml/. Therefore, a comparison of the forms within a grammatical model could also predict a preference for /spøf/, as opposed to/mløf/, if the model could calculate the number of violations as opposed to only distinguishing possible and impossible forms.

To summarize, the data and the analysis of the data are consistent with a mixed model in which both mechanisms (phonotactic grammar and lexical analogy) are active. If well-formedness judgment tasks engage both mechanisms, we would expect to find evidence of an effect of the analogical mechanism, as the above-mentioned studies found. However, the above studies did not test, or exclude, a combined model.

2.2.2 Grammatical Theories of Phonotactic Well-formedness

At the other endpoint of the studies that try to interpret the results of well-formedness tasks, we find those that argue for only a phonotactic grammar at work.

This view proposes that the grammar is the sole mechanism engaged in acceptability ratings. On this view, the lexicon itself plays no role in deeming a form (un)acceptable, despite the fact that grammatical constraints and/or rules are posited based on generalizations over words that are learned and stored in the lexicon. In other words, while there are certain inherent linguistic preferences that are lexicon-independent, there is also grammatical knowledge that is mediated by the lexicon and linguistic experience (e.g., Clements and Keyser, 1983; Hammond, 1999).

One type of a supporting argument for the grammatical account comes from rejected forms such as *bnick*. In these cases, analogy cannot explain robust judgments in rejecting these forms, in spite of the presence of close neighbors, such as *nick* or *brick*. Therefore, it may be that *bnick* is rejected by a phonotactic grammar that relies on well-formedness constraints or rules that are spelled out in the mental grammar.

From this perspective, the effective mechanism in predicting speakers' well-formedness judgments is the grammar. To this end, there have been attempts to provide a complete generative analysis that accounts for speakers' phonotactic knowledge. For example, in the phonological analysis offered by Clements and Keyser (1983), a novel form such as *bnik* is judged as ungrammatical on the basis of a syllable-structure condition (i.e., it enforces a certain syllable structure, as opposed to banning an illegal one). This condition requires syllable-initial clusters to have the following feature composition: [-sonorant][+sonorant, -nasal]. In the phonological analysis offered by Hammond (1999, 2004) the novel form *bnik* is ungrammatical, because it violates a highly-ranked constraint that bans stop-nasal sequences as syllable onsets in English.

2.2.2.1 Can a Grammatical Account Explain Observed Gradient Judgments?

A defect of many current grammatical accounts of phonotactics is that they render simple up-or-down decisions concerning well-formedness (i.e., they provide a categorical judgment) and, hence, cannot account for gradient judgments. But when judgments are elicited in a controlled fashion from speakers, they always emerge as gradient and include all intermediate values.

It might be the case that it is the nature of a judgments task (e.g., asking subjects to use a scale) that induces gradient judgments, and that is an issue independent of the underlying knowledge (Schütze, 1996). Frisch et al. (2000) collected well-formedness judgments of the same stimuli under two different conditions: 1) when speakers were given a scale of "1" to "7", and 2) when speakers were given a binary choice of "yes" and "no." They found that the binary acceptability results and the 1-to-7 scale of wordlikeness "essentially produce the same results."

As Schütze (1996) points out, one reason for our interest in grammaticality judgments stems from a need to access information that "scarcely exists within normal language use at all." For example, speakers in the present study rated novel forms [bɪs] and [glɑrm] on the higher end of the acceptability scale, as opposed to [ritk] which they judged as ill-formed (defined by the lowest possible ratings). Nevertheless, most speakers (over 80%) rated [bɪs] better than [glɑrm]. Even though speakers' performance has been repeatedly found to be gradient with respect to the processing of novel forms (Ohala and Ohala, 1986; Frisch, Large and Pisoni, 2000; Bailey and Hahn, 2001; Albright and Hayes, 2003; Hay Pierrehumbert and Beckman, 2004; Hammond, 2004; Vitevitch and Luce, 2005; Albright, 2006), the establishment of formal models that can predict such patterns of gradience remains an unsolved problem for generative phonology.

There has been a growing move towards developing probabilistic grammars that reflect speaker's gradient judgments. By probabilistic grammar, I mean a grammar that finds regularities and probabilities of segments, structures, and segment sequences in a particular structure, from the observed data, and uses variables and values that would predict such data. For example, Coleman and Pierrehumbert (1997) model speakers' well-formedness judgments using a stochastic context-free grammar that uses probabilistic measures. This grammar uses statistics based on phonological constituents (onsets and rimes) to predict gradient judgments. The details of their model, which is used in this study as a measure of phonotactic grammar, are discussed in section 3.1. The grammatical scores obtained by Coleman and Pierrehumbert's parser were highly correlated with ratings rendered by subjects. Coleman and Pierrehumbert (1997) claimed that their probabilistic grammar was a better ("more psychologically realistic") model of speakers' phonological competence than standard generative phonology, because it could account for a large portion of the variance observed in acceptability ratings.

Another example of a model that uses computational tools along with phonological theory in order to capture speakers' phonotactic knowledge is a maximum entropy model by Hayes and Wilson (forthcoming). Their model, which is, in fact, designed primarily as a phonotactic learner, uses a more elaborate phonological apparatus than Coleman and Pierrehumbert's model. In using features, autosegmental tiers, and metrical grids, Hayes and Wilson provide a grammar that learns phonotactic systems by learning a set of constraints and associating each constraint with a weight that reflects the strength of that constraint with respect to the full set of constraints. In Chapter 4, I will compare the predictions of the model proposed by Coleman and Pierrehumbert (1997) and the one proposed by Hayes and Wilson (forthcoming) for the stimuli used in the current study.

2.2.2.2 Does the Influence of Grammar Exclude an Analogical Influence?

There is evidence that speakers are sensitive to regularities and probabilities found in their language. For example, Treiman et al. (2000) have claimed that speakers are sensitive to the frequency of phonemes in particular structures. They identified VC pairs in English that could be transformed into less frequent pairs by switching the consonant members; that is, their stimuli were high frequency V1C1 – V2C2 pairs that could be transformed to low frequency V1C2–V2C1 pairs (e.g., /ɪm/ and /ɑb/ transformed to /ɪb/ and /ɑm/). Subjects in their experiment were presented with monosyllabic non-word stimuli that were constructed based on the above-described VC pairs. This allowed Treiman et al. (2000) to control for phoneme frequency, while they tested for speakers' implicit knowledge of grammatical structure frequencies. The subjects judged the high frequency rimes as better than the low frequency rimes. Treiman et al., therefore, claimed that native speakers are not sensitive simply to segment frequency, but are also sensitive to grammatical structures (e.g., rimes) and the frequencies of those structures. However, the results could not exclude analogy as a potential contributor to the task, because the study did not control for the effect of analogy. In other words, the cleverly transformed structures were not matched with respect to their analogical strength. Therefore, the results were not able to distinguish the effects of the analogical and grammatical mechanisms.

Similarly, the earlier results reported from the model of Coleman and Pierrehumbert (1997) could not exclude lexical analogy as an influential mechanism. However, the goal of that study was not to decide between the fundamentally different views on phonotactic well-formedness judgments outlined in this thesis. Consequently, the stimuli were not designed specifically to test the effects of the phonotactic grammar as opposed to lexical analogy. Their finding indicated that phonotactic grammar plays a significant role in predicting acceptability ratings, but it did not exclude the role of lexical analogy.

In the context of determining the predictive powers of analogy and grammar, Frisch, Large and Pisoni (2000) conducted their own well-formedness judgment task. Their stimuli were multisyllabic non-words made up of syllables that fell into High probability or Low probability categories. By combining the syllables into multisyllabic non-words, the stimuli covered a wide range of grammatical probabilities. The acceptability ratings observed in their study showed a high correlation with grammatical probability scores. Frisch et al. also compared their data with predictions of several models, one of which was an analogical model. They reported that the analogical model also showed a high correlation with the acceptability ratings and that the comparison of the correlations showed that the models' predictions were not significantly different from each other. The high correlations obtained from both models illustrated that either model, analogical or grammatical, could predict speakers' acceptability ratings to a great extent. However, they were not informative in distinguishing between the two models.

Yet, there are studies that do seek to distinguish between the two models. Vitevitch et al. (1997) sought to do so indirectly, by trying to determine whether phonotactic knowledge was independent of the lexical items stored in memory. They asked subjects to rate non-words based on how "English-like" they sounded. Their stimuli were bisyllabic words of the form CVCCVC, which varied with respect to the phonotactic grammaticality scores for each CVC syllable. The phonotactic scores were calculated based on two measures: 1) how often a segment occurred in a specific position within a word, and 2) bigram probability, i.e., the probability of co-occurrence of two adjacent segments. The stimuli fell into four groups based on the phonotactic probability of each syllable (High-High, High-Low, Low-High, Low-Low). This model of positional probability does not make use of any aspect of speakers' knowledge of phonological structure.

Combining the syllables to form bisyllabic CVCCVC non-words would require two additional grammatical measures. First, while a score for phonotactic probability was calculated for each syllable, the bisyllabic non-words were created without re-examining the probability of the resulting consonant sequences. For example, /tʃʌnfʌl/ and /fʌltʃʌn/ were both in the same condition that reflected the High-High category. Second, stress placement had no role in determining the category into which a stimulus fell, with respect to its grammatical probability score. For example, /ˈfʌltʃʌn/ and /fʌlˈtʃʌn/ were both in the High-High condition. These flaws in the calculation of phonotactic probability introduced complications into the interpretation of results. Most importantly, the items were not controlled for their analogical strength. We know that segment sequences with high grammatical scores tend to occur in dense lexical neighborhood (i.e., those with high analogical scores). As a result, while Vitevitch et al. reported that subjects' acceptability ratings were highly correlated with grammatical probability, the question of whether these effects were uniquely grammatical and independent from analogy remained unanswered.

2.2.2.3 Grammatical Knowledge May Not be Only Generalizations over the Lexicon

While it has been suggested that the source of the gradience in well-formedness judgments may be generalizations over the lexicon, this is not to say that all grammatical knowledge is derived from the lexicon. A grammatical account often assumes a further element, Universal Grammar, while an analogical view is generally taken to be purely inductive. There is supporting evidence for a view in which speakers show knowledge of certain linguistic properties that are independent from their lexicons. For example, Greenberg (1965) observed that initial consonant clusters occur within a universal hierarchy of clusters. The higher up in the hierarchy a cluster is (e.g., /bl/ › /bn/), the greater the number of languages that contain that cluster. If this kind of knowledge were innate, it would make a clear prediction. The prediction would be that this knowledge would be observable in speakers' judgments on stimuli that contain structures that are unattested in their language.

This is, in fact, what Pertz and Bever (1975) found. They used CCVC stimuli pairs that varied on the initial cluster (e.g., *mlit-lmit* and *rbeek-rneek*), in order to examine speakers' knowledge of certain co-occurrence restrictions on initial clusters. Participants showed sensitivity to Greenberg's proposed universal cluster hierarchy. Their sensitivity provides evidence for a bias that could not have been formulated based only on generalizations over linguistic experience.

This claim has been further supported by recent experimental results from Berent et al. (2006). In a series of tasks, they observed a similar preference by speakers regarding restrictions on the sonority profile of onsets. The first task in their study required English speakers to judge the number of syllables in auditory stimuli, e.g., how many syllables there are in *lbif*. In the second task, participants were asked to decide whether pairs of stimuli (e.g., *lbif-lebif*) were the same, or different. Finally, the same pairs were used as target-prime pairs in an auditory lexical decision task. Berent et al. claimed that speakers had preferences about linguistic structures that were independent of their lexicon and reflected inherent biases.

This preference has been observed for not only structure, but also linguistic processes. For example, results from Wilson (2003) indicated that speakers had biases when exposed to artificial languages that either had an assimilation/dissimilation rule that was attested in natural languages (but not in the speakers' language), or one which was random and not based on any formal linguistic relationship between segments involved in the process. The results indicated a particular type of bias, such that participants were more likely to learn the "natural" processes than the random ones.

In summary, the experimental literature on phonotactic well-formedness judgments supports the presence of an influential grammatical mechanism. However, the studies reported here did not test, or exclude, a mixed model.

2.2.3 Do Analogical and Grammatical Effects Co-exist?

So far, I have discussed two fundamentally different approaches to modeling phonotactic intuitions. One approach is based on analogy with existing forms, and the other is based on the rules or constraints of a grammar. However, there is no a priori reason to exclude the possibility that both factors play a role in native speaker judgments. In fact, there are reasons to seriously consider and investigate that possibility. For example, portions of the data reported by Ohala and Ohala (1986) were better predicted by an analogical model (based on Greenberg and Jenkins (1964)) than by a grammatical model (based on Chomsky and Halle (1968)). On the other hand, certain aspects of the data were best explained by a grammatical model (discussion in section 2.2.1). Therefore, the data as a whole are consistent with an account in which both mechanisms co-exist in judgments of phonotactic well-formedness.

Just this claim was made by Bailey and Hahn (2001), who attempted to determine whether the effects of analogy and grammar co-existed. They rightly pointed out that phonotactic grammar and lexical analogy might be confounded, because the tendency for items with highly grammatical patterns to have high analogical strength has been widely observed (e.g., Luce, Pisoni and Goldinger, 1990; Jusczyk et al., 1993; Charles-Luce and Luce, 1995; Vitevitch and Luce, 1998, 1999). They provided a sophisticated experimental design intended to tease apart the two mechanisms.

The grammatical mechanism was measured by using syllable-part probability (with onset, nucleus, and coda as the units used) in monosyllables. They used only monosyllables in their study, and decided not to use information from any other larger words in their computation of probability.[3] In order to measure the effects of lexical analogy, Bailey and Hahn used the Generalized Neighborhood Model (GNM), to be discussed below in section 3.2.

Bailey and Hahn first created a corpus of non-words that were two phonemes away from real words, and called this set "isolates". For a random set of "isolates," they then created additional non-words that had the following properties: 1) they had lexical neighbors that were one phoneme away, and 2) they were phonological neighbors to the "isolates." This new set was called "near-misses". The full set of stimuli included "isolates," "near-misses," and real words. This method of developing the stimuli was biased in allowing a greater influence for analogical strength because all stimuli had lexical neighbors at either one phoneme away or two phonemes away. In addition, the

[3] This points to their tacit assumption that only monosyllables contribute to grammatical knowledge about monosyllables.

stimuli were not controlled for their range of value in grammatical probability: the "isolate" stimuli were chosen at random, and the "near miss" items were based on these "isolate" forms. As a result of not controlling for grammatical probability, the study could not confidently investigate whether the stimuli at the endpoints of the spectrum for grammatical scores indicated a consistent contribution from the phonotactic grammar.

Despite this flaw, Bailey and Hahn measured the effects of analogy (using GNM) and grammatical probability (using syllable-part probability), and claimed that subjects' ratings reflected effects that were due to both phonotactic grammar and lexical analogy. They proposed that the effects of lexical analogy were not subsumed by grammatical probability. Instead, they claimed that lexical analogy had more influence than grammar on the well-formedness ratings. One reason for the inflated effect of analogy in their study may be that their non-word stimuli were selected based on whether forms had a neighbor or not. This thesis reports experiments whose results suggest that the grammatical mechanism may be the mechanism with a greater influence.

2.3 Experimental Literature on Other Effects of Analogy and Grammar

Grammatical and analogical effects on well-formedness tasks would suggest that we are likely to find similar types of effects on other linguistic tasks. In this section, I review the results from a range of psycholinguistic studies in which the effects of analogy or phonotactic grammar were investigated. In practice, there is an overlap, to a large degree, between forms with high grammatical probability and forms with high analogical strength. As a result, as discussed below, the results of studies were often ambiguous as to whether they provided evidence for effects of analogy or phonotactic grammar.

2.3.1 Evidence from Speech Production

Investigations into speech production is the arena in which the difference in behavior of non-words and words has allowed researchers to make some intriguing claims about the difference in the effect of analogy and grammar on speech processing. It has been suggested that analogy affects the production of words and non-words differently (Vitevitch, 1997, 2002; Harley and Brown, 1998; Vitevitch and Luce, 1998, 1999, 2005; Gordon 2002). This difference might be the first step in guiding us in making a distinction between these two mechanisms.

2.3.1.1 Repetition Tasks

In an immediate repetition task, participants are asked to repeat an auditory stimulus as quickly and as accurately as possible. This task has been most useful in illustrating the independent effects of analogy and grammar on words and non-words. Research on speakers' performance on this task has shown that each mechanism has a distinct effect on processing speed (e.g., Vitevitch and Luce, 1999). For example, Vitevitch and Luce (1998) have claimed that lexicality (i.e., whether a stimulus is word or not) interacts with analogy in repetition tasks, while an interaction is absent between lexicality and grammar. They used CVC words and non-words that fell into two categories: those with high analogy scores and those with low analogy scores. Words and non-words were presented to subjects in separate blocks. The interaction between analogical scores and lexicality was reflected in that words with high analogical scores were repeated more slowly than those with low scores (i.e., those from sparse lexical neighborhoods), while non-words showed the opposite pattern, such that items with high analogical scores were repeated more quickly than those with low scores.

Vitevitch and Luce (1998) used stimuli that had high scores for both analogy and grammatical probability. While this generally makes it unclear as to whether the observed effect is the effect of analogy above and beyond the effect of grammatical probability, they were able to illustrate the unique effect of each mechanism, depending on whether the stimulus was a word or a non-word. Their claim, as restated in Vitevitch and Luce (1999), was that grammatical probability had a facilitatory effect on speech production (e.g., repetition). This facilitatory effect was observed in the faster production of non-words with high grammatical probability. In contrast, they claimed that analogy had an inhibitory effect for words due to lexical competition, and that it was this competition that resulted in the slower response for words with high analogical scores. This effect was absent for non-words because presenting only non-words to participants does not result in lexical competition.

Their claim about the facilitatory effect of grammatical probability is also plausible from the point of learning new words. We do expect speakers to have mechanisms that allow them to learn and repeat words quickly. The above-observed facilitatory effect suggests that highly probable forms may be easier to learn than less probable forms, which is what has been found in some studies. For example, two year-olds have been reported to repeat forms with high grammatical probability more accurately than those with low probability (Zamuner, Gerken and Hammond, 2004). There is also some evidence suggesting that phonotactic probability may aid in learning new words in adults (Storkel, Armbruster and Hogan, 2006).

2.3.1.2 Same-Different Tasks

Auditory same-different tasks are those in which pairs of phonetically similar stimuli are presented to the subject to determine whether the pairs are perceived as the same or as different. Vitevitch and Luce (1999), using this method, argued that the effects of phonotactic grammar and analogy depended on the lexical status of the stimuli. They used CVC stimuli that comprised words and non-words that fell into High and Low groups based on their grammatical probability and analogical strength (the two measures were not manipulated independently). Participants heard either only words, or only non-words (i.e., lexicality was blocked by participant). The speed of processing was analyzed for correct "same" responses, and (similar to the results of the repetition tasks) it was found that there was an interaction between lexicality (word vs. non-word) and the category of items (High vs. Low). They reported that words in the High category were responded to more slowly than those in the Low category, while non-words in the High category were responded to more quickly than those in the Low category.

These results can be interpreted to mean that when speakers recognize that they do not need to access lexical items (e.g., when comparing non-words), they rely on the most useful information, i.e., the phonotactic properties of the stimuli. Since phonotactic probability has a facilitatory effect, the more phonotactically probable non-words are, the more quickly they are responded to. On the other hand, when listeners have to compare words, they are repeatedly engaged in a lexical search. Since words in the High category are most likely to be from dense neighborhoods, the lexical search results in competition among many neighbors, hence, leading to slower responses.

In a follow-up study in which words and non-words were mixed in the same block (i.e., lexicality was not blocked by participant), there was no interaction between lexicality and stimulus category. By mixing words and non-words in the same block of presentation in a same-different task, the category of the word no longer had any predictive power on speed of processing, while non-words continued to show a processing advantage for items in the High category. Vitevitch and Luce (1999) claimed that mixing the stimuli in the same block affected the speed of processing in this task only for words, because the task's design reduced the emphasis on word-recognition, while it increased emphasis on phonotactic processing. This meant that speakers recognized that discriminating words from non-words was irrelevant to the task; therefore, they focused on the phonetic forms of the stimuli, hence, emphasizing phonotactic processing. This strategy also meant that the effect of neighborhood density was decreased, which in turn reduced the inhibitory effect of lexical competition on word processing. Non-words were not affected because their repetition was not affected by lexical competition, regardless of whether they were presented mixed with words or not.

2.3.1.3 Tip-Of-the-Tongue States

Tip-of-the-tongue (TOT) refers to those instances in which one fails to retrieve a known word. Surprisingly, from data collected on spontaneously occurring instances of TOT, this state is not more likely to involve rare words (Reason and Lucas, 1984; Cohen and Faulkner, 1986; Burke et al., 1991). Research in this area also involves analysis of laboratory-induced TOTs. In TOT elicitation tasks, participants are given the definition of, generally, an uncommon word (referred to as a target word), and are asked to retrieve the word that matches that definition. If the participant indicates that they know the word but cannot retrieve it, that instance is labeled as being an instance of TOT. Sometimes, the given definition is followed by a potential blocking word, which could be related to the target, or not. The relation between the blocker and the target could be phonological, semantic, or both.

Using this type of technique, it has been proposed that TOT states are influenced by whether the potential blocking word is in the target's phonological neighborhood and by the target's neighborhood density. While the results of the earlier studies were assumed to suggest a blocking effect of analogy (Jones and Langford, 1987; Jones 1989, Maylor 1990a), more recent studies have suggested facilitatory effects of lexical analogy (Harley and Brown, 1998; Meyer and Bock, 1992; James and Burke, 2000; Abrams, White and Eitel, 2003). For example, James and Burke (2000) claimed that their studies indicated that phonological neighbors of the target behaved as facilitators in lexical retrievals, and not as blockers. They proposed that processing a phonological neighbor would consequently result in a decrease in TOTs, and an increase in the accuracy of lexical retrieval. It must be noted that these proposals were based on stimuli that were not controlled for grammatical probability, which has been suggested to have a facilitatory effect on speech production (section 2.3.1.1). Therefore, it is not clear whether it was the unique effect of analogy that caused the observed increase in ease and accuracy of retrieval.

2.3.2 Evidence from Speech Perception

There is considerable evidence that analogical strength and phonotactic grammar influence speakers' performance on a variety of language processing tasks that involve speech perception. This influence implies that listeners are, in general, sensitive to the similarity between a form and the lexical items in its phonological neighborhood, and that speakers use their phonotactic grammar in a variety of psycholinguistic tasks (Goldinger, Luce and Pisoni, 1989; Newman, Sawusch, and Luce, 1997; Wurm and Samuel, 1997; Luce and Pisoni, 1998; Vitevitch and Luce, 1998, 1999; Vitevitch, 2002; Lipinski and Gupta, 2003, 2005).

2.3.2.1 Phoneme Identification Tasks

In phoneme identification tasks, a subject is asked to identify a phoneme within an auditory stimulus. Using this task, it was discovered that phoneme identification was affected by the lexical status of the item within which the phoneme occurs. This influence was such that speakers were more likely to perceive an ambiguous phoneme as one that would result in the stimulus being a real word (Pitt and Samuel, 1993, 1995). Newman et al. (1997) claimed that, in addition to the lexical status of the stimuli, speakers' knowledge of lexical analogy also influenced performance. Their stimuli were CVC non-words that varied from an initial voiceless stop (e.g., /k/) to a voiced one (e.g., /g/), with varying degrees of voice onset time that made segments' voicing properties ambiguous. It was found that speakers were more likely to perceive the ambiguous segment as one that would result in a form with many neighbors. For example, when the stimulus pair was *gice-kice*, with *gice* having more neighbors than *kice*, listeners were more likely to identify the ambiguous initial segment as /g/. However, when the pair was *gipe-kipe*, with *kipe* having more lexical neighbors, listeners were more likely to recognize the target phoneme as /k/ (Newman, Sawusch and Luce, 1997).

While Newman et al. claimed that neighborhood density affects phoneme identification, the results were ambiguous because lexical analogy and grammatical probability are highly correlated. Therefore, the results could be interpreted as showing that speakers assume they are more likely to hear forms that have high grammatical probability.

In fact, grammatical effects also have been reported on these types of tasks (Pitt and McQueen, 1998; Pitt, 1998; Samuel and Pitt, 2003). For example, Massaro and Cohen (1983) asked subjects to identify phonemes that were ambiguous between /r/ and /l/. Speakers' responses showed an influence of phonotactic grammar, such that speakers were more likely to identify the phoneme as one that did not violate any phonotactic restrictions. For example, following a /t/ in word-initial position, subjects were more likely to identify the ambiguous phoneme as /r/. On the other hand, following a word-initial /s/, more ambiguous phonemes were identified as /l/ than /r/. Massaro and

Cohen proposed that speech perception is influenced by phonotactic grammar. Moreover, the grammatical effect could be observed even when the ambiguous phoneme did not result in a form that would be violating strict phonotactic restrictions. For example, when speakers were asked to identify phonemes in auditory stimuli, target phonemes were detected more easily when they occurred in syllables with higher scores of grammatical probability than those with lower scores (Pitt and Samuel, 1995; Pitt and McQueen, 1998). But highly probable sequences tend to be in words with many neighbors (i.e., with high analogical scores). As a result, this finding could also be interpreted as showing the effect of analogy.

In light of these inconclusive results, Samuel and Pitt (2003) propose that the full pattern of results can be accounted for by allowing a role for both analogy and grammar to influence speech perception, and that due to this co-contribution, the effects are hard to isolate.

2.3.2.2 Lexical Decision Tasks

In addition to affecting phoneme recognition, analogy and grammar affect also whole-word recognition. In a word recognition task, a subject is presented with a sequence of speech sounds, or a string of letters, and has to decide, as quickly and as accurately as possible, whether the stimulus is a word.

It has been commonly reported that the analogical strength of a stimulus affects reaction time and accuracy in word recognition (Goldinger, Luce and Pisoni, 1989; McQueen et al. 1995; Luce and Pisoni, 1998). For example, Luce and Pisoni (1998) reported the results of an auditory word recognition task in which the stimuli varied with respect to their neighborhood density. Words that had high analogical scores (i.e., had many phonologically related neighbors) were recognized more slowly than those that had few neighbors. In addition, they found that words that had low frequency neighbors were recognized more quickly and more accurately than those that had high frequency neighbors. These results have been replicated numerous times (e.g., Luce and Pisoni, 1998; Vitevitch and Luce, 1999; Dell and Gordon, 2003).

Luce and Pisoni (1998) have proposed the Neighborhood Activation Model (NAM), in order to account for the observed data. This model assumes that an acoustic input activates all words phonologically related to the input. Luce and Pisoni's computations of the effects of phonological neighbors on processing included two separate variables: one for the number of neighbors, and another for the frequency of those neighbors. Both measures were assumed to have inhibitory effects on words and non-words, such that items in the High category (i.e., with many neighbors and with frequent neighbors) would be responded to more slowly than those in the Low category. However, their model did not consider the effects of grammatical probability. Therefore, their findings are ambiguous as to whether the observed effects could be attributed entirely to lexical analogy or to grammatical probability, even though the study aimed at examining the effects of the analogical mechanism.

In addition to the effect of analogy, the effect of grammatical probability on word recognition has also been investigated (e.g., Vitevitch and Luce, 1998, 1999). Using magnetoencephalography (MEG), Pylkkänen, Stringfellow and Marantz (2002) examined the presence of facilitatory effects of grammatical probability on lexical decision tasks. As a tool for determining the presence or absence of such effects, they used the M350, which is a response in the left temporal cortex that peaks at 300-400 ms after the visual presentation of a stimulus. If the M350 is a reflection of linguistic activation before a lexical selection is made, then forms with high grammatical scores should show decreased M350 latencies (i.e., the M350 would be observed earlier).[4] Since the M350 peak was observed earlier for stimuli with high grammatical probability, Pylkkänen et al. stated that it must have been a reflection of grammatical probability. This claim implies that the decreased M350 latency (due to the facilitatory effects of grammatical probability) is an effect that must be independent from the increased response latencies (due to inhibitory effects of analogy).

2.4 Experimental Literature on Age-specific Effects

One aim of this thesis is to examine the effects of aging on the two mechanisms (analogy and grammar). Older adults differ from young adults in the following relevant respects: 1) their exposure to language has been longer; 2) they have been using words for a longer period of time; and 3) they have a larger (and perhaps continually growing) lexicon. It is plausible that these characteristics would affect the analogical mechanism but not the grammatical one. Hypothesis 2 predicted that the influence of the analogical mechanism on acceptability ratings would show an increase with age, indicating a positive age-related change.

---

[4] The study does not report whether they considered any effects of orthography, even though the stimuli were presented visually.

Research on older adults has focused mainly on lexical access difficulties or TOTs, and slowed responses. However, extended use of language and linguistic knowledge may have influences that have not been carefully studied (Bybee 2003). Clearly, the presence or absence of these types of influence has implications for our understanding of the language faculty and its processing. There is evidence that healthy aging affects only certain components of the language faculty, such as syntax and semantics (e.g., Kempler and Zelinski, 1994; Friederici, Schriefers and Lindenberger, 1998; Johnson et al., 2001). For example, it has been reported that not only do semantic priming effects remain present in older adults, but that the effects of semantic priming increase with age (Laver and Burke, 1993; Laver, 2000; Taylor and Burke, 2002).

In addition, there is no evidence that lexical knowledge, that is, information about words and their meaning, in adults is affected by healthy aging (e.g., Light 1991; Wingfield and Stine-Morrow, 2000; Thornton and Light, 2006). On the other hand, older adults report difficulty in lexical retrieval as the most pronounced effect of aging (e.g., Schaie and Parham, 1977; Zelinski, Gilewski and Thompson; 1980; Albert, Heller and Milberg, 1988; Brown and Mitchell, 1991). As a result, most of the studies on linguistic changes in older adults have centered on the decreased ease of lexical retrieval and/or increased instances of TOT (Maylor 1990b; Burke et al., 1991; Kemper, 1992; James and Burke, 2000; White and Abrams, 2002; Cross and Burke, 2004; Gollan and Acenas, 2004). Many of the studies in this area use TOT elicitation tasks. It has been widely confirmed that an increase in age highly correlates with an increase in TOT instances (Cohen and Faulkner, 1986; Burke et al., 1991; Brown and Nix, 1996; James and Burke, 2000; White and Abrams, 2002; Vitevitch and Sommers, 2003). However, while the existence of retrieval difficulties in older adults is not argued, the cause of it is highly debated. Theories proposed to explain this phenomenon include general slowing theories (e.g., Salthouse 2000), Inhibition Deficit theories (i.e., deficit in the inhibitory mechanisms (e.g., Zacks and Hasher, 1997), and Transmission Deficit theories (i.e., weakened connections in a linguistic network (e.g., Mackay and Burke, 1990).

As mentioned previously, I have hypothesized that older adults would show an increased effect of the analogical mechanism. I interpret some studies as supporting this hypothesis. For example, using a TOT elicitation task, Vitevitch and Sommers (2003) examined the effect of analogy on word retrieval in older adults. The effect of analogy on speech production has been well-established (see section 2.1 above). However, by comparing the influence of analogy on speech production in older and younger adults, Vitevitch and Sommers were able to demonstrate that this effect changes in normal aging, even though they provided no explanation for what the properties of these changes are, or why they occur.

The subjects in their study participated in a TOT elicitation task where the monosyllable targets fell into eight conditions based on the target's frequency, the target's number of neighbors, and the frequency of those neighbors. While, as expected, older adults showed more instances of TOT, they showed more instances of TOT for words with low analogical scores, while younger adults showed more TOTs for words with high analogical scores. In other words, having many neighbors or having high frequency neighbors resulted in less TOT states for older adults. An explanation of these results requires a model of language processing that can account for the observed age-related effects. Nevertheless, the results imply that older adults show a strengthened effect of analogy.

Gollan and Brown (2006) have proposed that age-related changes in language processing are both positive and negative. Using a TOT elicitation task, they claimed that an increase in age resulted in an increase in instances of TOTs, when TOTs were not adjusted for errors and successful retrievals; however, the increase in age also resulted in a decrease in lexical retrieval failures (i.e., more successful retrievals). In other words, older adults were more likely to be able to retrieve the lexical item that had caused TOT. Gollan and Brown compared these results with the data from bilinguals in which bilinguals showed more TOTs and more retrieval failures than monolinguals. In an effort to account for both sets of patterns, they proposed that retrieval advantages observed in older adults were a result of increased experience and use. As a result, bilinguals who split their language use between two languages did not show a use advantage, while older monolinguals showed an advantage in lexical retrieval, because of the increase in language use and experience with lexical items, and larger lexicons.

Some studies on older adults' reading abilities have also suggested that they are more sensitive to lexical frequencies than are young adults. For example, Spieler and Balota (2000) reported a study in which they compared younger and older adults' performance in reading over 2800 monosyllabic English words. They investigated the influence of three variables: 1) word frequency, 2) orthographic length, and 3) orthographic neighborhood density. Their results indicated that the effect of word frequency increased with age. This effect could be another example of the result of increase in language use and experience with lexical items. There is, indeed, growing evidence that the effect of word frequency increases with age (Balota et al., 2004; Gollan et al., 2006). Spieler and Balota viewed word frequency as a lexical factor, because its influence is possible only when "whole-word" comparisons are performed. Therefore, they concluded that older adults rely less on processing the components of a word, as opposed to whole-word processing. It is not clear whether their measure of density was an analogical factor, because it was based on orthographic neighborhood. Also, the nature of the task involved visual processing, which might slow down in aging, hence, impacting the results without having a linguistic basis. Nevertheless, I expected to find increased effects of lexical analogy in older adults.

2.5 Experimental Literature on Task-specific Effects

One of the studies in this thesis was designed to investigate whether the presence of real words in the stimulus set affects well-formedness judgments (section 2.1). In collecting well-formedness judgments from speakers, some studies have presented subjects with a stimulus set that included both words and non-words (e.g., Greenberg and Jenkins, 1964; Bailey and Hahn, 2001). Other studies have used stimulus lists that comprised only non-words (e.g., Vitevitch et al., 1997; Frisch, Large and Pisoni, 2000). To test hypothesis 3, I wanted to investigate whether the inclusion of real words in the stimulus set would affect the acceptability ratings. The addition of the lexical decision task to the original task could have quantifiable consequences. In a lexical decision task in which a speaker has to decide whether a stimulus is a word or not, the speaker is repeatedly involved in a search of the lexicon before rejecting a non-word. This kind of lexical search causes the activation of the phonological neighbors of the target, because any acoustic input activates multiple phonological forms that compete with each other (Forster, 1979; Marslen-Wilson, 1989; Luce, Pisoni and Goldinger, 1990; Vitevitch et al., 1997; Norris, McQueen and Cutler, 2000). This activation affects both speed and accuracy of processing during word recognition, because the competition among neighbors slows down the processing and reduces its response accuracy. In fact, it might be that lexical decision tasks are the best way to activate phonological neighbors of a target.

The lexical search and the resulting activation of phonological neighbors of a target could be particularly consequential when eliciting well-formedness judgments. The reason is that neighborhood activation is more likely to impact the analogical mechanism than the grammatical one. As a result, the inherent nature of the task is likely to magnify the effect of lexical analogy in well-formedness judgments. The logical question to ask is whether reinforcing a lexical search biases speakers' acceptability judgments to show a greater influence of lexical analogy, and whether removing lexical searches from the task results in a smaller effect of lexical analogy. One hypothesis of this thesis is that including or excluding real words in the stimulus set will have an increasing or decreasing effect on the strength of the analogical mechanism respectively.

2.6 Goal of the Dissertation

Given all the findings that were reviewed in the earlier sections regarding the effect of analogy and grammar on various psycholinguistic tasks, it would not be surprising to expect effects of both mechanisms on phonotactic well-formedness judgments. Distinguishing the effects of lexical analogy and phonotactic grammar is, however, not a trivial matter. The results of most studies are ambiguous as to what effect they are measuring, and how that particular effect is independently influential. This ambiguity, as Vitevitch and Luce (1999) pointed out, generally arises from the fact that segments with high scores from the phonotactic grammar tend to have many neighbors in the lexicon, and hence, have a high analogical score as well.

The problem has persisted in past studies because there has not been a real effort put forward to distinguish the effects of phonotactic grammar and analogy. In order to answer whether well-formedness judgments are a reflection of analogy or grammar or both, we need to compare the two mechanisms' predictions explicitly in the same task, and using the same stimuli. Moreover, the stimuli should be selected such that each mechanism's predictions are distinguishable. In the following chapter, I will describe the formal models of the contributing mechanisms (analogy and grammar) and give details of the method of stimulus selection of the study. Equipped with the formal models and the stimuli, we will be able to accomplish the goal of this dissertation, which is to investigate the nature of phonotactic well-formedness judgments.

Chapter 3: Stimuli and Experimental Design

In the last chapter, I proposed three hypotheses:

1) Underlying phonotactic well-formedness judgments is a combination of two mechanisms: grammar and analogy.
2) The influence of analogy increases with age.
3) Including real words in the stimulus set results in acceptability ratings that show an increased influence of analogy.

Testing these hypotheses required the ability to monitor the effects of analogy and grammar independently in a phonotactic well-formedness judgment task. This necessitated stimuli that varied independently along these two dimensions, in order to observe the strength of each mechanism separately. To this end, I chose two explicit models that provided a measure to quantify the effect of each mechanism. Using a large corpus of non-words, I was able to be selective with respect to the appropriate stimuli. In the following sections, a grammatical and an analogical model that were used in the selection of the stimuli are introduced, followed by the specific properties of the selected stimuli.

3.1 A Formal Model of Phonotactic Grammar: Coleman and Pierrehumbert (1997)

In selecting the stimuli, I used the probabilistic grammar of Coleman and Pierrehumbert (1997). This model employed a rudimentary grammatical representation, based on parsing words into syllables with onsets and rhymes as unanalyzed strings.

While this model did not offer a complete analysis of acceptability judgments, it was able to capture the fact that phonotactic judgments are gradient, as opposed to categorical.

Coleman and Pierrehumbert developed their model for specific purposes that are orthogonal to the investigation in this thesis; specifically, they sought to show the flaws of an approach in which a single violation of any highly ranked constraint in the grammar would determine the evaluation of a non-word (section 2.2.2). The core of their model was a method for calculating the probability of novel forms. The evaluation of the probability of a form was based on the product of the likelihood of each constituent, where likelihood was defined as the number of times the constituent was observed, divided by the number of times the constituent could be expected, based on counts from a representative corpus.

The constituents used were onsets and rhymes of the syllables of each word. The constituents were classified according to: 1) type (either onset or rime); 2) stress (presence or absence); and 3) the position of the structure with respect to word edges. Table 3.1 summarizes the roles:

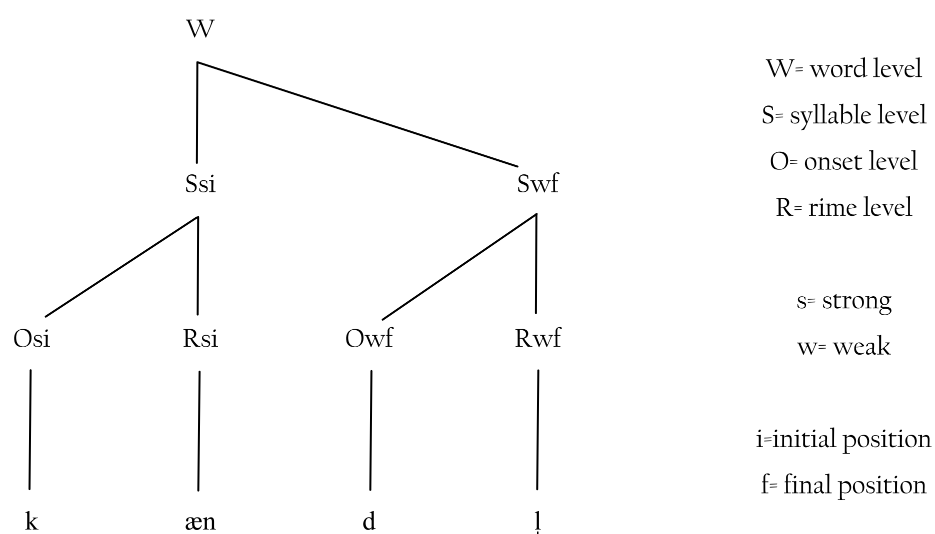
Table 3.1 – Summary of the role parameters used in parsing an English corpus

Measures for roles	**Levels**
Constituent type	Onset | Rime
Presence of stress	Strong | Weak
Location relative to word edge	Initial (and not final) | Final (and not initial) | Initial and Final (monosyllables)

Based on the above categories, each stressed syllable could be one of three types: 1) a strong syllable which was initial (Ssi); 2) a strong syllable which was final (Ssf); or 3) a strong syllable which was both initial and final (Ssif). Unstressed syllables could come in only two forms, namely (1) and (2), because a monosyllable cannot be unstressed (in English). Furthermore, there were four different ways of counting observed onset types: 1) onset (all onset roles), 2) initial onset (osi+osif+owi), 3) stressed onset (osi+osif+osf), and 4) initial and stressed onset (osi+osif). Similarly, there were four different ways of counting observed rime types: 1) rime (all rime roles), 2) final rime (rsf+rsif+rwf), 3) stressed rime (rsi+rsf+rsif), and 4) final and stressed rime (rsf+rsif). A full list of role labels is provided in Appendix 1.

Below is an example of a parse, labeled for category:

Figure 3.1 – Example of a parse for 'candle'



Based on the type of parsing presented above, the model requires the following computations. First, the number of times a particular onset or rime (referred to as "filler" (F)) appeared in a particular role (R) must be calculated, based on a given corpus. The corpus used by Coleman and Pierrehumbert was a subset of Mitton (1992) that contained only monosyllables and disyllables. Therefore, for example, for calculating the probability of /**kændl̩**/, Coleman and Pierrehumbert calculated the number of times each constituent (/k/, /æn/, /d/, and /l̩/) appeared in its current role in that corpus.

The next step was to calculate the number of times the role had been available in the corpus. To compute the grammatical probability (p) of having the filler (F) in the role (R), the number of observations was divided by the number of possible R positions:

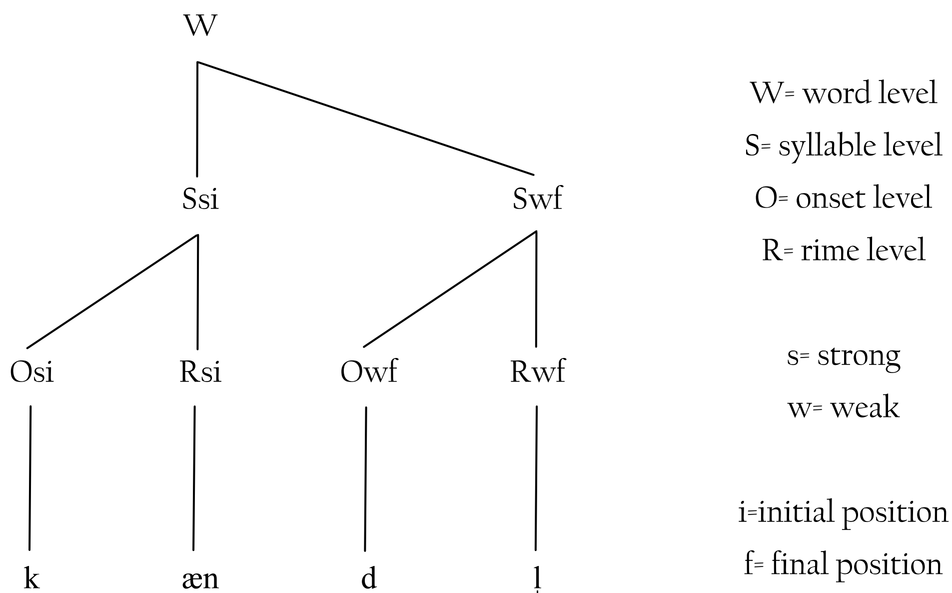
$$p(F|R) = \frac{\text{Number of times F appears in role R (observed)}}{\text{Total number of role R (expected)}}$$

After computing the probability of each filler in its role, the probability of the word was calculated as the product of the filler probabilities. For a disyllabic word, the probability would be:

$$p(w) = ((p(F_{o1} \mid R_{o1})) * (p(F_{r1} \mid R_{r1})) * (p(F_{o2} \mid R_{o2})) * (p(F_{r2} \mid R_{r2})))$$

For example, the probability of 'candle' would be the product of the probability of the four fillers: /**k**/, /**æn**/, /**d**/ and /**l̩**/. Finally, the grammatical score of each word was the logarithm of the product of the probability of each constituent. Therefore, the grammatical probability of a form like “candle” was based on the probability of /k/ as an initial onset, the probability of /æn/ as an initial stressed rime, the probability of /d/ in a final onset position, and the probability of /l̩/ in a final unstressed rime position. An example of the parse, along with the calculations for ‘candle,’ is given in Figure 3.2.

Figure 3.2 – Example of a parse and probability calculation for ‘candle’



1444/22993	29/7227	1627/28345	1785/27345	Calculate the probability for each filler p (F|R)[5]
0.628	0.0025	0.057	0.063	

06.28 * 0.0025 * 0.057 * 0.063 = 5.65 E-07 — Calculate the product of probabilities

Log (5.65E-07) = - 6.25 — Grammatical score is the logarithm of the product

[5] The probabilities given here are based on the corpus used in this study (section 3.1.1), rather than Coleman and Pierrhumbert’s figures.

The probability of the whole word also depended on how much influence one would attribute to the presence or absence of stress, or to the syllable-position with respect to word-edges. These issues are further explored in the following section.

3.1.1 Adapting Coleman and Pierrehumbert's Model for the Present Study

The present study used only monosyllabic stimuli. In this section, I briefly discuss how the model was scaled to fit this purpose. In order to determine the grammatical probability of novel forms and select stimuli for the study, I had to establish segments or segment sequences (i.e., fillers) and their roles that were observed in the existing data (i.e., words in the lexicon). Therefore, a corpus of English words was parsed along the lines of the Coleman and Pierrehumbert's model. The English corpus used in this study was the online Carnegie Mellon University dictionary (2004), with over 60,000 entries. All compounds, contractions, abbreviations, and listed affixes were removed, leaving over 45,000 words to be parsed. The corpus was parsed into onsets, nuclei, and codas. Based on the parsed English corpus, the fillers were determined by examining what segment or segment sequences appeared in onset and rime positions.

The result of the parsing was a full list of fillers found in the corpus along with the number of times they occurred in each role. Table 3.2 illustrates an example of the parser's output:

Table 3.2 – Example of the parser's output for all types of onset roles (raw frequencies)

Onset	Osf	Osif	Osi	Owf	Owi
(null)	513	239	4274	2628	3232
t	573	288	763	4325	94
l	624	329	812	2675	81
d	281	276	1398	1627	847
s	449	317	1400	1559	487
k	302	307	1444	1108	995
r	151	317	1175	1154	1256
m	271	277	1406	1113	269
st	130	188	310	588	22

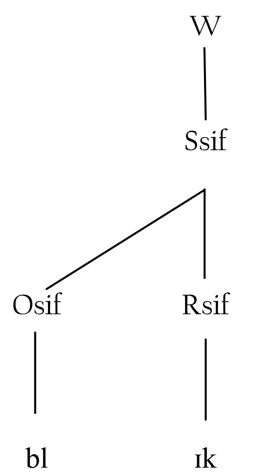
It must be noted that, in calculating the probability score for a form, there were various role possibilities. For example, the likelihood of an onset like [bl] could be determined by examining how often [bl] occurred as an onset in the corpus, or how often it occurred as an initial onset. The latter choice allowed for there to be an influence of position with respect to word edges. Allowing for all role possibilities left me with 16 possible ways of evaluating the probability score for a given form.

In order to determine which role combination would provide the most valid scores, I used the data collected in Albright and Hayes (2003). In their study, participants were asked to inflect a novel form for past tense. As a control for the study, participants were also asked to repeat the novel form aloud and rate the naturalness of the stem on a scale of 1 to 7. Using the Albright and Hayes' ratings collected for 92 novel forms, the best correlation was achieved by using initial onset and stressed rime.

The probability of each constituent of a form was based on the probability score it received after parsing the online dictionary. To give an example, the grammatical probability of *blick* was determined in the following way. After parsing the word, I counted the number of times [bl] appeared as an initial onset (213). Next, to compute the grammatical probability (p) of having the filler (F) in the role (R), the number of observations had to be divided by the number of possible R positions (e.g., 213/39589).

The grammatical probability of each novel string (p(w)) was determined by calculating the product of probabilities (as discussed in section 3.1). As Coleman and Pierrehumbert did in their model, I used the logarithm of the product of the constituents' probabilities, and refer to this measure as **PLog**. For example, for any monosyllable there were two constituents: onset (F_o) and rime (F_r). Therefore, the **PLog** of a monosyllable word (w) was calculated as: $p(w) = \log((p(F_o \mid R_o)) * (p(F_r \mid R_r)))$. Below is an example of how a probability score for *blick* would be calculated:

Figure 3.3 – Example of a parse for *blick* and its PLog (The probabilities given here are based on the corpus used in this study (section 3.1.1), rather than Coleman and Pierrhumbert's figures)



W= word level
S= syllable level
O= onset level
R= rime level
s= strong
w= weak
i=initial position
f= final position

213/39589 106/37321
0.0053 0.0028

Calculate the probability for each filler p (F|R)

0.0053 * 0.0028 = 1.48E-05

Calculate the product of the probabilities

PLog (1.48E-05) = - 4.8

PLog is the logarithm of the product

The stimuli were extracted from a corpus of all possible monosyllabic words in English,[6] which contained close to 20,000 pseudo-words. This corpus contained some forms that could be rendered acceptable only if they were analyzed as bi-morphemic. For example, /zd/ and /spt/ appear in coda positions in English only in suffixed forms. In order to avoid the problem of participants' rating a form more acceptable due to its morphological composition, I removed all forms that contained codas that could be analyzed only as suffixed, leaving about 18,000 pseudo-words. A complete list of codas that were eliminated is provided in Appendix 2. The resulting corpus is referred to as the "stimulus corpus." I calculated a probability score (PLog) for each entry in the stimulus corpus.

3.2 A Formal Model of Analogy: Bailey and Hahn (2001)

As a model of analogy in phonotactics, I used the Generalized Neighborhood Model (GNM) developed by Bailey and Hahn (2001). This model was developed as an adaptation of Nosofsky's Generalized Context Model (1986), which reflects how, in general, people assess similarity (for discussions of this model, see Kruskal, 1983; Bailey and Hahn, 2001; Albright and Hayes, 2003). The intuition behind the model is that a form is highly similar to other words, if there are many words that it is similar to, and if it is very similar to those words.

The advantages of GNM that made it appropriate for this study were as follows. First, the model's measure of similarity reflected similarity to all lexical items. While

[6] The list was generously provided by Mike Hammond (University of Arizona).

common measures of lexical neighborhood are based on the number of neighbors that are one phoneme away (Luce, 1986), there is no arbitrary sharp distinction between a neighbor and a non-neighbor in GNM. Instead, neighbors are considered to be on a continuous scale and a form's lexical similarity is calculated based on all lexical items.

In addition to computing a form's similarity based on all lexical items, this model's calculations of similarity are sensitive to phonetic similarity between forms, and the cost of substitution in the model is based on the phonetic similarity between the phonemes (section 3.2). While distance between forms is usually used with equal weights attached to all substitutions (see section 2.2.1), the GNM aimed for a more realistic measure, which used varying weights that were based on the phonological nature of the substitution. For example, in the model, 'duck' was not the same distance from 'tuck' and 'suck,' because /d/ and /t/ are less phonologically different than /d/ and /s/.

The model's method for calculating the similarity of a form *i* to forms in the lexicon are best understood if we begin by understanding the similarity relations between segments. We, then, develop the full analogical measure, going step by step.

Similarity between two segments was based on the natural class lattice distance metric, proposed by Frisch (1996) and Frisch, Broe and Pierrehumbert (1997):

$$\text{Similarity} = \frac{\text{Shared natural classes}}{\text{Shared natural classes} + \text{Non-shared natural classes}}$$

Identical segments would have identical "shared" and "non-shared" natural classes; therefore, their similarity value would be "1."

The next step in calculating similarity between two forms is to calculate the phonetic distance between their segments. The distance between segments was calculated by deducting the similarity value from "1":

Dissimilarity = 1 – (similarity)

The model also requires a value for the dissimilarity of a segment to null. This value was set at 0.7. The next step is to use these segmental dissimilarity values to calculate the dissimilarity of entire strings. This was based on string edit distance, which is a metric of how alike two strings are to each other (Jurafsky and Martin, 2000). The first step in measuring the distance was finding the lowest-cost alignment of the strings (for discussions of optimal alignment see Kruskal, 1983; Jurafsky and Martin, 2000; Albright and Hayes, 2003).

Dissimilarity is equated with the summed penalties under this optimal alignment. For example, the best alignment between "shine" and the novel form *scride* (from Albright and Hayes (2003)) is given below, along with the computed dissimilarities:

shine	ʃ		null		null		a		ɪ		n	
penalty	.155	+	.7	+	.7	+	0	+	0	+	.667	= 2.222 ($d_{shine-scride}$)
scride	s		k		ɹ		a		ɪ		d	

The next step is to convert dissimilarity into similarity, using the equation below. Here, *D* is a sensitivity parameter, which regulates the relative analogical influence of highly similar versus highly dissimilar forms. When D is high, the score will be dominated by closely similar forms, and when D is low, the influence is spread more widely. This coefficient (which was set to 5.75) indicated how quickly similarity increased as a function of *d*:

$$\text{Similarity} = \exp(-D * d_{i-j})$$

The similarity of a form *i* to the lexicon was based on assessing the similarity of item *i* to every other lexical item, and adding the similarity scores together:

$$\text{Similarity Score}_i = \sum_j \mathit{Similarity}\ (i, j)$$

In the full model, similarity is assumed to be influenced by the frequency of lexical items. In order to avoid zeros, a constant (2) was added to all word frequencies. The GNM model used by Bailey and Hahn (2001) allowed for non-monotonic effects of frequency by using a quadratic frequency-weighting term, as seen in the equation below.

$$S_i = \sum_j (Af_j^2 + Bf_j + C)^{e^{-D.dij}}$$

Where A is negative, the most frequent forms of all will have less influence than somewhat frequent forms. B is a straightforward token frequency term, and C expresses type frequency. Therefore, a similarity score for *blick* would be calculated based on comparing it to all existing words in the lexicon and adding the similarity scores together:

$$S_{blick} = ((-.47 * f_{black}^2) + (.17 * f_{black}) + 1.87)^{e^{-5.75 * d_{blick.black}}} + \ldots + ((-.47 * f_n^2) + (.17 * f_n) + 1.87)^{e^{-5.75 * d_{blick.n}}}$$

The hypothesis of this model is that the phonotactic well-formedness of a word is determined by its aggregated similarity to the lexicon, as just defined. In the following section, I discuss how the model was adapted for this study.

3.2.1 Adapting the GNM for the Study

The version of the Generalized Neighborhood Model used in this study was an implementation by Adam Albright of M.I.T. The model was adapted in the following way. The Bailey and Hahn model's computation of similarity included a frequency-weighted term, which allowed for the effect of a neighbor's frequency in calculating similarity. As discussed above (section 3.2), the GNM allowed for non-monotonic effects of frequency. However, in the adaptation used in this study, the effect of frequency was a linear effect, rather than a quadratic one (i.e., *A* was set to zero). This was unlikely to make a significant difference in the predictions of the model, because in Albright's replication of Bailey and Hahn's results, it emerged as very small (Albright, personal communication). The values for *B* and *C* were set to "1." The equation used in this study is given below:

$$\mathrm{SIM}_i = \sum_j (Bf + C)_j e^{-D.dij}$$

The lexical data base used for computing similarity scores was CELEX (Baayen, Piepenbrock and Gulikers, 1995). Below is a schematic calculation for a novel form *swide*, starting with the terms for the two most similar words, "side" and "slide":

$$\mathrm{SIM}_{\mathrm{swide}} = (1 + 1 * f_{\mathrm{side}} e^{-5.75 * d_{\mathrm{side.swide}}}) + (1 + 1 * f_{\mathrm{slide}} e^{-5.75 * d_{\mathrm{slide.swide}}}) + \ldots + (1 + 1 * f_n e^{-5.75 * d_{swide.n}})$$

SIM swide = .0000661

Below are more examples of the output from the GNM model:

Table 3.3 – Example of scores (SIM) by GNM for items in the stimulus corpus

Stimulus	SIM
biss [bɪs]	1.85E-05
swide [swaɪd]	6.61E-06
balt [bælt]	7.74E-06
Kloomp [klump]	7.98E-07

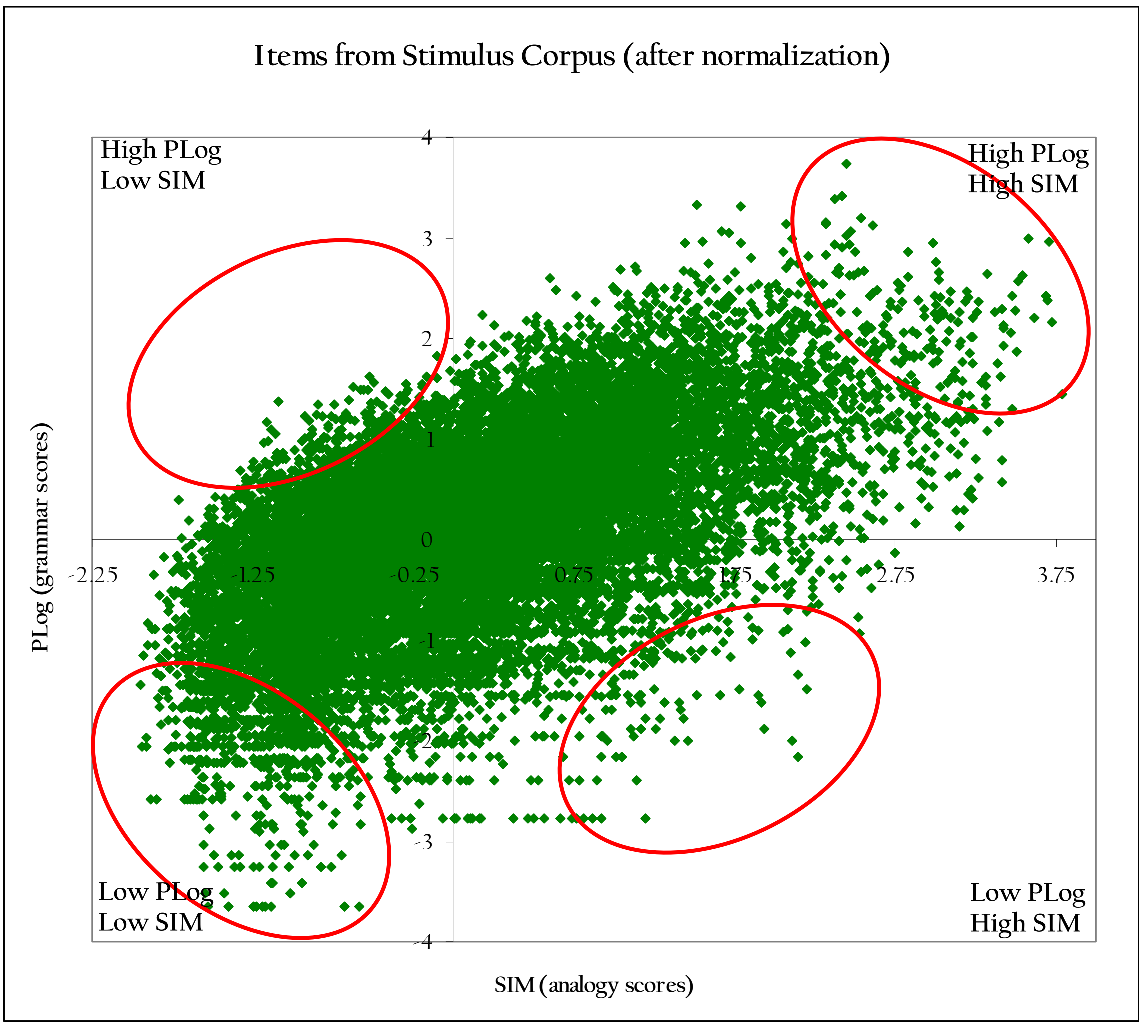
3.3 Phonotactic Violations

In addition, the stimuli included forms that contained outright phonotactic violations, as in traditional generative analytical practice. The presence of these items had two purposes. First, it was important to know whether there was a distinction among items with low PLog scores (LP) and items that contain a phonotactic violation. Second, it was possible that phonotactically illegal forms would provide better insight as to the role of lexical analogy. To this end, a set of 500 phonotactically illegal non-words were created as the starting point in developing the experimental stimuli in this category. I determined the violations based on the phonological restrictions in English proposed by Hammond (1999). For example, restrictions such as "lax vowels cannot occur word-finally" (e.g., */krɪ/) were used in creating the illegal non-words. Items in this subcategory did not originate from the same corpus as the other stimuli. In these conditions, one constituent violated a phonotactic restriction, while the other constituent was a high probability constituent. I selected the illegal items such that the PLog values for the legal constituent were in the same range as the Hi-PLog items. The items in the 500-"word" corpus were classified as Lo-SIM (LS) or Hi-SIM (HS) based on their lexical analogy score.

3.4 Stimulus Selection

Nonce words – In order to select stimuli that covered the full range of variation with respect to the predictions of the two models, both sets of scores for the stimulus corpus were normalized. Figure 3.4 shows the distribution of the items after normalization. The regions of interest are those in which either the two models make the same predictions, or they make opposite predictions.

Figure 3.4 – The distribution of items in the stimulus corpus after normalization of the scores, and the regions of interest



From the stimulus and illegal corpora, I chose 84 monosyllabic non-words for the stimulus set. Monosyllables were used in this study in order to remove the location of stress as a factor in determining the acceptability rating of the stimuli. Table 3.4 summarizes all experimental conditions. Fourteen items were selected for each category (for value ranges in each category see Table 3.5). A complete list of stimuli is provided in Appendix 3.

Table 3.4 – Examples of novel-stimuli in each experimental condition

Grammatical Probability / *Lexical Analogy*	**High Plog (HP)**	**Low PLog (LP)**	**Violation (V)**
High Sim (HS)	***HSHP:*** kowz, prum	***HSLP:*** zɛk, pɛls	***HSV:*** pɛrð, krɪ
Low Sim (LS)	***LSHP:*** flɪst, strin	***LSLP:*** glɑrm, smʌsp	***LSV:*** læfʃ, plɪθf

In practice, it was difficult to maintain the independent effects of grammatical probability and lexical analogy (cf. Vitevitch and Luce, 1999). High probability segments are found in many words, and a word that contains high probability segments tends to have many words in the lexicon that are phonologically similar to that word. As a result, words with high probability segments tend to be in dense neighborhoods. This meant that in offering either effect (grammatical probability or lexical analogy) as an explanation for phonotactic well-formedness judgments, one must first distinguish the effect of grammatical probability from the effect of analogy in the experimental stimuli. To this end, in the two SIM categories (Hi-SIM and Lo-SIM) in the stimulus corpus, items could be selected that belonged to either Hi-PLog (HP) or Lo-PLog (LP) categories. Therefore, a stimulus could be selected such that it had a low analogy (i.e., lexical similarity) score (Lo-SIM), but a high grammatical score (Hi-PLog), and would be labeled as LSHP. Nevertheless, these two factors (grammatical probability and lexical analogy) were not completely independent from each other.

The close relation between the two factors meant that HSLP or LSHP items were rare. As a result, for each factor there were overlaps in the range of values. For example, the range of grammatical probability for HSHP items was [-4.5, -3.1], while the range of grammatical probability for LSHP items was [-5.1, -3.6]. This meant that a novel form that had a PLog of -4.9 (i.e., in the overlap range) was labeled as HP only if it had a low SIM score. The reason was that for the HS items (i.e., with high analogical scores) -4.9 was, in fact, a rather low grammatical score. Below is a chart where the cells show the range of value for each factor, lexical similarity (SIM) and phonotactic probability (PLog):

Table 3.5 – Experimental conditions and their range of values

	Hi PLog (HP)	**Low PLog (LP)**	**Violation (V)**
High Similarity (HS)	PLog (-4.5, -3.1) SIM (13.6, 42.2)	PLog (-6.8, -4.8) SIM (3.9, 83)	PLog – Ø SIM (0.79, 11)
Low Similarity (LS)	PLog (-5.1, -3.6) SIM (2.7, 12.6)	PLog (-7.5, -6.2) SIM (0.02, 2.6)	PLog – Ø SIM (0.008, 0.7)

Real Words – In two of the experiments reported here (Experiments 1 and 2) both novel and real words were presented to the participants. The real words were chosen to fall into the same initial four conditions, HSHP, HSLP, LSHP, and LSLP, as the novel forms. No forms appear in HSV and LSV condition because no real word that is not a borrowed word contains a phonotactic violation. Below are some examples of the real words used as stimuli. A complete list of real words in the stimuli is provided in Appendix 4:

Table 3.6 – Real word stimuli and their experimental conditions

Grammatical Probability / *Lexical Similarity*	**High PLog (HP)**	**Low PLog (LP)**	**Violation (V)**
High Sim (HS)	***HSHP:*** *rise*	***HSLP:*** *crush*	*NA*
Low Sim (LS)	***LSHP:*** *chop*	***LSHP:*** *swamp*	*NA*

3.5 Experimental Design

The stimuli for all three experiments were digitally recorded by a female trained phonetician who was a native speaker of English and was naïve with respect to the purposes of the study. The speaker repeated every item three times, and using PRAAT (Boersma and Weenink), the third token of each was cut and stored.

Stimuli were presented using PsyScope 1.2.5 (Cohen, MacWhinney, Flatt, and Provost, 1993). The participants were asked to rate the typicality of each stimulus on a scale from 1 ("completely normal, this would make a fine English word") to 7 ("completely bizarre, this is impossible as an English word"). They were instructed to ignore the spelling in their scoring. In addition to verbal instructions, written instructions appeared on the screen, before the start of the experiment. The written instructions were as follows:

> "Thank you for participating in this study. You will hear a series of words. You will hear each word twice, and will see its spelling on the screen. Please be advised that the spelling is only an estimate and is

provided only to guide you to confirm the sounds you heard. IGNORE the spelling when giving a score.

For each word, your task is to answer the following: if we were to select some words to add to the English language, what is the likelihood of this word being selected. In other words, 'how good (or typical) an English word would this word make?' Some of the words you hear will be real English words. Judge those the same way – 'how good (or typical) an English word is this word?'

Please use a scale from 1 to 7, with the following endpoints:

1 = Completely bizarre (this is impossible as an English word)
7 = Completely normal (this would be a fine English word)

Please use the numbered keys to enter your response. Please enter your answer as FAST as you can, but maintain ACCURACY. Please press any key when you are ready to START the session. There is no practice session for this experiment. If you have any questions, please ask the experimenter now. Once the experiment starts, you cannot pause."

The order of presentation of stimuli was randomized for each participant. Each participant heard every stimulus twice, with a 500 ms interval, via headphones. The second auditory presentation was accompanied by the presentation of a string of letters (a close orthographic approximation of the stimulus) on the computer screen in front of the participants. The results of the experiments are discussed in the following chapter.

Chapter 4: Experiment 1

The Effects of Grammar and Analogy

This experiment was designed to determine which mechanisms are responsible for phonotactic well-formedness judgments. In Chapter 2, I considered three possibilities: 1) a grammatical mechanism, which rates a form based on the phonological constituents found in the language, and the probability of occurrence of each constituent, 2) an analogical model, which rates a form based on its phonological neighbors, the degree of phonetic similarity among them, and the frequency of the neighbors, and 3) a combination of (1) and (2).

Hypothesis 1 predicted that phonotactic well-formedness judgments would be a reflection of two mechanisms (grammar and analogy). In order to test this hypothesis, stimuli were selected (Chapter 3) to monitor the effects of each mechanism, and Experiment 1 was carried out.

Stimuli – The stimuli comprised 84 non-words and 24 words (as given in section 3.4). The presentation of the stimuli was randomized for each subject.

Participants – Twenty-four undergraduates (17 female, 8 male) from the University of California, Los Angeles participated in this study as part of their Introduction to Linguistics course. All participants were native speakers of American English, right-handed, and under 30 years of age (M=19.6, range = 18-27). None of the subjects reported a history of hearing difficulties.

4.1 Results

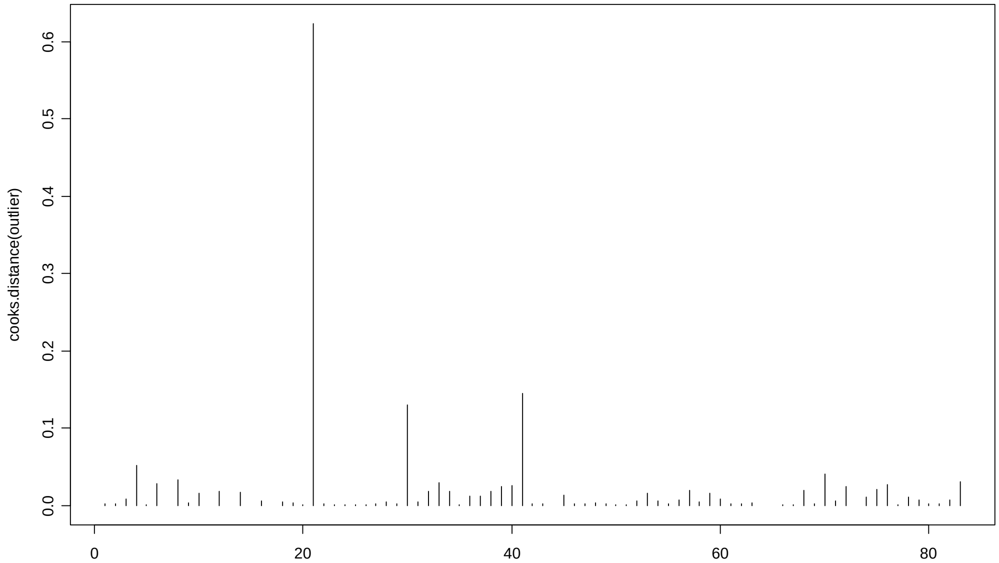
Real words were excluded from the analysis. One non-word stimulus was not recorded correctly and was also dropped from the analysis. Raw ratings for the remaining 83 items were transformed into an approximately normal distribution, using an arcsine transformation:

$$Ratings = \frac{2\arcsin\sqrt{\frac{rawRatings}{7}}}{\pi}$$

The transformation resulted in acceptability ratings that ranged from 0 (ill-formed) to 1 (most well-formed). In order to determine whether there were any items that would significantly impact the analysis, the means for ratings and Cook's distance for each item were calculated.

Cook's distance is a measure for an item in an analysis that shows the change in the multiple regression coefficients that would occur if that case were dropped from the analysis (Stevens, 1996). Therefore, Cook's distance reveals the item(s) that is most influential in affecting the analysis. Figure 4.1 shows the Cook's distance plot for the stimuli in the study. The outlier was *ruck* (which also had a very high SIM value outside the range of other items), which was subsequently dropped from the analysis.

Figure 4.1 – Cook's distance plot: the effects on the coefficients are on the y-axis, and the identity numbers for the stimuli are on the x-axis



4.1.1 Do Analogical and Grammatical Effects Co-exist?

After the outlier was removed, the 82 remaining items were used in all remaining analyses. In order to determine whether both mechanisms (lexical analogy and grammatical probability) contributed to phonotactic well-formedness judgments, I used a mixed-effects model (McCullagh and Nelder, 1989). Because these models are a relatively recent development in statistical analysis, a brief outline of this type of model follows.

A mixed-effects model is an extension of the generalized linear model and is used for analysis of correlated data, such as repeated observations, multiple dependent variables, or longitudinal data. An important feature of mixed-effects models is that they incorporate random effects, as well as fixed effects; therefore, they allow for multiple sources of variation (Baayen, in press).

I selected a mixed model as a tool for analysis for this study because it allowed for analysis of the data using individual ratings. Often, in psycholinguistic studies, one does not want to average across items because it would result in eliminating item variability from the statistical tests (Clark 1973). A central point of interest in most psycholinguistic studies is determining how the variability associated with items affects the speaker's response. Therefore, eliminating that variability would be undesirable. In addition, while most psycholinguistic studies use values that are averaged across subjects, this averaging would eliminate variability among subjects (Lorch and Myers, 1990). Therefore, a mixed model is preferred because it allows for subject variability, as well as item variability.

It must be noted that the use of the term "model" is only referring to a statistical model in which a stimulus' rating can be predicted and calculated based on a mathematical formula. Ultimately, all the reported findings have to be explained by some psycholinguistic theory that explains how well-formedness judgments arise as the simultaneous result of two distinct mechanisms. This thesis has a more limited goal of simply demonstrating this blend of mechanisms. It is to be hoped that fulfilling this goal constitutes a step towards the ultimately correct psycholinguistic processing theory.

The current study had a repeated-measures design, because I had various subjects responding to various stimuli. In this design, the assumptions were that "stimulus" and "subject" were sampled randomly (within their group), and any interpretation of the data could be extended to other participants and other stimuli (i.e., the study could be replicated). For each stimulus, the study had 24 repeated measures (one for each subject). Also, for each subject, there were 82 repeated measures (one for each non-word). I was interested in the following fixed factors: 1) the stimulus grammatical probability (GRAMMAR), and 2) the stimulus analogical strength (ANALOGY).[7] This meant that I needed to test a model in which there were two "random" effects and at least two "fixed" effects.

In the model, as presented in Table 4.1, there are four fixed effects whose coefficients are labeled with β *i*n the table. There are also two random effects. The variances provided in the summary table are the parameters in the model that allow it to calculate the "best linear unbiased predictors" (or BLUPs) for the by-word adjustment (*bi*) and the by-item adjustment (*bj*).

Table 4.1 – Mixed-model results for the acceptability ratings

Linear mixed-effects model fit by maximum likelihood
number of observations: 1968, *groups*: Stimulus, 82; Subject, 24

Random effects:

Group Name	Variance	Std.Dev.
Stimulus (Intercept)	0.0075594	0.086945
Subject (Intercept)	0.0076046	0.087204
Residual	0.0374506	0.193522

Fixed effects:

	Estimate	Std. Error	t -value
(INTERCEPT)	β_0 =0.6604232	0.0517690	12.757
ANALOGY (SIM)	β_1=-0.0060383	0.0040381	-1.495
GRAMMAR (PLOG)	β_2 =0.0446728	0.0061994	7.206
ANALOGY: GRAMMAR	β_3 =-0.0019523	0.0008227	-2.373

[7] We will see in the following chapter that these data are part of a larger data set. However, for purposes of clarity, this subset of the data is presented first – without reference to the larger set of data.

Using the coefficients for the fixed effects' parameters (β s) and the BLUPs (b_i and b_j) extracted from the model[8], I could calculate the model's fitted rating for any stimulus. Below, I have provided the formula for how the model predicted the score (*Mrating*=Model's rating) given for item *i* by subject *j*.

$$Mrating = \beta_0 * 1 + \beta_1 * SIM + \beta_2 * PLog + \beta_3 * SIM * PLog + b_i + b_j$$

For example, for the stimulus *lunk* [lʌŋk], the by-item adjustment is 0.10. For subject *204*, the by-subject adjustment is -.009. Using the coefficient in the above table and the analogy (SIM) and grammar (PLog) scores, we have[9]:

(0.66*1) + (-0.006*6.48) + (0.045* -4.39) + (-0.002*6.48*-4.39) + 0.10 + -0.009 = 0.58

The transformed rating given for stimulus *lunk* by subject *204* is 0.61, which is close to the model's rating above (0.58).

In order to determine which effects significantly contributed to the model's predictive power, we take the following steps. First, one of the parameters is removed from the model, creating a model that is nested within the original model. Next, the two models' performances are compared using ANOVA (Raudenbush and Bryk, 2002). The question to answer is whether each parameter's inclusion in the model is justified.

In the current analysis, the interaction between GRAMMAR and ANALOGY was significant (χ^2 (1)= 5.45, p<.02), which indicated that both effects, ANALOGY and GRAMMAR, were significant. It is noteworthy that the coefficient for analogy and for the

---

[8] All statistical analysis were run using the R statistical programming environment (2007)

[9] All numbers are rounded to the closest third decimal point.

interaction term were both negative. This might suggest that analogical strength was only effective for a certain range of grammatical probability, and not for the whole range of variation. This proposal is explored in Chapter 7.

4.1.2 Comparison of Analogy and Grammar

In order to determine which mechanism's effect was larger, one can compare the sizes of their coefficients. However, this comparison is only meaningful when the two variables (PLog and SIM) are on the same scale and have similar ranges. To this end, the grammatical and analogical scores were transformed to standard scores (also called z-scores) by using the mean (μ) and standard deviation (σ), and using the formula below (where x is the raw score):

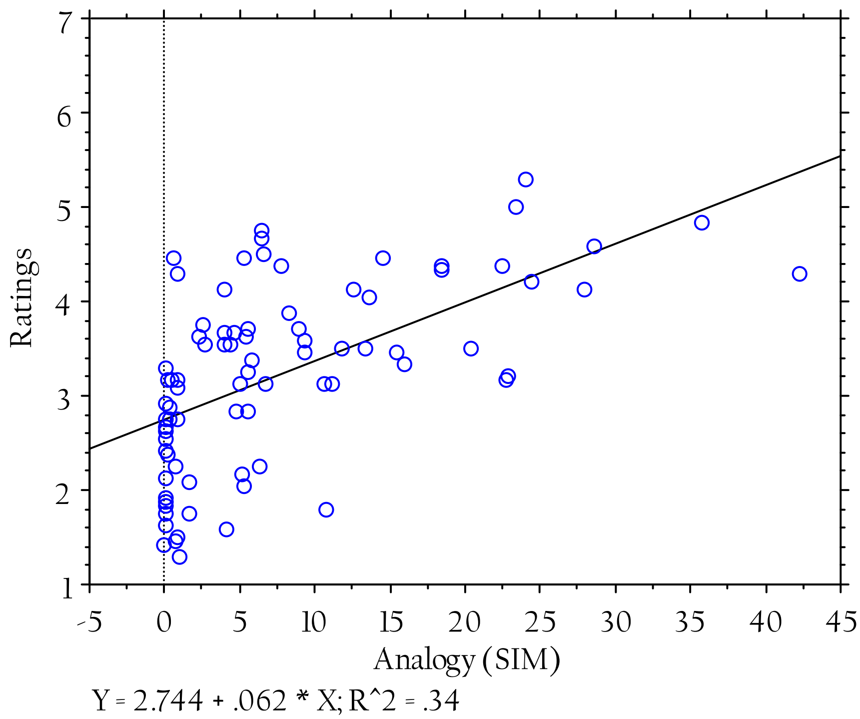
$$z = \frac{x - \mu}{\sigma}$$

Throughout this thesis, when effect sizes were compared, only standardized scores were considered. The standardized scores revealed that while both effects remained significant, the difference in their effect sizes (relative to each other) was not large. The effect of the grammatical mechanism (β=0.07) was slightly larger than that of the analogical mechanism (β=0.06), suggesting that grammatical probability had a slightly larger impact on the ratings than did the analogical score.

Examining the correlations of the ratings with scores from the analogical model (Figure 4.2) and scores from the grammatical model (Figure 4.3) further showed that the ratings were better correlated with grammatical probability than lexical analogy.

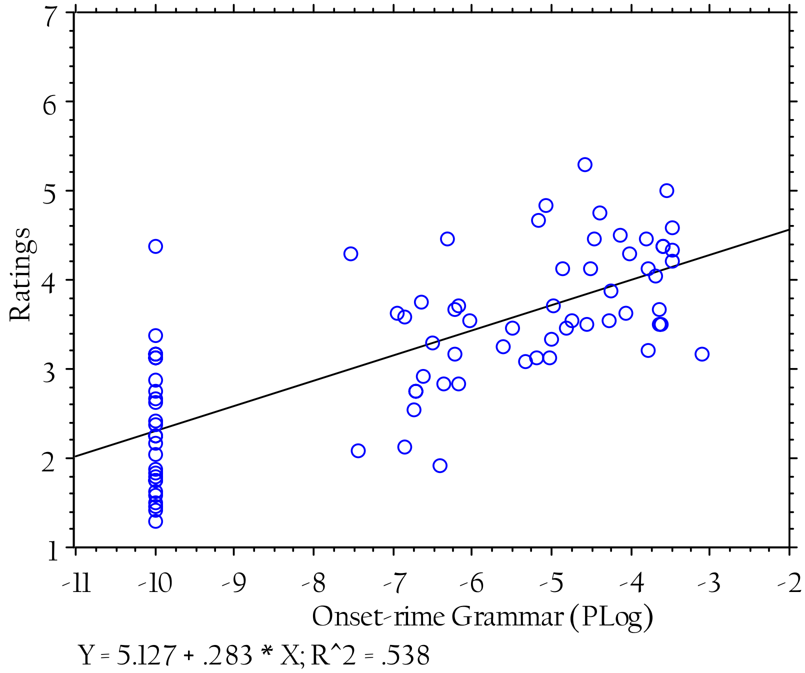
Items clustered in the left corner of Figure 4.2 were items with low analogical scores. The graph demonstrated that the analogical model might not make a fine enough distinction among items with low similarity scores.

Figure 4.2 – Correlation between ratings from subjects (on the y-axis) and scores from the analogical model (on the x-axis)



Similarly, items gathered in the left corner of Figure 4.3 (on the following page) were the “illegal” stimuli, and were not distinguished in the Onset-rime grammatical model, which would assign a probability of zero to constituents that never appear in the lexicon.

Figure 4.3 – Correlation between ratings from subjects (on the y-axis) and scores from the Onset-rime grammar (on the x-axis)



4.1.3 Comparison of the Onset-rime and the Constraint Grammar

In Chapter 2, I discussed two computational models of grammar. I used the Coleman and Pierrhumbert model in the analyses presented in the previous sections of this chapter. I refer to this model as the "Onset-rime" grammar. In this section, I will report on the Hayes and Wilson model's performance. I refer to their model as the "Constraint" grammar.

The analysis using the Constraint grammar showed that the interaction between (CONSTRAINT) GRAMMAR and ANALOGY was significant (χ^2 (1)= 5.5, p<.02), as it had been for the Onset-rime grammar. After the scores were normalized, the effect size of CONSTRAINT-GRAMMAR and ANALOGY were compared. The effect of Constraint grammar was larger (β=0.15) than the effect of analogy (β=0.02). This difference in effect size between analogy and grammar is much larger than the difference observed in the analysis of the Onset-rime grammar.

As Figure 4.4 and Figure 4.5 below illustrate, while both models' grammatical scores highly correlated with the acceptability ratings, the Constraint grammar is better correlated with the ratings. Items clustered in the right corner of Figure 4.4 were the stimuli that did not contain any phonotactic violations, and were not distinguished by the Constraint grammar, which assigns a score to a form based on the constraints (and their weights) that the form violates. Figure 4.5 suggests that the Onset-rime grammar outperformed the Constraint grammar for items with high phonotactic probability. On the other hand, the Constraint grammar made a distinction among phonotactically ill-formed non-words that Onset-rime grammar lacked.

Figure 4.4 – Correlation between ratings from subjects (on the y-axis) and scores from the Constraint grammar (on the x-axis)

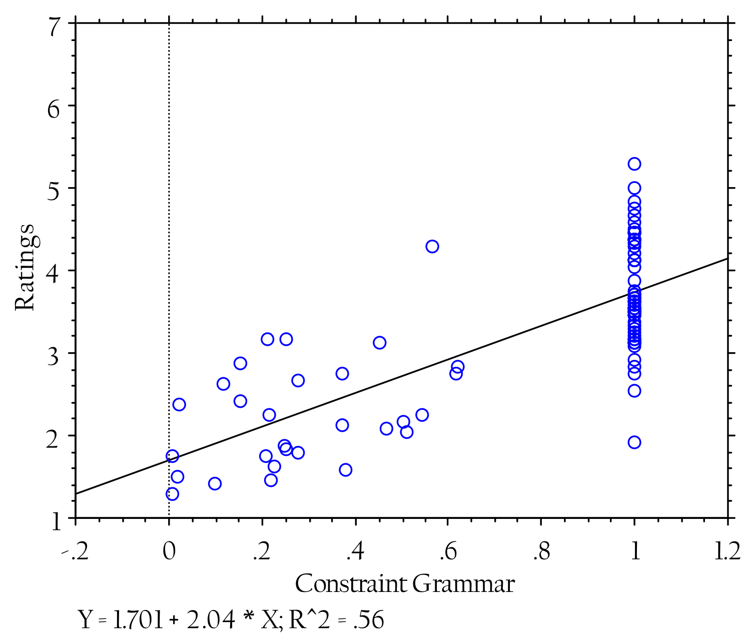
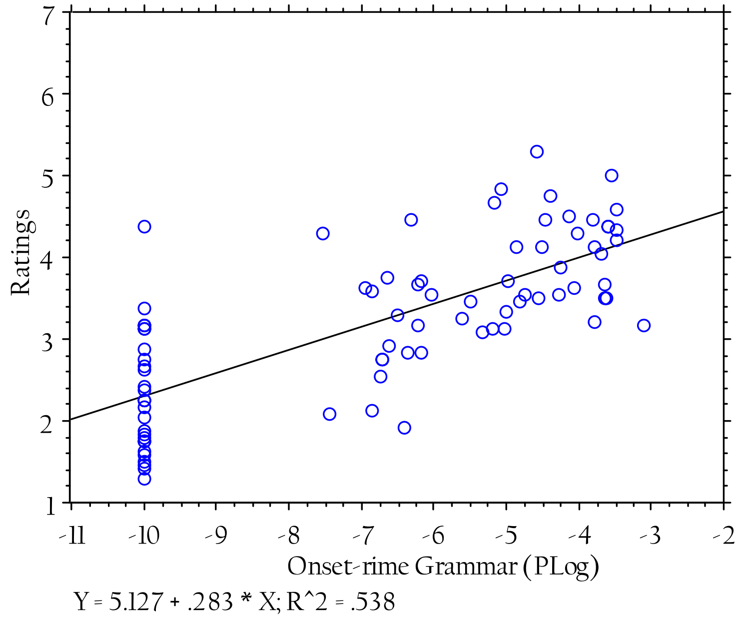


Figure 4.5 – Correlation between ratings from subjects (on the y-axis) and scores from the Onset-rime grammar (on the x-axis)



4.2 Discussion

The central findings of this chapter were the following:

1) Two mechanisms contribute to phonotactic well-formedness judgments: grammatical probability and lexical analogy.

2) The two mechanisms interact in an intriguing manner.

The present findings demonstrated that two mechanisms affected phonotactic well-formedness judgments: an analogical mechanism and a grammatical one. The stimuli used in previous studies were such that they would yield a dichotomous high-low distribution of the factors that were under investigation. Therefore, both mechanisms would make the same prediction regarding an item's acceptability because grammatical probability and analogical strength of forms tend to be highly correlated.

This meant that the results neither ruled out the other factor as having the main effect on the observed data, nor showed that the effects co-exist. The only study that attempted to investigate the co-existence of the two mechanisms was Bailey and Hahn (2004). However, the basis of their stimuli design was the analogical strength of the stimuli.

The current findings showed that each effect (grammatical probability and lexical analogy) was highly correlated with the phonotactic well-formedness judgments. However, the real contribution of the results presented here is that they indicate that the two mechanisms co-exist when the stimuli are carefully selected to represent the whole range of variation. This study demonstrated this effect collecting acceptability ratings on stimuli that varied independently with respect to the effects of grammatical probability and lexical analogy.

The stimuli were selected using the probabilistic grammar of Coleman and Pierrehumbert (1997), and the Generalized Neighborhood Model of Bailey and Hahn (2001). The independent manipulation of the two mechanisms allowed me to monitor the contribution of each mechanism. In doing so, I observed that not only did both mechanisms contribute to well-formedness judgments, but I discovered that lexical analogy had less influence on acceptability ratings. Moreover, the data indicated that the effect of lexical analogy interacted with the effect of grammatical probability. We saw that lexical analogy had a weaker correlation with the ratings than grammatical probability. This difference was also illustrated by its coefficient, which was smaller than the coefficient for grammatical scores, regardless of which grammatical model was used in the analysis.

One intriguing result was the presence of the interaction between the two factors. Recall that the coefficient for the interaction term was negative. This suggested that analogy's influence might be visible only within a certain range of grammatical probability. In order to determine what range of data was mostly influenced by analogy, I had to discern the nature of the interaction between grammatical probability and analogy. The interpretation of this interaction was not straightforward. One plausible explanation was that the negative coefficient meant that the effect of analogy was reduced when the effect of grammatical probability was very high. In other words, when a non-word was predicted to be highly acceptable by its grammatical score, the effect of analogy was insignificant.

In order to investigate whether this explanation was a probable one, I selected a subset of the stimuli that included the top 25% of items with respect to their grammatical probability. For this subset, the correlation of ratings and the analogical scores was insignificant (R^2=.01). Next, I selected a subset of the stimuli that covered the bottom 25% with respect to grammatical probability. For this subset, the ratings were also not significantly correlated with the analogical score (R^2=.16), even though the correlation had slightly improved.

The interesting aspect was that when the items in the middle range of grammatical probability (37 to 63 percentile) were examined, the correlation with analogical scores increased (R^2=.31). This could suggest that the effect of analogy is significant only in a certain range of the data; however, conclusive interpretation of the interaction requires further investigation (Chapter 7). Regardless of the precise influence of the interaction between lexical analogy and grammatical probability, the existence of the interaction indicated that both factors are influential in phonotactic well-formedness judgments.

Chapter 5: Experiment 2

Age-specific Effects

In the last chapter, we saw that phonotactic well-formedness judgments involve two mechanisms, an analogical one and a grammatical one. In this chapter, I focus on phonotactic knowledge in older adults.

As discussed in Chapter 2, older adults have been reported to show more sensitivity than young adults to word frequency in naming tasks (Spieler and Balota, 2000), as well as more sensitivity to neighborhood frequency in TOT elicitation tasks (Vitevitch and Sommers, 2003). Therefore, the subtle effects of a growing lexicon, and extended use and experience with words, both of which are associated with an increase in age, could have quantifiable effects on language processing. In particular, lexical effects could be more robust in older adults' performance, than younger adults'. I was interested in examining this age-related effect in phonotactic well-formedness judgments. Therefore, this chapter centers on answering three questions:

1) Does phonotactic knowledge in older adults reflect a similar combination of the two mechanisms, as it does in younger adults?
2) Do older adults show an increased effect of lexical analogy on well-formedness judgments?
3) Is there any observable change in the effect of grammatical probability in older adults' acceptability ratings?

I had hypothesized that older adults would show a greater effect of lexical analogy in well-formedness judgments, reflecting their increased experience and use of the lexicon. Moreover, I had proposed that there was no reason to expect a change in the effect of grammatical probability. To confirm these hypotheses, Experiment 2 was carried out.

Stimuli – Same as Experiment 1 (Chapter 4).

Participants – 28 older adults participated in the study. All participants were paid for their participation in the experiment. One participant's responses were not recorded due to a technical error. All participants were recruited as right-handed native speakers of English; however, one participant disclosed he was left-handed at the end of the experiment. In addition, all subjects also participated in a lexical decision task, and for two participants levels of accuracy on that task were below 70%. Therefore, the responses of these 4 participants were dropped from the analysis. The remaining 24 participants (15 female and 9 male) were native speakers of American English, right-handed, and over 65 years of age (M=77.4, range = 65 – 87).

5.1 Results

The stimuli used in the analysis and the transformation of the raw ratings were the same as in Experiment 1. In order to determine how aging affects the acceptability ratings, the data from Experiments 1 and 2 were pooled and analyzed using a mixed model (section 4.1.1).

5.1.1 The Mixed-effects Model

This model was discussed in the previous chapter. The data in the current analysis, however, included acceptability ratings from both "young" and "old" groups.[10] Therefore, the model had one additional fixed effect: 1) whether the subject belonged to the "young" or "old" group (AGE), 2) the stimulus grammatical probability (GRAMMAR), and 3) the stimulus analogical strength (ANALOGY). Table 5.1 includes the significant effects in the model, which were determined by comparing nested models (section 4.1.1).

Table 5.1 – Mixed-model results for both age groups

Linear mixed-effects model fit by maximum likelihood
number of observations: 3936, *groups*: Stimulus, 82; Subject, 48

AIC	BIC	logLik	MLdeviance	REMLdeviance
-1350	-1300	683.1	-1366	-1309

Random effects:

Groups Name	Variance	Std.Dev.
Stimulus (Intercept)	0.0069005	0.08307
Subject (Intercept)	0.0128679	0.11344
Residual	0.0378961	0.19467

Fixed effects:

	Estimate	Std. Error	t-value
(INTERCEPT)	β_0 =0.5982370	0.0504523	11.857
AGEYOUNG (AGE)	β_3=0.0555741	0.0337452	1.647
ANALOGY	β_1 =-0.0057060	0.0037237	-1.532
GRAMMAR	β_2 =0.0441574	0.0056927	7.757
AGEYOUNG: ANALOGY	β_5=-0.0014955	0.0006805	-2.198
ANALOGY: GRAMMAR	β_4=-0.0022719	0.0007555	-3.007

Fitted rating = β_0*1 + β_1*SIM + β_2*PLog + β_3*Age + β_4*SIM*PLog + β_5*Age*SIM + b_i + b_j

[10] Note that the results presented in Chapter 4 could be drawn from results presented here. The earlier results were presented first, in order to make the implications of these findings clearer.

Using the coefficients for the fixed effects' parameters and the BLUPs found in the model's summary in Table 5.1, we could calculate the model's fitted rating for a stimulus (for a discussion of the formula and the model's details, see section 4.1). As before, we can compare nested models, in order to determine the significance of each effect. The three-way interaction between AGE and ANALOGY and GRAMMAR was not significant (χ^2 (1) = 1.74, p>0.1). In the following sections, I will discuss the significance of other effects in the model.

5.1.2 Does Phonotactic Knowledge in Older Adults Reflect a Combination of Two Mechanisms, as it does in Younger Adults?

The first step was to determine whether phonotactic well-formedness judgments in older adults were similar to young adults, such that they reflected a mechanism that combined the effects of lexical analogy and grammatical probability. The mixed model analysis demonstrated that older adults and young adults have these same underlying mechanisms influencing their well-formedness judgments.

This conclusion was based on the fact that the model's prediction regarding the combined effect of grammatical probability and lexical analogy remained the same: the interaction between ANALOGY and GRAMMAR was significant (χ^2 (1)= 8.58, p<.005), as it had been in Experiment 1. The interaction indicated that, as before, both effects made a significant contribution to the predictive power of the model.

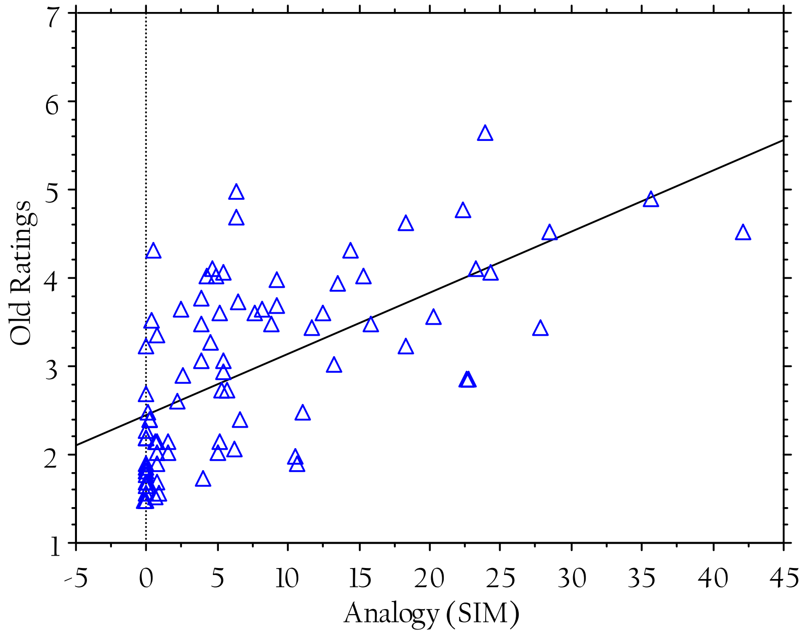
In order to compare the effect size of the two mechanisms, the standardized scores (section 4.1.1) were analyzed. This analysis showed that when both groups were considered, the coefficient for ANALOGY (β=0.08) was in fact slightly larger than the one for GRAMMAR (β=0.07). This meant that while the "old" group illustrated that both mechanisms are at work, the balance between the two mechanisms had shifted in favor of the analogical mechanism. However, the coefficient for the interaction between GRAMMAR and ANALOGY remained similar when the "old" group was added to the analysis, indicating that the two effects were significant, and interacted in the same way for both "young" and "old" groups. I will examine the results of the interaction between age and each mechanism's effect below.

5.1.3 Do Older Adults Show an Increased Effect of Lexical Analogy?

The issue central to this chapter is whether the increase in age has any effect on the contribution of lexical analogy. The main focus was to determine whether there was a crucial parameter in the model that indicated an interaction between the two factors AGE and ANALOGY. In order to establish whether these two factors interact, I had to answer the following question: is the inclusion of the parameter for the interaction of AGE and ANALOGY in the model significant for the model's predictive power? The results indicated that the interaction between AGE and ANALOGY was significant ($\chi^2(1)$= 5.99, $p<0.02$).

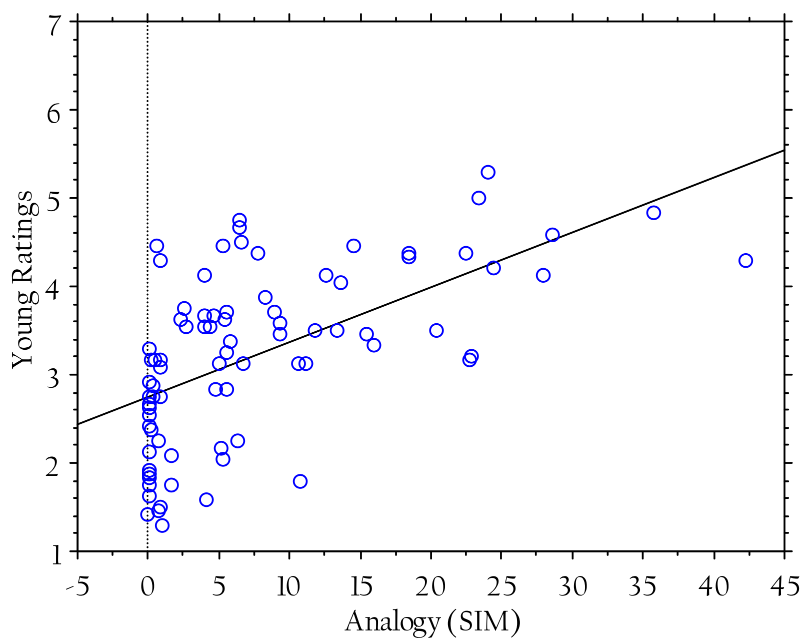
Examining the correlation of the analogical scores with the acceptability ratings from the "old" group (Figure 5.1) and with the ratings from the "young" group (Figure 5.2) also showed that analogical scores were better correlated with the ratings from the "old" group than with the "young" group. Throughout the thesis, "old" data points are represented with △ in graphs, and the "young" data (in Experiment 1) with O.

Figure 5.1 – Correlation between ratings from “old” subjects (on the y-axis) and analogical scores (on the x-axis)



Y = 2.449 + .069 * X; R^2 = .38

Figure 5.2 – Correlation between ratings from “young” subjects (on the y-axis) and analogical scores (on the x-axis)



Y = 2.744 + .062 * X; R^2 = .34

In addition, recall that the stimuli were selected such that there would be items for which the two factors, ANALOGY and GRAMMAR, would make opposite predictions. A visual presentation of the ratings for these items illustrates the interaction. In Figure 5.3 below, the mean ratings are presented for items in the LP category (i.e., items that had low PLog scores). Within this category, there were two subcategories based on the stimuli's analogical scores: HSLP (HiSIM-LoPLog) and LSLP (LoSIM-LoPLog). The difference in mean ratings for the two categories was larger for the "old" group than the "young" group.

Figure 5.3 – Interaction plot for "old" and "young" ratings: responses to HSLP (HiSIM-LoPLog) and LSLP (LoSIM-LoPLog)

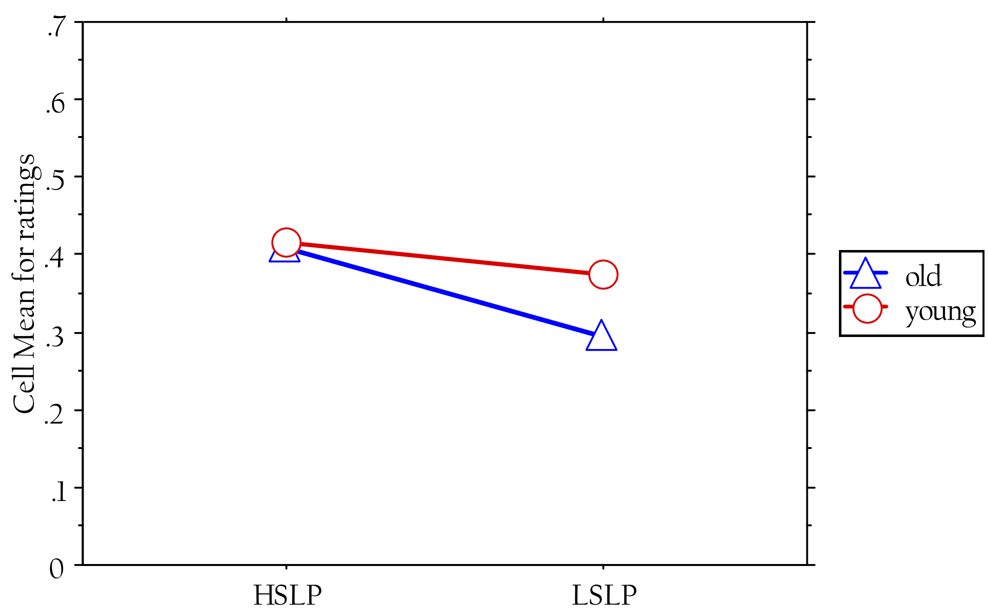
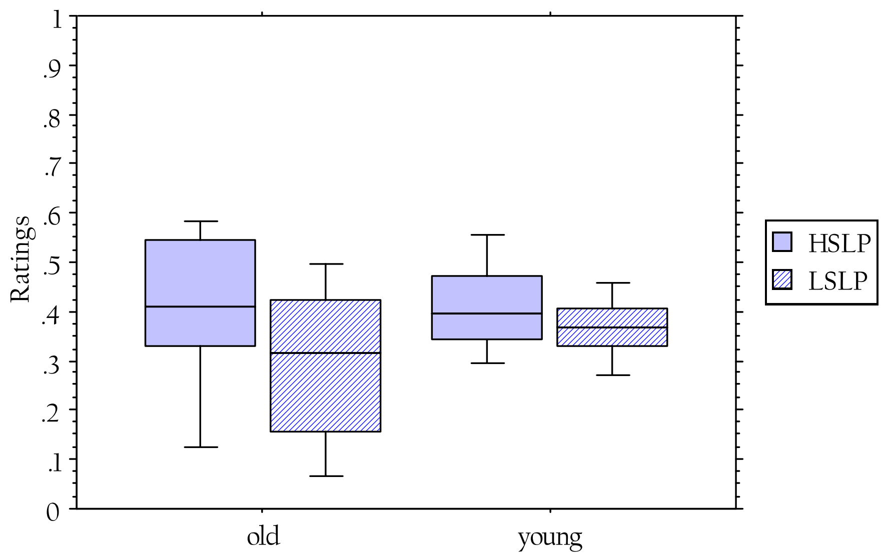


Figure 5.4 shows that, in addition to the AGE and ANALOGY interaction, the ratings from the "old" group had more variation within each category than did the ratings from the "young" group. The negative coefficient for the interaction between ANALOGY and

AGE (YOUNG) indicated that the effect of lexical analogy was different in the two age groups (as was also seen in Figure 5.3) in the following way: the effect of the analogical mechanism on the "young" ratings was smaller than on the "old" ratings.

Figure 5.4 – Comparing the mean ratings (24 subjects in each age group) for LP (Lo-PLog) conditions for old and young groups (HS= Hi-SIM, LS=Lo-SIM)

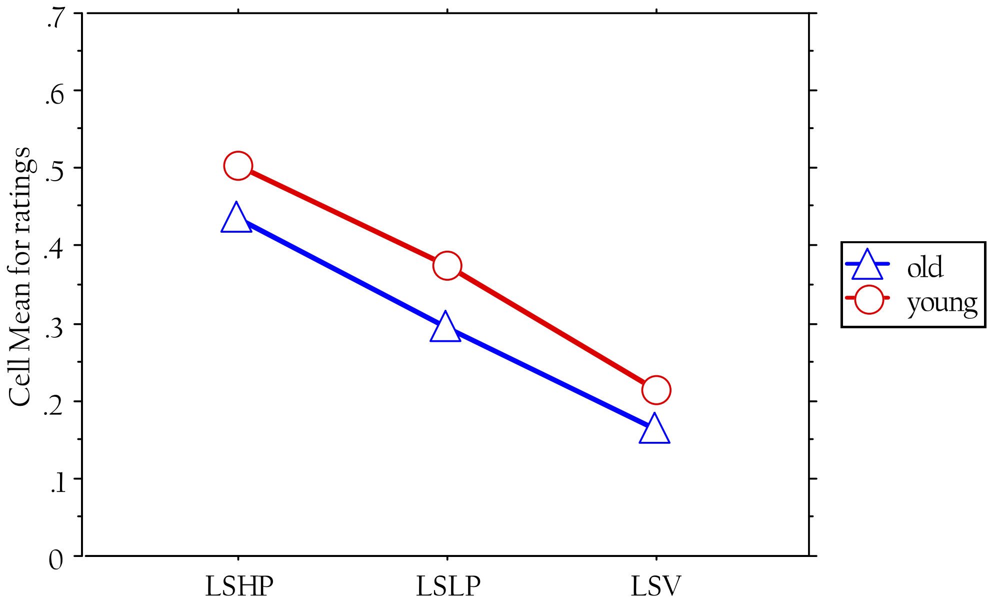


5.1.4 Is There Any Observable Change in the Effect of Grammatical Probability in Older Adults' Acceptability Ratings?

While the interaction between AGE and ANALOGY was significant, I did not expect the effect of grammatical probability to be affected differently in the two age groups because all adults would have the same grammar regardless of their lexicon size and their experience. The interaction between AGE and GRAMMAR was non-significant ($\chi^2(1)$= 1.34, p>.2). This meant that grammatical probability affected the acceptability ratings in the two groups consistently. Figure 5.5 shows the mean ratings for items in the LS condition (i.e., items with low analogical scores). Within this category, there were three subcategories in which the effect of grammatical probability varied: 1) LSHP

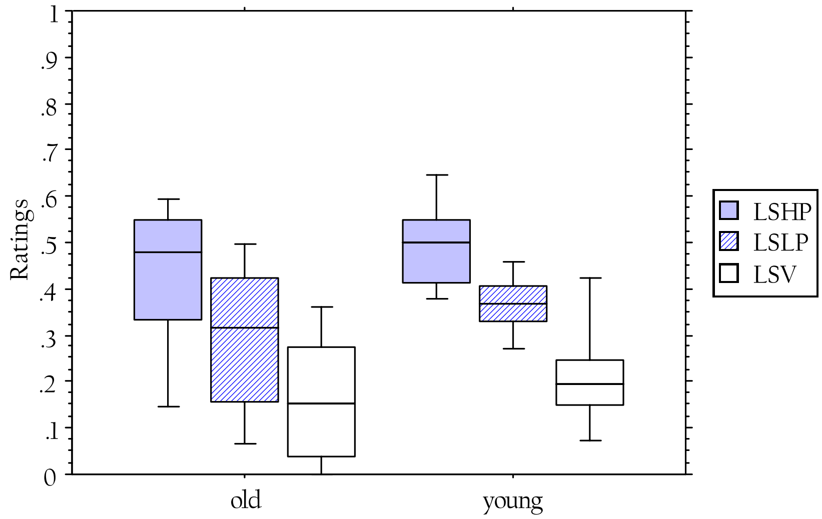
(LoSIM-HiPLog), 2) LSLP (LoSIM-LoPLog), and 3) LSV (LoSIM-Violation). The difference in mean ratings for the three categories is consistent across the two groups.

Figure 5.5 - Interaction plot for "old" and "young" ratings: responses to LSLP (LoSIM-HiPLog), LSLP (LoSIM-LoPLog), and LSV (LoSIM-Violation)



The positive coefficient for AGE (YOUNG) indicated that the "young" group rated items higher than the "old" group, as we saw in Figure 5.5. In addition, similar to the ratings for items in the LP (LoPLog) categories (Figure 5.4), the ratings in the LS (LoSIM) subcategories also showed a greater variation in the "old" group than in the "young" group, as seen in Figure 5.6:

Figure 5.6 – Comparing the mean ratings for LS (Lo-SIM) conditions for old and young groups (HP= Hi-PLog, LP=Lo-PLog, V=Violation)



5.1.5 Comparison of the Onset-rime and the Constraint Grammar

For the young group (Experiment 1), we saw that the two grammatical models (Onset-rime grammar and Constraint grammar) behaved similarly in that there was an interaction between ANALOGY and GRAMMAR, regardless of which grammatical model was used. A similar comparison was not possible for the current set of data, because there was a three-way interaction between AGE and ANALOGY and CONSTRAINT-GRAMMAR (χ^2 (1)= 7.5, p<.007). This meant that none of the two-way interactions could be removed from the model; thus, their significance could not be specifically examined.

The correlations of the "old" ratings with the two grammatical models are shown in the figures below. For the "old" group, the correlations of the ratings with the Onset-rime model (Figure 5.7) and with the Constraint model (Figure 5.8) are both worse than they were for the "young" group (section 4.1.3).

For the "young" group (Experiment 1), we saw that the Constraint grammar was better correlated with the acceptability ratings. However, the "old" ratings were slightly better correlated with the Onset-rime grammar than with the Constraint grammar, which may be a reflection of the lexical influence on the Onset-rime model:

Figure 5.7 – Correlation between ratings from "old" subjects (on the y-axis) and scores from the Onset-rime grammar (on the x-axis)

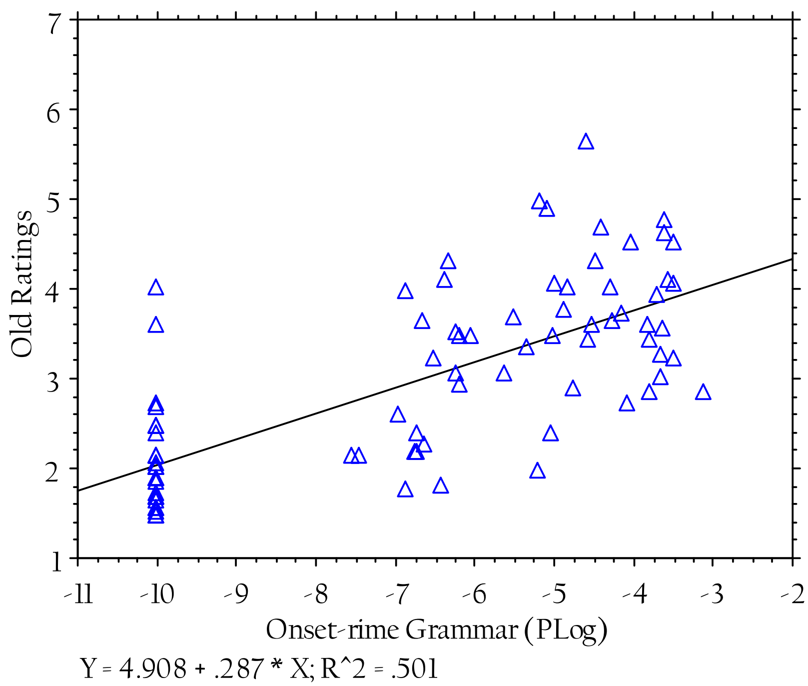
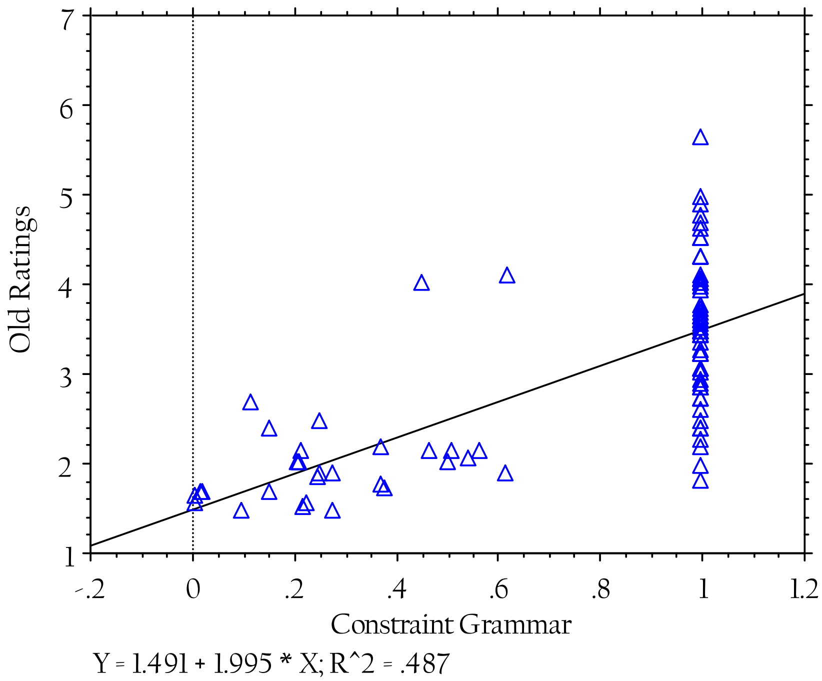


Figure 5.8 – Correlation between ratings from "old" subjects (on the y-axis) and scores from the Constraint grammar (on the x-axis)



5.2 Discussion

5.2.1 Does Phonotactic Knowledge in Older Adults Reflect a Combination of Two Mechanisms, as it does in Younger Adults?

In the last chapter, we saw that well-formedness judgments were best predicted by a model in which not only were analogy and grammatical probability the mechanisms involved, but the two mechanisms were found to interact. In this section, I was able to investigate well-formedness judgments from older adults (over 65). The results indicated that, with respect to the active mechanisms in phonotactic well-formedness judgments, older adults' acceptability ratings exhibited the same property as younger adults': both grammatical probability and lexical analogy contributed to speakers' well-formedness judgments, regardless of age.

The negative coefficient for the interaction between the two fixed effects also remained the same when the two sets of "young" and "old" data were analyzed (Table 5.1). In the last chapter, I suggested that the interaction between the two factors could mean that analogy exercises its effect only within a certain range of grammatical probability. Since the coefficient for this interaction remained the same for the older adults, I considered the "old" data in specific ranges of grammatical probability (PLog), similarly to the subsets examined in section 4.2.

For the subset that contained the top 25% of grammatically probable items, the correlation between the "old" ratings and analogical scores was non-significant (R^2=.06). The correlation between the analogical scores and the "old" ratings for the bottom 25% of grammatically probable items was also not significant (R^2=.14). However, for the subset that contained stimuli with PLog values in the mid range, the correlation increased (R^2=.26). This strengthened the plausibility of the earlier proposal that lexical analogy has the strongest effect when the values of PLog are not at the extreme edges. Also related to this proposal is the difference in the strength of the two mechanisms. If grammatical probability has a much greater influence on well-formedness judgments, then I would not expect lexical analogy to be an influential factor when the values for grammatical probability are at the extremes. With respect to the extremely improbable stimuli, it seems that reducing the effect of lexical analogy is a reasonable strategy in disallowing phonotactically illegal forms. This is the mechanism that results in speakers' rejecting *bnick*, in spite of its neighbors and their frequencies.

I have shown that the main results discussed in the last chapter are consistent across age groups. Phonotactic well-formedness judgments are the result of a mechanism that has two components, an analogical one and a grammatical one. In addition, the two components interact, such that the analogical mechanism's influence seems to be active only when the effect of grammatical probability is not at either endpoint.

5.2.2 Do Older Adults Show an Increased Effect of Lexical Analogy on Well-formedness Judgments?

The central finding of this chapter is that phonotactic well-formedness judgments in older adults show an increased effect of lexical analogy. One example of a positive age-related change in linguistic knowledge is that the lexicon continually grows, and older adults have a larger lexicon than younger adults (Varhaeghen, 2003; Birren and Shaie 2006). Therefore, I hypothesized that this change associated with increased age would result in a stronger effect of lexical analogy. The results confirm that older adults (over 65) exhibit evidence of a strengthened analogical mechanism.

On some psycholinguistic tasks (e.g., word-recognition, TOT-elicitation, sentence production) older adults' performance exhibits differences from that of younger adults' (Kemper at al, 2001; Vitevitch and Sommers, 2003). These differences are often reported as negative age-related changes (e.g., slower processing, more TOTs, less complex sentence structure). There are two possible interpretations of these effects: 1) these age-related changes in performance are not particular to language processing; 2) there are changes in older adults' linguistic performance that are specific to language processing. If age-related changes are the consequences of a general cognitive slowing (i.e., not specific to the language faculty), I would have no reason to expect that these changes would affect lexical analogy differently than grammatical probability. However, a general cognitive slowing cannot account for certain asymmetries observed in older adults' language processing. For example, older adults show increased effects of semantic priming (Laver, 2000), even while they exhibit increased latencies in word recognition tasks (Burke, White and Diaz, 1987).

The starting point of the current investigation was that an increase in age is associated with a continued exposure to language, more experience with words, and a larger lexicon. As a result, I expected to find that the analogical mechanism has a greater influence on older adults' acceptability ratings. The increased lexical effect proposed in this thesis could also account for some unexplained results of earlier studies. For example, Vitevitch and Sommers (2003) found that low frequency words with high frequency neighbors resulted in more TOTs in younger adults than older adults. This effect was also observed for words that had high frequency neighbors, but were in sparse neighborhoods. Vitevitch and Sommers did not account for this observed difference. However, the current findings offer an account for these data in the following way. Older adults have lived longer and have used their lexical entries for a longer time. In addition, neighborhood density and neighborhood frequency have been shown to have facilitative effects on speech production (Vitevitch and Luce, 1999). As a result, an increased effect of lexical analogy means increased facilitation in speech production, which is indeed the observed data (i.e., fewer TOTS for words with frequent neighbors for older adults) that were left unexplained in Vitevitch and Sommers (2003).

The hypothesis that well-formedness judgments in older adults would exhibit a greater influence of an analogical mechanism is supported by the results presented here. This effect is unique in that it is not the result of a general effect on aging because the grammatical component of phonotactic knowledge is not affected in the same way. I discuss this issue in the following section.

5.2.3 Is There Any Observable Change in the Effect of Grammatical Probability in Older Adults' Acceptability Ratings?

As predicted, the present study found no effect of increased age on the influence of grammatical probability. I had proposed that increased experience with one's lexicon would not have an effect on the rule-based mechanism. In other words, the grammatical system remains stable in adulthood and is not subject to later age-related changes. The proposal that grammar remains stable in adulthood is supported, first by the fact that grammatical acquisition is not a life-long process, while lexical acquisition is. Humans acquire grammar within the first several years of their lives, but continue to add lexical items throughout the life-span.

Second, grammatical acquisition is subject to a critical period, after which its acquisition is significantly limited; whereas, lexical acquisition is not subject to this developmental limit. No one has suggested an age limit for learning new words. Finally, the acquisition of grammar is the essence of language development and successful communication. While it is necessary to know a minimum number of lexical items in order to communicate (obviously it is not possible to communicate linguistically without any "words"), there is no a priori requirement on the size of the lexicon. Having a small vocabulary might result in less articulate expressions, but it would not hinder communication in the same manner that an incomplete grammar would. Therefore, it is not surprising that the analogical component of well-formedness judgments is affected by aging because it is dependent on aspects of language that can change in a life-span. In contrast, the grammatical component is not affected by aging because its constancy in healthy adults is one of its hallmarks. The results presented here are a reflection of the stability of the grammatical mechanism past any critical period for its acquisition.

Chapter 6: Experiment 3

Task-specific Effects

The experiment reported in this chapter focused on determining whether the exclusion of real words from the stimulus set would have consequences on the effect of lexical analogy. I had hypothesized that removing real words from the stimulus set would result in a reduced effect of lexical analogy on acceptability ratings. This effect would arise because judging the phonotactic well-formedness of items in a set that contained real words and non-words would mean that, simultaneously, the participants were engaged in discriminating lexical items from non-words. This discrimination would initiate a lexical search in order to access lexical units. Lexical searches that are triggered by auditory stimuli cause the activation of a word's phonological neighbors. Therefore, during a well-formedness judgment task various lexical units would be activated. Acceptability ratings would reflect this activation, hence, an increased effect of lexical analogy. To confirm this hypothesis, Experiment 3 was carried out.

Stimuli – In this experiment, the stimuli consisted of the same 84 non-words used in Experiments 1 and 2 (as given in section 3.4). Real words were excluded from the stimulus set.

Participants – 24 undergraduates (5 males and 19 females) from the University of California, Los Angeles participated in the study. The students participated in the experiment as part of their Introduction to Linguistics course. All participants were native speakers of American English, right-handed, and under 30 years of age (M=19.4, range = 18-29).

6.1 Results

The stimuli used in the analysis and the transformation of the raw ratings were the same as in Experiment 1 (discussed in section 4.1). The current experiment was designed to test whether the balance between grammatical probability and lexical analogy that was observed in Experiment 1 was sensitive to the presence of real words in the stimulus set. The first step was to determine whether, similar to the earlier experiments, acceptability ratings reflected a combination of the effects of lexical analogy and grammatical probability.

6.1.1 The Mixed-effects Model

In Experiment 1, participants saw the full stimulus set (I refer to that subject group as the "Control" group). In Experiment 3, participants saw only the non-words from the stimulus set, and real words were excluded from the presentation (I refer to this set of participants as the "Nonce-only" group). In the current model, the data comprised these two sets: Control and "Nonce-only."

This meant that the model had the following fixed factors: 1) GROUP: whether the subject was in the Control group or in the Nonce-only group, 2) GRAMMAR: the stimulus grammatical probability, and 3) ANALOGY: the stimulus analogical strength. Table 6.1 includes the parameters that were found to be significant in the model, which were determined by comparing the nested models (section 4.1.1)

Table 6.1 – Mixed-model results for both groups (Control and Nonce-only)

Linear mixed-effects model fit by maximum likelihood
number of observations: 3936, *groups*: Stimulus, 82; Subject, 48

AIC	BIC	logLik	MLdeviance	REMLdeviance
-1308	-1258	661.9	-1324	-1270

Random effects:

Groups Name	Variance	Std.Dev.
Stimulus (Intercept)	0.0099520	0.099759
Subject (Intercept)	0.0086386	0.092944
Residual	0.0382303	0.195526

Fixed effects:

	Estimate	Std. Error	t –value
(INTERCEPT)	β_0=0.7859394	0.0567764	13.843
GROUP (CONTROL)	β_3=-0.1093928	0.0321516	-3.402
ANALOGY	β_1=-0.0072866	0.0043839	-1.662
GRAMMAR	β_2=0.0507264	0.0068427	7.413
GROUPCONTROL: GRAMMAR	β_4=-0.0042333	0.0024719	-1.713
ANALOGY: GRAMMAR	β_5=-0.0021039	0.0008931	-2.356

Fitted rating = β_0*1 + β_1*SIM + β_2*PLog + β_3*Group + β_4*SIM*PLog + β_5*Group*PLog + b_i + b_j

As before, I compared nested mixed models in order to determine which effects were significant. The three-way interaction was non-significant (χ^2 (1) = 0.3887, p>0.5). In the following sections, I will discuss the significance of the other effects in the model.

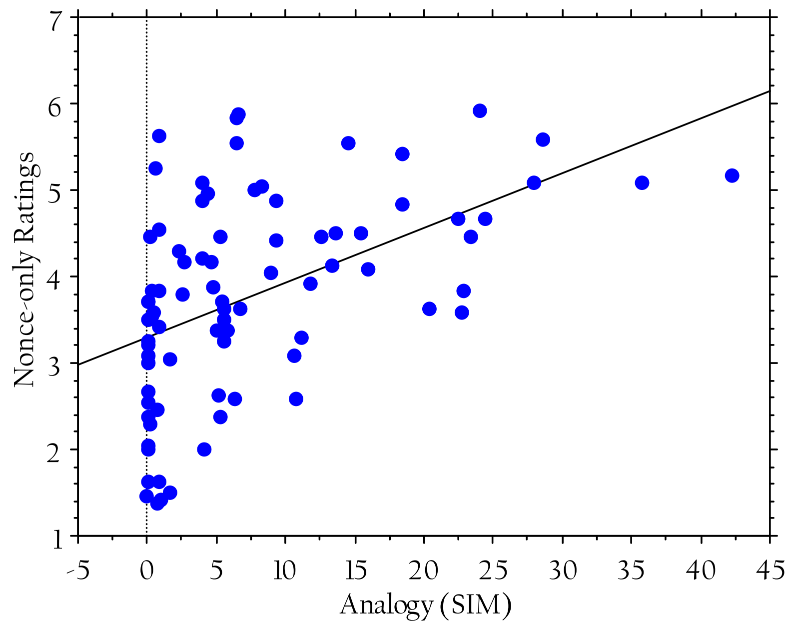
6.1.2 Does Phonotactic Knowledge Reflect a Combination of Two Mechanisms, When Real Words are Excluded?

The interaction between ANALOGY and GRAMMAR was significant (χ^2 (1) = 5.37, p<0.02), indicating that both effects made a significant contribution to the predictive power of the model. This is the same consistent result that all three experiments have shown. In order to compare the effect size of the two factors (GRAMMAR and ANALOGY) for the current set of data, the standardized scores were examined (section 4.1.1). The analysis showed that in this data set, grammar's effect (β=0.08) is larger than that of analogy (β=0.06). This pattern is different from the observed pattern for the data examined in the previous chapter (old vs. young). This change in the difference between the coefficients suggests that the balance between the two mechanisms has shifted due to the ratings from the Nonce-only subject group. Below, I will investigate the presence of an interaction between each mechanism and GROUP.

6.1.3 Is There Any Observable Change in the Effect of Lexical Analogy on Well-formedness Judgments When Real Words are Excluded?

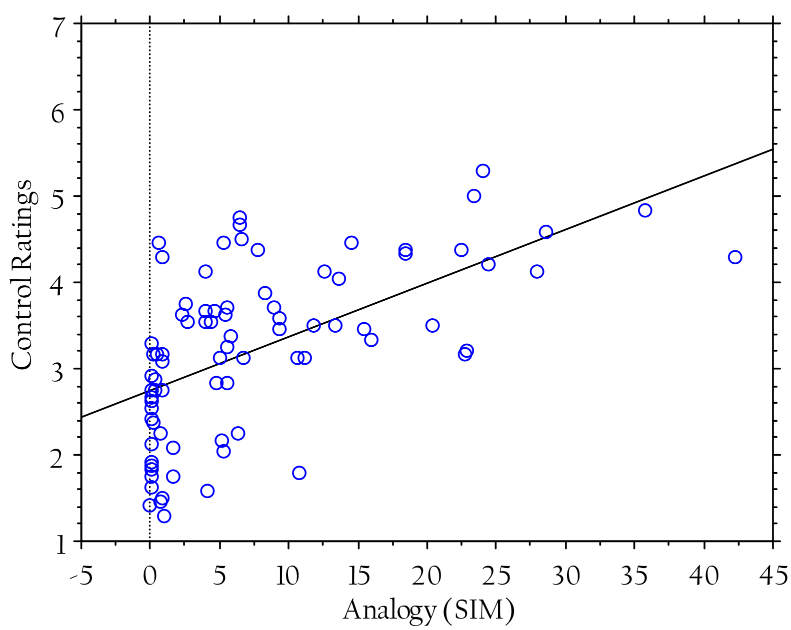
The central issue in this experiment was to determine whether the exclusion of real words from the stimulus set impacted the effect of lexical analogy on acceptability ratings. The interaction between GROUP and ANALOGY was non-significant (χ^2 (1) = 1.66, $p>0.1$). On the other hand, comparing the correlations of the analogical scores with the acceptability ratings from the Nonce-only group (Figure 6.1), and with the ratings from the Control group (Figure 6.2), indicated that the analogical scores were better correlated with the Control ratings. Throughout the thesis, "Nonce-only" data points are represented with ● in graphs, and the Control data (in Experiment 1) with ○.

Figure 6.1 – Correlation between the analogical scores (on the x-axis) and the ratings from the Nonce-only group (on the y-axis)



Y = 3.293 + .064 * X; R^2 = .246

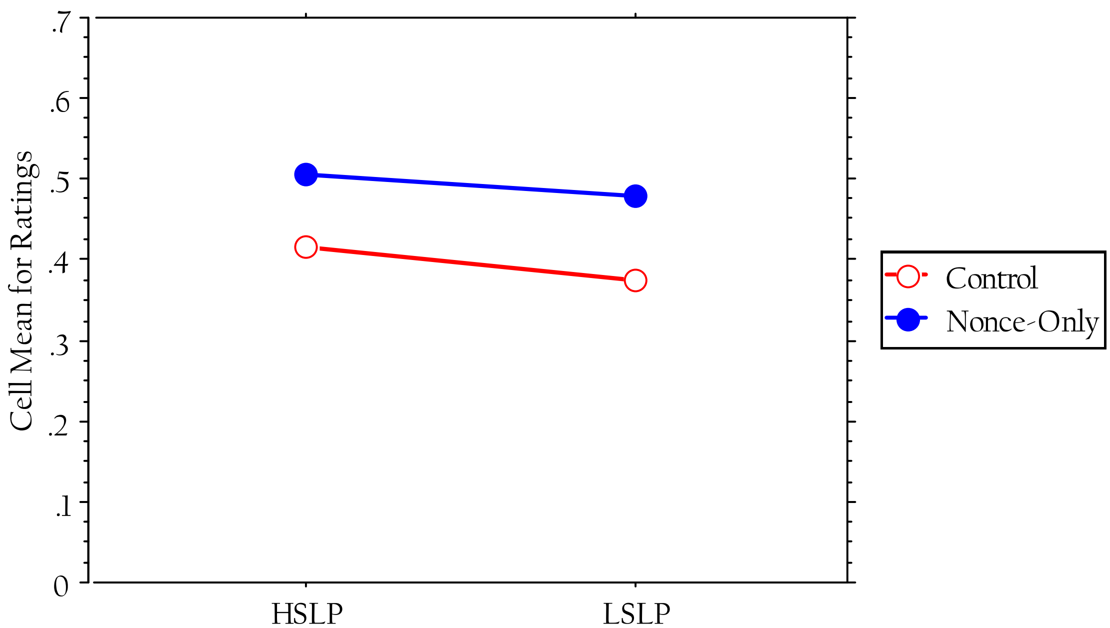
Figure 6.2 – Correlation between the analogical scores (on the x-axis) and the ratings from the Control group (on the y-axis)



Y = 2.744 + .062 * X; R^2 = .34

While the Nonce-only ratings were better correlated with analogical scores (SIMs), examining the subsets of the stimuli confirmed a lack of interaction between GROUP and ANALOGY. Within the LP (Lo-PLog) category, I compared the mean acceptability ratings of the subcategories in which the effect of analogy varied: LSLP (Lo-SIM) and HSLP (Hi-SIM). As the lack of interaction between GROUP and ANALOGY suggests, the difference in the mean ratings of the two stimuli categories did not vary greatly across the two subject groups (Figure 6.3).

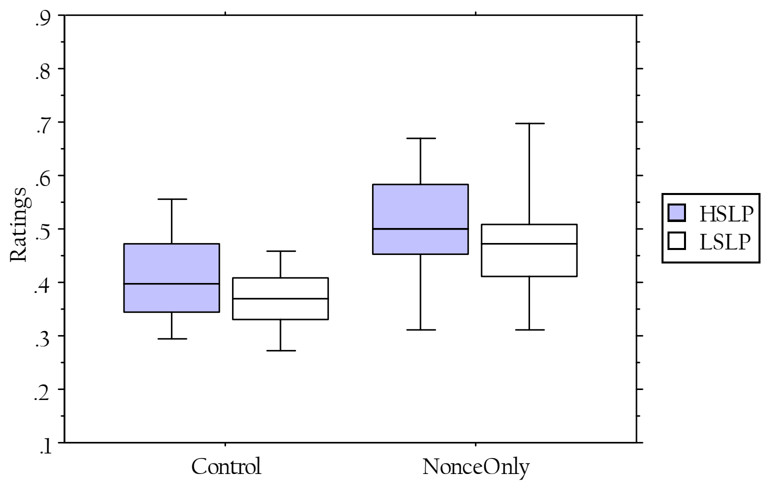
Figure 6.3 – Interaction plot for the effect of analogical strength for the Control and the Nonce-only groups: responses to HSLP (HiSIM-LoPLog) and LSLP (LoSIM-LoPLog)



The above figure also shows the main effect of GROUP. The negative coefficient for the effect of (CONTROL) GROUP indicated that the inclusion of real words had a negative effect on the ratings (i.e., the stimuli were rated lower by this group). This result was expected because the Control subjects would give the real words the highest acceptability ratings. Therefore, the non-words would be expected to be consistently rated lower than real words. This is the source of the main effect of GROUP.

In addition, it appeared that the ratings from the Nonce-only group showed more variation than the ratings from the Control group (Figure 6.4).

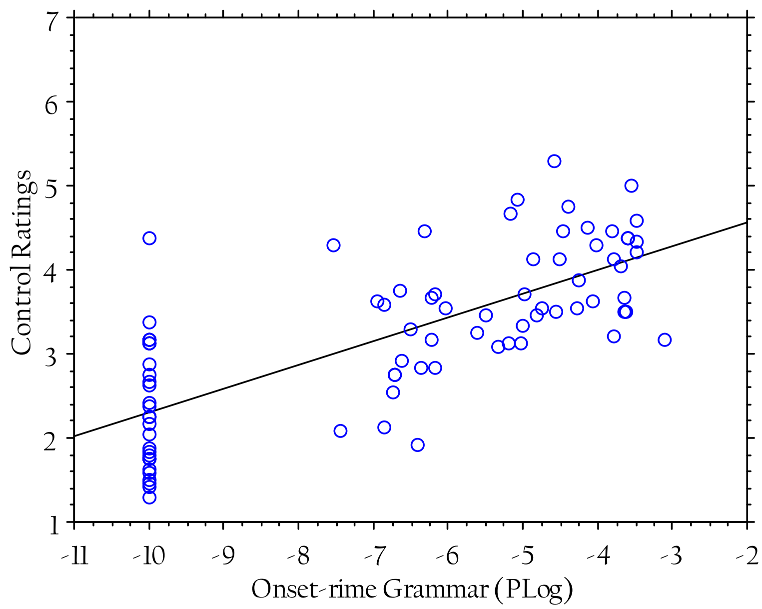
Figure 6.4 – Means ratings for LSLP and HSLP conditions for Control and the Nonce-only groups



6.1.4 Is There Any Observable Change in the Effect of Grammatical Probability When Real Words are Excluded?

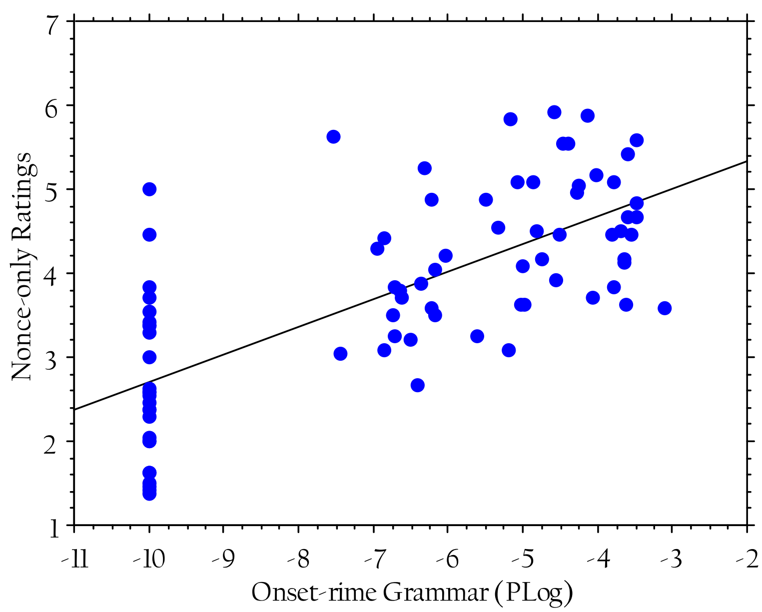
While the interaction between GROUP and ANALOGY was non-significant, there was a significant interaction between GRAMMAR and GROUP (χ^2 (1) = 4.6, $p<0.05$). Examining the correlations between grammatical probability and the ratings from the two groups indicated that grammatical probability was better correlated with the Control ratings (Figure 6.5), than the Nonce-only ratings (Figure 6.6):

Figure 6.5 – Correlation between ratings from the Control group (on the y-axis) and scores for grammatical probability (on the x-axis)



Y = 5.127 + .283 * X; R^2 = .538

Figure 6.6 – Correlation between ratings from the Nonce-only group (on the y-axis) and scores for grammatical probability (on the x-axis)



Y = 5.978 + .327 * X; R^2 = .497

This interaction between GROUP and GRAMMAR meant that the exclusion of real words from the stimulus set had a significant impact on the influence of grammatical probability on the ratings. To better understand this interaction, I examined the subset of the stimuli that fell within the LS category (i.e., items that had low SIM scores). I compared the mean ratings for the subcategories: HP (Hi-PLog), LP (Lo-PLog), and V (Violation). As Figure 6.7 illustrates, the mean ratings for the LSV items showed that the Nonce-only group was more sensitive to grammatical violations than the Control group:

Figure 6.7 – Interaction plot for Control and Nonce-only ratings: responses to LSHP (LoSIM-HiPLog), LSLP (LoSIM-LoPLog), and LSV (LoSIM-Violation)

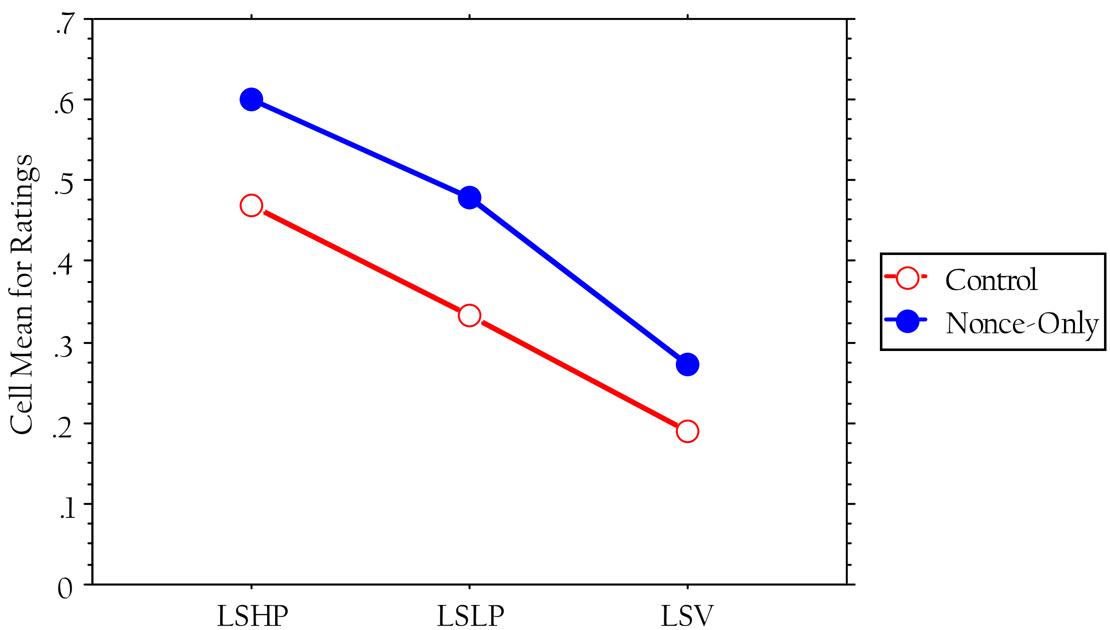
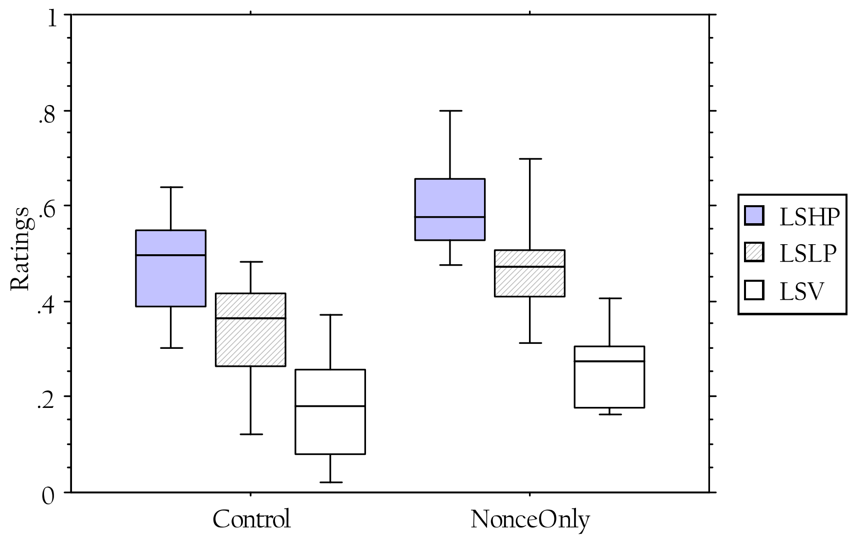


Figure 6.8 below further illustrates the main effect of GROUP (also discussed in section 6.1.3) in that the Control ratings were consistently lower than the Nonce-only ratings. We can also see that there is more variation within the LSV category for the Control group:

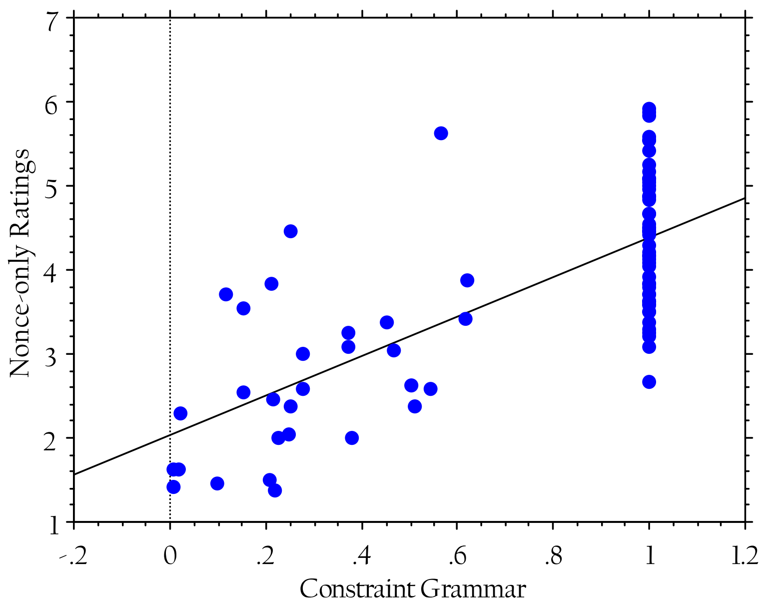
Figure 6.8 – Comparing the mean ratings for LS (Lo-SIM) conditions for Nonce-only and Control Groups (HP=Hi-PLog, LP=Lo-PLog, and V=Violation)



6.1.5 Comparison of the Onset-rime and the Constraint Grammar

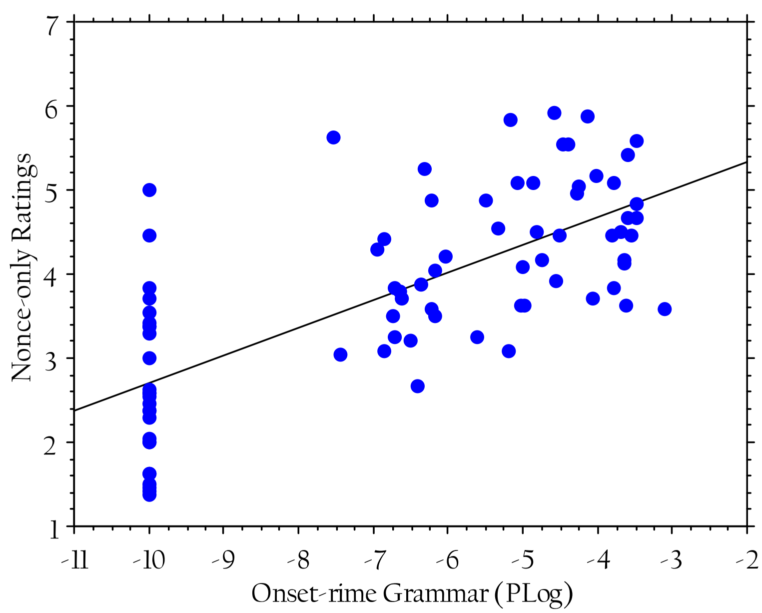
In comparing the two grammatical models, we see that the three-way interaction between GROUP and ANALOGY and CONSTRAINT-GRAMMAR is non-significant (χ^2 (1) = 0.47, p>0.4), as was the case with the Onset-rime grammar. In addition, neither the interaction between GROUP and ANALOGY (χ^2 (1) = 0.62, p>0.4), nor the interaction between GROUP and CONSTRAINT-GRAMMAR were significant (χ^2 (1) = 1.1, p>0.2). Comparing the correlations of the Nonce-only ratings with the scores from the two grammatical models showed that the Nonce-only ratings were slightly better correlated with the Constraint grammar scores (Figure 6.9), than with the Onset-rime grammar scores (Figure 6.10):

Figure 6.9 – Correlation between ratings from the Nonce-only group (on the y-axis) and scores from the Constraint grammar (on the x-axis)



Y = 2.031 + 2.346 * X; R^2 = .514

Figure 6.10 – Correlation between ratings from the Nonce-only group (on the y-axis) and scores from the Onset-rime grammar (on the x-axis)



Y = 5.978 + .327 * X; R^2 = .497

The interaction between CONSTRAINT-GRAMMAR and ANALOGY remains significant, as before, and as it does when the grammatical model is the Onset-rime grammar. The unchanged significance of the interaction confirmed that the relationship between the two mechanisms (lexical analogy and grammatical probability) were stable in the following respects: 1) both mechanisms contributed to acceptability ratings, even though lexical analogy might have had an effect only within a certain range of grammatical scores; and 2) the influence of grammatical probability on ratings appeared to be larger than that of lexical analogy.

6.2 Discussion

6.2.1 Do Phonotactic Judgments Reflect a Combination of Two Mechanisms When Real Words are Excluded?

As seen in the previous two chapters, acceptability ratings were a reflection of a combination of lexical analogy and grammatical probability. Moreover, these two effects interacted. The findings in this chapter illustrated that these observations remained true, regardless of whether or not real words were excluded from the stimulus set.

In order to determine whether the interaction between analogy and grammar had the same properties for the Nonce-only group, I considered similar subsets as before. Examining the stimuli at the top 25% range of grammatical probability indicated that the correlation of the ratings and the analogical scores was non-significant (R^2=.001). I refer to this subset as the "Top 25" subset. The correlation for Top 25 items and analogical scores was worse for the Nonce-only group, than it was for the Control group (R^2=.01). The correlation between the bottom 25% grammatically probable items and analogical scores was also non-significant (R^2=.07). I refer to this subset of stimuli as the "Bottom 25" subset.

When I considered the subset of the stimuli that contained the items with the mid range of PLog values (R^2=.14), to which I refer as the "Mid 25" subset, the correlation improved. The changes in the correlation suggest, as before, that the analogical mechanism may have the most influence when the grammatical probability values are in the mid range value. This issue is further explored in Chapter 7.

6.2.2 Is There Any Observable Change in the Effect of Lexical Analogy on Well-formedness Judgments When Real Words are Excluded?

As discussed in section 2.5, I hypothesized that the exclusion of real words from the stimulus set would result in a decreased effect of lexical analogy. Different tests have yielded different results for this experiment: 1) the analogical scores were better correlated with the Control ratings than the Nonce-only ratings (as was seen in Figure 6.1 and Figure 6.2 in section 6.1.3), 2) examining the subset of the stimuli within the mid range of grammatical scores (as discussed above in section 6.2.1) further suggested that analogy had a bigger influence on the Control ratings (r^2=.31) than the Nonce-only group (r^2=.14), 3) there was no interaction between GROUP and ANALOGY in the mixed model regardless of whether the Constraint grammar or the Onset-rime grammar was used in the analysis, and 4) the negative coefficient for the interaction between GROUP and ONSET-GRAMMAR indicated that the effect of grammar was reduced for the Control group, while correlations suggested that the Control ratings were better correlated with the grammatical scores than were the Nonce-only ratings. As a result, while the final fitted model for the data indicated a lack of interaction between GROUP and ANALOGY, there were reasons to believe that the results were inconclusive at this point.

When the full set of stimuli, and subsets of it, were examined, the Nonce-only ratings showed a decreased correlation with lexical analogy. These changes in correlation suggested that the presence of real words in the stimulus set had some effect on the ratings; however, this effect did not reach a statistical significance, as the lack of interaction in the model indicated. Therefore, the results seemed inconclusive in this respect, and required further studies.

6.2.3 Is There Any Observable Change in the Effect of Grammatical Probability When Real Words are Excluded?

The current findings suggested that the presence or absence of real words in the stimulus set affected the influence of grammatical probability. This experiment was designed in order to examine whether the exclusion of real words had any observable effects on acceptability ratings. The well-formedness judgments confirmed the influence of both mechanisms (analogy and grammatical probability), but I failed to find an interaction between GROUP and ANALOGY. Instead, the data showed that there was an interaction between GROUP and GRAMMAR. As discussed in section 6.1.1, the interaction term had a negative coefficient that indicated the reduced effect of grammatical probability for the Control group. This result was puzzling in light of the fact that, as mentioned in section 6.1.4, PLog values were better correlated with the ratings from the Control group (Figure 6.5), than with those from the Nonce-only group (Figure 6.6).

It is possible that the decreased effect of grammatical probability was a reflection of the increased influence of analogy. I had proposed that the inclusion of real words in the stimulus set would cause subjects to discriminate between real words and non-words, hence, initiating a lexical search. Recall that in section 5.2.1, I suggested that the analogical mechanism's influence was limited to items that had PLog values in the mid range of variation. Now, I considered the possibility that grammatical probability might also have the most influence within certain ranges of analogical variation. Therefore, I examined subsets of ratings from the Nonce-only and the Control group. Table 6.2 below shows the correlations of the ratings and PLog values for the subsets:

Table 6.2 – Correlations of ratings with PLog values for subsets of the stimuli, for all three groups (r^2 values are given in the cells)

	Subject Groups	"Nonce-only"	"Control"	"Old"
Subsets of stimuli based on SIM				
Top 25		.10	.08	.07
Mid 25		.56	.60	.35
Bottom 25		.29	.26	.33

While the ratings were best correlated with grammatical scores for the Mid 25 subset, the difference in correlations among the subsets do not clearly point to a direction (e.g., compare the correlations for the "old" subject group). The findings of Experiment 3 are inconclusive as to how the exclusion of real words affects the mechanism involved in phonotactic well-formedness judgments. While I have shown that this change in the stimulus set had an effect that was reflected in the interaction between GROUP and GRAMMAR, I have not been able to provide a conclusive interpretation of the data. A better understanding of the effect of including real words in the well-formedness judgment task requires further investigation. I discuss some suggestions for further studies in section 8.2.

Chapter 7: Is Analogy Used in "Hard" Cases?

One issue that has remained unexplained thus far is the nature of the interaction between ANALOGY and GRAMMAR. There are two features regarding this interaction that have yet to be explained. First, an interpretation of how the two mechanisms interact is still needed. The second issue concerns the negative coefficient for the interaction term. These characteristics have been found across the three experiments; therefore, they seem to be consistent properties of the two mechanisms, and how they interact in phonotactic well-formedness judgments. As I suggested in the previous chapters, lexical analogy may have had less influence on the stimuli that had extreme values for grammatical probability. In this chapter, I offer further analysis of the interaction between the analogical and the grammatical mechanisms.

7.1 Investigation of the Subsets of the Data

The negative coefficient for the interaction between ANALOGY and GRAMMAR implied that, at least for some of the stimuli, ANALOGY may have had a negative effect on acceptability ratings. Therefore, I explored two possibilities: 1) whether lexical analogy has a negative effect on acceptability ratings for forms with extreme PLog values; or 2) whether the analogical mechanism is most influential when PLog values are not at the extreme. To this end, I considered subsets of the data, based on the stimuli's PLog values. The "Top 25" subset included items whose PLog values were in the Top 25%

range. The "Mid 25" subset comprised stimuli with the mid range of PLog values (37 to 62 percentile), and the "Bottom 25" subset contained those at the bottom 25% of the probability scale. For these sets, correlations between the mean ratings and analogy scores (SIMs) were calculated for the three subject groups.

In Table 7.1, I have provided a summary of the correlations. As the table shows, the analogy scores are best correlated with the ratings of the Mid 25 subset (shaded box). The worst correlation was yielded by the Top 25 subset. This is the case for all three subject groups.

Table 7.1 – Correlations with analogical scores with subsets of the stimuli, for all three groups (R^2 values are given in the cells)

Subject Groups	"Nonce-only"	"Control"	"Old"
Subsets of stimuli based on PLog			
Top 25	.001	.01	.06
Mid 25	.14	.31	.26
Bottom 25	.07	.16	.14

The correlations suggested that it was in the Mid 25 subset that lexical analogy had the most influence on acceptability of a form. This is plausible considering that if a word contains an illegal sequence (e.g., *bnick* which would be in the Bottom 25 subset), the existence of lexical neighbors (e.g., *nick* or *brick*) would likely not significantly improve the acceptability of the form. Also, in none of the subsets was analogy negatively correlated with the ratings, which excludes the possibility that analogy had a negative effect for any subsets. Therefore, I expected that analogy would have the most influence on well-formedness judgments when grammatical probability was least decisive about a form's well-formedness, i.e., the form was neither highly acceptable, nor highly unacceptable. In order to test this proposal, another analysis using a mixed-effects model (section 4.1.1) was performed.

7.2 Analysis

Consider the three ranges of PLog that have been presented. When the novel forms contained a violation (i.e., grammatical scores were at the lowest), the prediction was that having high analogical scores would not significantly influence well-formedness judgments. This also would have been the case when the grammatical probability scores were at the highest. On the other hand, when PLogs were not decisive (i.e., for items that were "hard" to judge), lexical analogy was used as an additional strategy in determining the acceptability of a form. This prediction was also in line with the previous conclusion that grammar had a larger impact than analogy on acceptability ratings.

If analogy had the most influence on an item when its PLog value was mid range, one needed to determine a coefficient for analogy to reduce its influence as PLog values got smaller or larger than the mean. The formulation of this coefficient is best understood by starting with identifying a mathematical formulation of an item's PLog distance from the mean. The first step is identifying a form's PLog ($PLog_i$). The next step is to calculate how far $PLog_i$ was from the mean ($\overline{PLog}$), by deducting its value from the mean: $PLog_i - \overline{PLog}$

The third step is constructing a coefficient that would result in the same value, regardless of whether $PLog_i$ has a value lower than the mean, or higher. In other words, analogy is meant to have the same effect as $PLog_i$ moved away in either direction. To this end, a squared value of the distance is used: $(PLog_i - \overline{PLog})^2$

The fourth step is to use the coefficient to *reduce* the influence of analogy as $PLog_i$ moved away from the mean. Therefore, an inverse function is used:

$$\frac{1}{(PLog_i - \overline{PLog})^2}$$

Finally, in order to avoid having a zero in the denominator, the value "1" is added to the denominator. The resulting parameter is termed "Gated.SIM":

$$Gated.SIM = SIM(\frac{1}{1 + (PLog_i - \overline{PLog})^2})$$

Note that when $PLog_i$ is exactly at the mean ($PLog_i = \overline{PLog}$), the above coefficient will be "1", as shown below:

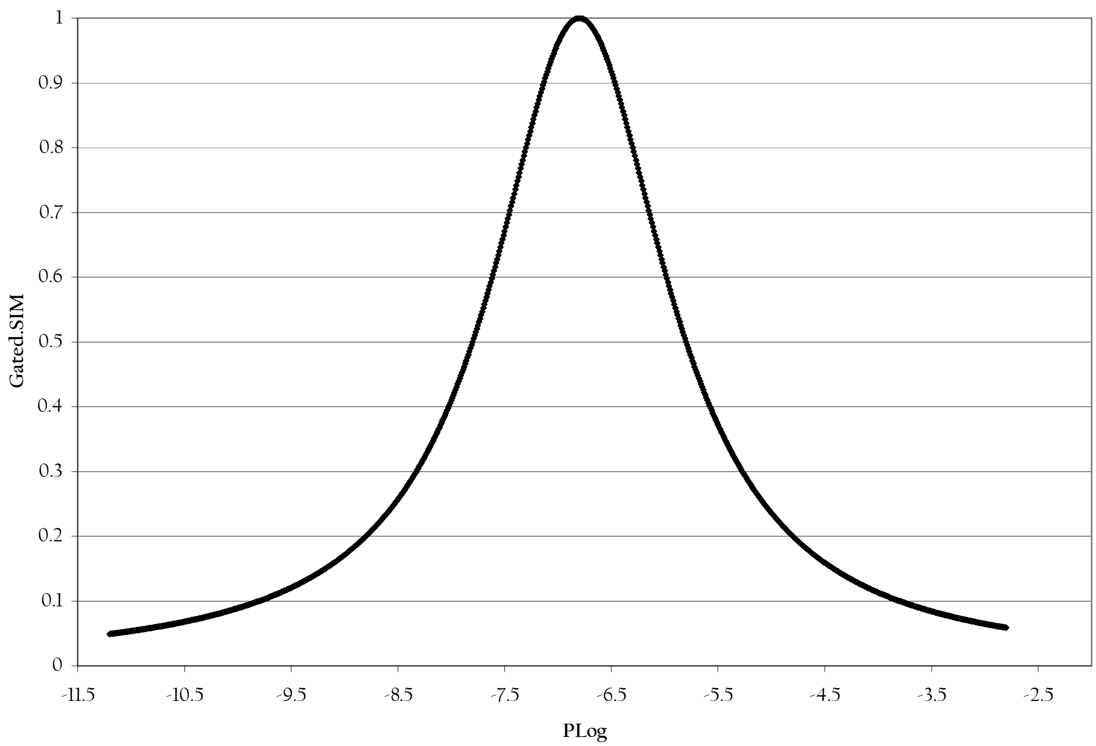
$$Gated.SIM = SIM(\frac{1}{1 + (PLog_i - \overline{PLog})^2})$$

$$Gated.SIM = SIM(\frac{1}{1 + (0)^2})$$

$$Gated.SIM = SIM(1) = SIM$$

The above coefficient results in analogy having full influence when $PLog_i$ is at the mean, and having less and less influence as $PLog_i$ gets farther and farther from the mean. I refer to this mediated analogy as "Gated-analogy." A graph of the function that the coefficient represents is provided below (Figure 7.1). In the following section, an analysis of the data using a mixed-effects model, in which Gated-analogy (and not analogy) is a fixed effect, is presented.

Figure 7.1 – The function assumed for Gated.SIM



7.2.1 A Mixed-effects Model with Gated-analogy

The hypothesis to be tested was whether analogy had the most influence on acceptability of the stimuli with the mid range of PLog values. Gated-analogy was formulated to represent this interaction. If this formulation were correct, the results of the current mixed-model analysis would reflect two changes, compared to ones presented in the previous chapters. First, Gated-analogy would have a *positive* coefficient because it, in fact, would have a positive effect in its region of influence. Second, the interaction between ANALOGY and GRAMMAR would no longer be significant, because the coefficient that was introduced to create Gated-analogy (section 7.2) contained the correct characterization of the interaction between ANALOGY and GRAMMAR.

To test the hypothesis, the data set that contained results from Experiments 1 and 2 (data from young and old groups) were examined. All effects used in the current model were the same as discussed in section 5.1.5, except that instead of ANALOGY, the expression GATED-ANALOGY (as defined above) was one of the fixed effects. The summary of the model is provided in Table 7.2.

Table 7.2 – Mixed-model results for using Gated-analogy

Linear mixed-effects model fit by maximum likelihood
number of observations: 3936, *groups*: Stimulus, 82; Subject, 48

AIC	BIC	logLik	MLdeviance	REMLdeviance
-1353	-1309	683.3	-1367	-1328

Random effects:

Groups Name	Variance	Std.Dev.
Stimulus (Intercept)	0.0074707	0.08643
Subject (Intercept)	0.0128727	0.11345
Residual	0.0378344	0.19451

Fixed effects:

	Estimate	Std. Error	t-value
(INTERCEPT)	β_0 = 0.558738	0.042740	13.073
AGEYOUNG (AGE)	β_3= 0.060264	0.033693	1.789
GATED-ANALOGY	β_1 = 0.024863	0.006094	4.080
GRAMMAR	β_2 = 0.037511	0.004406	8.513
AGEYOUNG:GATED-ANALOGY	β_4=-0.010884	0.003276	-3.322

rating = β_0*1+β_1*SIM*(1/(1+(PLog-Mean)^2) + β_2*PLog + β_3*Age + β_4*Age*SIM*(1/(1+(PLog-Mean)^2) + b_i + b_j

I used the same analysis of comparing the nested mixed models in order to determine which parameters were significantly contributing to how well the model fit the data. Similar to the model with ANALOGY as a fixed effect, the three way interaction (AGE and GATED-ANALOGY and GRAMMAR) was not significant (χ^2 (1) = .61, p>0.4). In addition, as before, AGE and GRAMMAR did not interact (χ^2 (1) = 1.25, p>0.2). In these respects, the two models (using analogy vs. Gated-analogy) behaved the same.

7.2.2 Does Gated-analogy Reflect the Interaction between Analogy and Grammar?

Recall that across all three experiments the interaction between ANALOGY and GRAMMAR had been significant (sections 4.1.1, 5.1.2, and 6.1.2). After normalizing the analogical and grammatical scores, the earlier model indicated that the effect of ANALOGY (β=0.06) was only slightly smaller than GRAMMAR (β=0.07) for the data in Experiment 1.

In the current model, after normalizing the analogical and grammatical scores, the interaction between GATED-ANALOGY and GRAMMAR was no longer significant (χ^2 (1) = 0.17, p>0.6), indicating that the coefficient used in defining Gated-analogy may be the correct characterization of the interaction. In addition to removing the interaction as a significant effect from the model, for Experiment 1, using Gated-analogy resulted in a larger effect of GRAMMAR (β=0.097), which was now about twice as large as that of GATED-ANALOGY (β=0.045).

7.2.3 Is there an Interaction between Age and Gated-analogy?

The interaction between age group and analogical scores, which was now represented by an interaction between AGE and GATED-ANALOGY, was still highly significant in the current model (χ^2 (1) = 11.02, p<0.0001). Moreover, the effect was in the same direction, i.e., the "old" group showed an increase in the influence of GATED-ANALOGY on acceptability ratings.

Another advantage of using Gated-analogy was that the effects that had been marginal in earlier analyses were shown more clearly now. Recall that including the "old" data in the earlier analysis had only slightly affected the balance between the two fixed effects, such that ANALOGY (β=0.08) had a slightly larger effect than GRAMMAR (β=0.06). In using Gated-analogy, the shift of balance between the two effects was shown more clearly: GATED-ANALOGY (β=0.14) had an effect more than one and a half times larger than GRAMMAR (β=0.085), when ratings from "old" subject group were included in the analysis.

7.2.4 Is there a Task-specific Effect?

A similar analysis for the data in Experiments 1 and 3 (Control and Nonce-only groups) remained inconclusive, as it had been when using ANALOGY as a fixed effect (sections 6.2.2 and 6.2.3). In using GATED-ANALOGY instead of ANALOGY as an effect in the model for the data that combined Control and Nonce-only ratings, all interaction terms could be removed from the model: 1) the interaction between GROUP and GATED-ANALOGY was non-significant (χ^2 (1) = .018, p>0.8), and 2) the interaction between GROUP and GRAMMAR was non-significant (χ^2 (1) = 2.93, p>0.08). Moreover, the effect of GATED-ANALOGY could be removed from the model (χ^2 (1) = 3.71, p>0.05). While this result suggests that including the Nonce-only data removed the effect of the analogical mechanism, the lack of interaction between GROUP and ANALOGY is puzzling, and requires further studies to determine how the presence of the real words affected well-formedness judgments.

7.3 Discussion

Across all three experiments in the earlier chapters, it was found that analyzing analogy and grammar required an interaction term with a negative coefficient. In an effort to understand this interaction, three subsets of stimuli were examined. It was found that analogy best correlated with the ratings for stimuli with the mid-range PLog values. I provided a mathematical formulation of this interaction (referred to as "Gated-analogy") and reanalyzed the data. The interaction between GATED-ANALOGY and GRAMMAR was found to be no longer significant. These results suggest that Gated-analogy represents a close approximation of the interaction between ANALOGY and GRAMMAR.

7.3.1 Does Grammar have more Influence on Acceptability Ratings than Analogy?

As discussed earlier (section 4.2), Bailey and Hahn (2001) claimed that the effects of the two mechanisms (analogy and grammar) on well-formedness judgments were such that lexical analogy had the greater effect. Except the data from the "old" group, the other experiments reported in this thesis do not support that view. The question is why the two studies would differ so vastly in the size of the effect of the two mechanisms. A possibility was that the stimuli from Bailey and Hahn did not cover a full range of PLog values. Another related possibility was that the stimuli were mostly "hard" to judge because their PLog values were in the mid range.

7.3.1.1 How to calculate the Mean PLog Value for Gating Analagy?

The above-mentioned possibilities about the differences in the two studies raised an issue. What is the mean range of PLog to be used in "gating" analogy? In the earlier section of this chapter, I reported an analysis in which Gated-analogy operated based on the mean of PLog, calculated based on the set of stimuli used in the current study (M=-6.7). In order to investigate whether the specific values in this study had a special property, I examined the range of values in the "stimulus" corpus (section 3.4). In that corpus, the median was -5.66, and the mean was -5.67. The data was reanalyzed based on the mean from the stimulus corpus. This change did not affect the analysis because: 1) interaction between AGE and GATED-ANALOGY and GRAMMAR remained non-significant (χ^2 (1) = .101, p>0.7); 2) the interaction between AGE and GRAMMAR was non-significant (χ^2 (1) = .127, p>0. 2); and 3) the interaction between GATED-ANALOGY and GRAMMAR also was non-significant (χ^2 (1) = 0.055, p>0.8). The interaction between AGE and GATED-ANALOGY remained significant (χ^2 (1) = 7.36, p<0.007), indicating that lexical analogy had an increased effect on the "old" group. The summary of this analysis is presented in Table 7.3.

Table 7.3 - Mixed-model results for using Gated-analogy gated by the mean PLog from the stimulus corpus

Linear mixed-effects model fit by maximum likelihood
number of observations: 3936, *groups*: Stimulus, 82; Subject, 48

AIC	BIC	logLik	MLdeviance	REMLdeviance
-1348	-1304	680.8	-1362	-1320

Random effects:

Groups Name	Variance	Std.Dev.
Stimulus (Intercept)	0.0076151	0.08726
Subject (Intercept)	0.0128732	0.11346
Residual	0.0378705	0.19603

Fixed effects:

	Estimate	Std. Error	t-value
(INTERCEPT)	β_0 =0.559171	0.044150	12.665
AGEYOUNG (AGE)	β_3=0.055367	0.033598	1.648
GATED-ANALOGY	β_1 =0.010787	0.002955	3.650
GRAMMAR	β_2 =0.036498	0.004666	7.822
AGEYOUNG:GATED-ANALOGY	β_4=-0.004100	0.001507	-2.721

rating = β_0*1+β_1*SIM*(1/(1+(PLog-Mean)^2) + β_2*PLog + β_3*Age + β_4*Age*SIM*(1/(1+(PLog-Mean)^2) + b_i + b_j

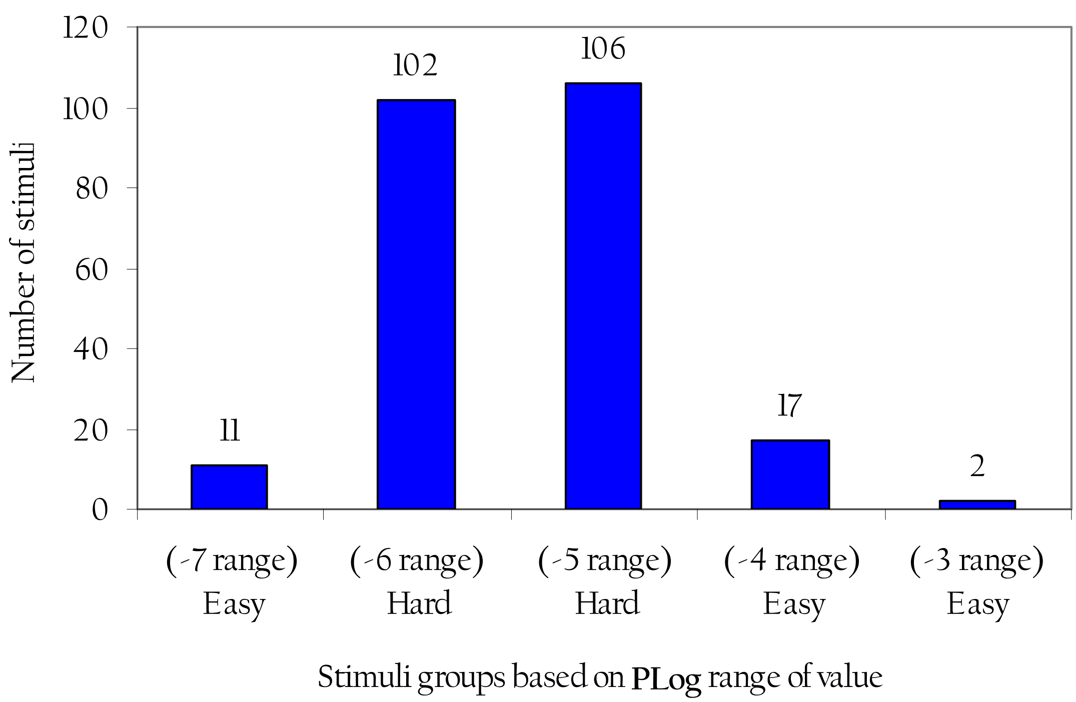
Based on the fact that the mean calculated based on the present study's stimulus set (M= -6.7) and the mean calculated based on the stimulus corpus (M= -5.67) could function as the value used for Gated-analogy, In the following section, I have labeled items with PLog values that range from -5 to -6.99 as "hard" items.

7.3.1.2 The Stimuli from Bailey and Hahn (2001) were "hard"

Let us return to the method of stimulus selection employed by Bailey and Hahn. Their method of stimulus selection (section 2.2.3) had two consequences. First, there were no forms in the stimulus set that contained an outright phonotactic violation;

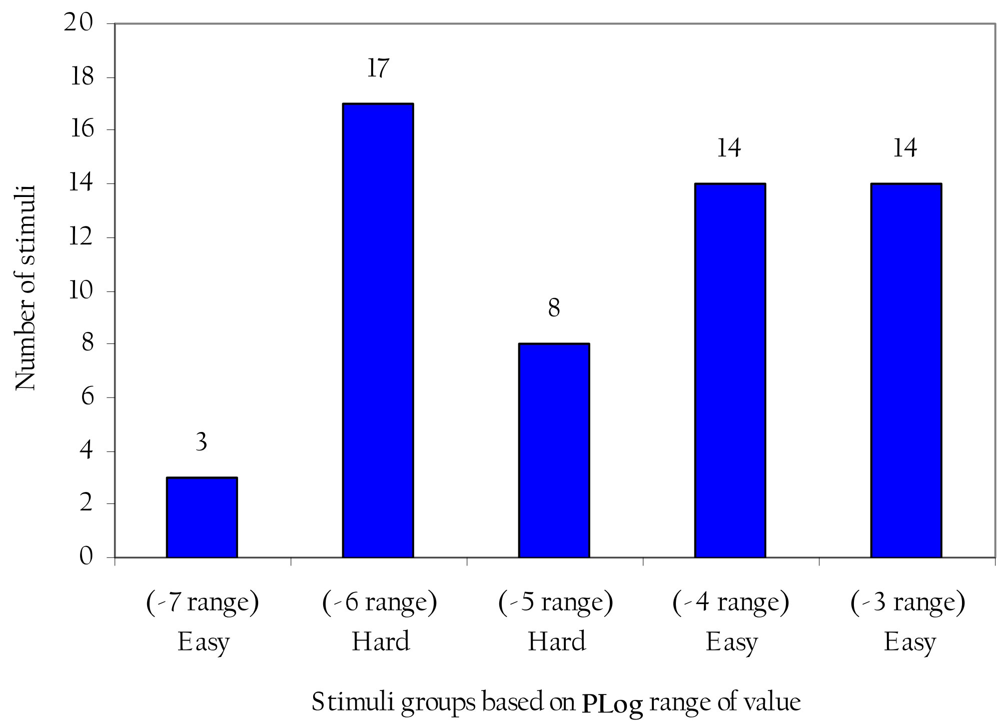
therefore, one set of forms for which the grammatical mechanism would be highly decisive was eliminated. Second, selecting the forms based on having lexical neighbors resulted in a stimulus set that was biased towards an inflated analogical effect. In fact, examining their stimuli, it appeared that most forms were "hard" to judge, because they did not have any extreme PLog values. In order to determine whether that was the case, I calculated PLog values for all the stimuli reported in Bailey and Hahn (2001). I grouped the stimuli based on their PLog values. Figure 7.2 shows the distribution of the stimuli. As the graph illustrates, most of the stimuli fell into the "hard" category, which allowed for the most influence from analogy.

Figure 7.2 – The distribution of stimuli (from Bailey and Hahn (2001)), grouped based on PLog values



The same grouping was done for the stimuli used in this thesis, in order to determine whether the stimuli used in the three experiments had, indeed, a different distribution with respect to their PLog values, which would result in a more balanced influence of analogy.

Figure 7.3 – The distribution of stimuli from the present study, grouped based on PLog values



As can be seen from the graphs, the range of PLog values for the present study's stimuli covers a broader range than those used in Bailey and Hahn's study. This difference further confirms that their results were dues to the stimuli's properties with respect to their mid-range PLog values. That range resulted in an inflated effect of analogy.

7.3.2 Summary

In summary, I have made the following proposal: lexical analogy is a strategy that speakers use when the grammatical mechanism does not have a strong preference for accepting or rejecting a form. This suggests that whether a form is "easy" or "hard" is determined by the grammatical mechanism. "Easy" cases are those for which the grammar needs no help in judging their well-formedness, because its judgments are strong for those cases. The grammatical mechanism's judgment is strongest when PLog is at the extreme high, or extreme low. "Hard" cases, where the grammar does not have a strong judgment about a form, are those that invoke the analogical mechanism. In summary, well-formedness judgments are an intricate combination of the two mechanisms.

Chapter 8: Summary

8.1 Summary

This thesis investigated the nature of phonotactic well-formedness judgments. Its goal was to test three hypotheses:

Hypothesis 1. Phonotactic well-formedness judgments reflect a combination of two mechanisms: a grammatical mechanism and an analogical one.

Hypothesis 2. The analogical mechanism shows an increased influence on acceptability ratings in older adults.

Hypothesis 3. The analogical mechanism shows a decreased influence on acceptability ratings when real words are excluded from the stimulus set.

To this end, real word and non-word stimuli were selected to monitor the independent effects of lexical analogy and grammatical probability, by varying along the two dimensions. As described in Chapter 3, the effects of the two mechanisms (analogy and grammar) were measured based on two formal models: the stochastic phonological grammar of Coleman and Pierrehumbert (1997) and the Generalized Neighborhood Model of Bailey and Hahn (2001). In that chapter, I also discussed the experimental design of the studies, in which subjects rated the stimuli's acceptability on a scale of "1" (impossible English word) to "7" (fine English word).

Three experiments and their results were reported. The investigation of Hypothesis (1) was discussed in Chapter 4. The data demonstrated that phonotactic well-formedness judgments are a reflection of a combination of lexical analogy and grammatical probability. In addition, the two effects were shown to interact, such that the analogical mechanism has the most influence when the grammatical mechanism's predicted acceptability ratings are in the middle rang, i.e., not decisive. Therefore, my first hypothesis was confirmed.

Hypothesis (2) was tested in Chapter 5, where the results of my second experiment, using older adults, were reported. These results replicated the combinatory effect of the two mechanisms in well-formedness judgments found in Experiment 1. The main finding of Experiment 2, however, was that lexical analogy has more influence on the acceptability ratings of older adults, confirming my second hypothesis.

In Chapter 6, Hypothesis (3) was examined by reporting the results of Experiment 3, in which real words were removed from the stimulus set. The results of this experiment, however, were inconclusive with respect to confirming or refuting Hypothesis (3). The results showed that the presence or absence of real words does impact acceptability ratings. It is the exact nature of this effect that was not revealed by the experiment because different tests yielded different results.

8.1.1 The Effects of Grammar vs. Analogy

Taken together, the results of all three experiments demonstrate that there are two interactive mechanisms at work in phonotactic well-formedness judgments. While the results corroborate the conclusion by Bailey and Hahn (2001), that there are two mechanisms at work, they make three additional contributions. First, I developed a method for testing the independent effects of each mechanism by implementing formal models for each mechanism and carefully selecting the stimuli to cover a full range of variation with respect to both mechanisms. This allowed for extending Hypothesis (1) to all novel forms, especially those for which the two mechanisms would disagree on the degree of acceptability.

By doing this, moreover, I was able to demonstrate that the influence of grammatical probability is greater than that of lexical analogy. The strength of grammatical probability seen in my data might seem in conflict with the results reported by Bailey and Hahn, who found a greater effect of lexical analogy than phonotactic probability. However, the stimulus set used in their study comprised items that always had real word neighbors. This method of stimulus selection causes the effect of lexical analogy to be magnified. This issue was explored in Chapter 7.

A third contribution of this dissertation is the finding that lexical analogy and grammatical probability interact for all subject groups. To account for this result, I proposed that lexical analogy is most influential for stimuli that cover the middle range of values for grammatical probability scores. For items with the most extreme values for grammatical probability, the influence of lexical analogy is significantly reduced. I proposed a coefficient for lexical analogy to reflect this interaction, and referred to the new expression as "Gated-analogy." A mixed model analysis showed that Gated-analogy does not interact with grammatical probability, suggesting that Gated-analogy may be the correct characterization of the interaction between the two mechanisms (grammar and analogy).

Recall that, in Chapter 2, I reported on the conflicting accounts of mechanisms that underlie phonotactic well-formedness. Some of these conflicts arose because stimuli used in previous studies were not controlled with respect to *both* mechanisms. However, it is also plausible that some of the studies that failed to find an effect of lexical analogy did so because their stimuli were selected by dividing forms into two classes of grammatical probability (e.g., high vs. low). If the proposal on the limited range of analogical influence is correct, one would predict that stimuli partitioned into high versus low probability would not show a significant influence of lexical analogy. On the other hand, selecting stimuli based on only extreme values of analogical scores would show an inflated influence of analogy.

In addition, the validity of the proposed account is boosted by the statistical technique that was used in data analysis. This technique, a mixed-effects model that analyzes the individual ratings of each subject for each stimulus, avoids some commonly practiced methods that are disadvantageous in psycholinguistic studies. For example, most analyses discard the precise numerical information of the stimuli; instead, they use a factorial contrast (e.g., high vs. low, long vs. short). While this type of factorization can be useful in some situations, it can also reduce the power of the analysis in generalizing to all items (Baayen, 2007). Also, most data sets are routinely averaged across subjects before analysis, which reduces the power of the analysis to generalize to all subjects. However, the ability to generalize to all speakers and across all linguistic input is a goal often at the heart of most psycholinguistic studies. I believe it is the combination of the carefully selected stimuli and the use of a mixed-effects model that has allowed the current study to report novel findings on phonotactic well-formedness judgments.

8.1.2 The Age-specific Effect on Analogy

Most studies of the effects of aging report negative effects, such as decrements in performance accuracy or speed. However, another contribution of this thesis has been to demonstrate that some changes in linguistic knowledge of older adults are positive changes, such as a continually growing lexicon. Older adults have a larger vocabulary and more experience with the words they know. As a result, I had hypothesized that they would show more sensitivity to lexical properties of forms. This meant that one would expect to find reflections of positive changes associated with an increase in age (e.g., an increased sensitivity to neighborhood density). As shown in section 6.1.1, phonotactic well-formedness judgments in older adults (over 65 years old) reflect an increased influence of lexical analogy, indicating that they are more sensitive than younger speakers to a form's phonological neighborhood in rating its phonotactic well-formedness.

Exploring this view shows that this interpretation could be extended to account for puzzling data observed in other studies, such as Vitevitch and Sommers (2003). In section 5.1.3, the details of how the findings in this thesis could be extended were discussed. In summary, the increased effect of lexical analogy in older adults also explains why they show fewer TOTs than younger adults for words that had high frequency neighbors, when the target words were low frequency words, or were from sparse lexical neighborhoods (Vitevitch and Sommers, 2003).

Importantly, the increased influence of lexical analogy on phonotactic well-formedness is accompanied by no change in the effect of grammatical probability in older adults. The unchanged influence of the grammatical mechanism indicates that at least some aspects of the grammatical system remain stable as people age.

8.1.3 The Task-specific Effect on Grammar

Investigating Hypothesis (3) resulted in inconclusive findings. While in the mixed model the presence of real words did not show an interaction with the influence of lexical analogy, the changes in the correlations between ratings and analogical scores were in the predicted direction (i.e., an increased influence of lexical analogy emerged as the result of the presence of lexical items in the stimulus set).

The hypothesis hinged on the assumption that the inclusion of real words in the stimulus set would result in two simultaneous tasks: providing well-formedness judgments and discriminating words from non-words. The findings indicated that the effects of lexical analogy were not a by-product of a lexical search, and perhaps not *easily* affected by it, either.

On the other hand, I found that the presence of real words interacts with the influence of grammatical probability, such that it has a smaller effect when real words are included in the stimulus set. One possible explanation is that the grammatical mechanism may have a limited range of influence, similar to the one proposed for the analogical mechanism. However, the data regarding this explanation were also inconclusive. It is possible that this interaction is partially the result of using a rudimentary grammatical model in the study. This possibility is further explored in section 8.2.2.

8.2 Future Research

This thesis has provided evidence for some novel effects in phonotactic well-formedness judgments. However, it has also not only left the interpretation of some data unresolved, but has raised new questions. In the following sections, I explore some of these issues.

8.2.1 An Improved Analogical Model

The analogical model used in the study has several flaws. First, it only considers the similarity between elements that occur in the same position. However, it has been shown that there are other factors that contribute to similarity judgments in people. For example, different prosodic positions may have unequal roles in computing similarity (e.g., onsets may be more prominent than codas). Second, the model computes the same cost for deletion and insertion, and computes only substitution based on segment similarity. However, it is possible that in some instances substitution is more costly than insertion or deletion. For example, speakers are reported to judge /stɪp/ closer to /stɪmp/ than /stɪm/ (Bailey and Hahn, 2004). Third, it treats all segments as equal with respect to deletion and insertion. Modifications of these aspects of the model could improve the model's performance.

Finally, as seen in section 4.1.2, the analogical model's performance on making distinctions among items that have low similarity scores could be improved. As seen in Figures Figure 4.2 and Figure 5.1, items with low similarity are clustered in one corner of the scale. Ideally, one would want to develop an analogical model that incorporates all that is known about metrics of similarity (e.g., GNM (Bailey and Hahn, 2001)) and dissimilarity (e.g., PLASS (Bailey and Hahn, 2004)).

8.2.2 An Improved Grammatical Model

The predictions of two grammatical models were compared with respect to the acceptability ratings. While the scores from both models are correlated with well-formedness judgments and contribute in a similar fashion to the mixed model's predictive power, each model's predictions are unable to account for a subset of the stimuli. The Onset-rime grammar (discussed in section 3.1) does not make a distinction among "illegal" forms (see Figure 4.4). Recall that the illegal forms (i.e., items with probability of zero) were selected such that their legal constituents were all highly probable. Therefore, the distinction that the subjects made among these forms could not be due to the difference in their legal constituent's probability. This grammatical model, however, cannot account for distinctions within this class of stimuli. In addition, because this model bases its scores on probabilities that are generalized over the lexicon, distinguishing its effects from lexical effects is always a challenge.

In contrast, the Constraint grammar (section 2.2.2) makes distinctions among illegal forms, but it does not make fine enough distinctions among legal forms (see Figure 4.3). The reason for this flaw seems to be that legal forms do not violate any constraints;

therefore, the model does not embody any means to predict distinctions within the class of possible words.

Ideally, one would develop a grammatical model that succeeds where these two models fail. Such a model would make distinctions within the legal set of forms, such that it reflects native speaker intuitions regarding which forms are more typical instances of a word in their language, and would also make distinctions within illegal forms, such that it reflects native speaker intuitions regarding how violating some constraints are more costly than others.

8.2.3 Investigating Older Adults' Responses to a Nonce-only Stimulus Set

The results of Experiments 2 and 3 call for an additional experiment in which older adults respond to a stimulus set consisting of only non-words (i.e., replicating Experiment 3 with older adults). Ratings for a Nonce-only stimulus set by older adults could shed light on the effects of the presence of real words on acceptability ratings. For this proposed experiment, there are three possible outcomes: 1) older adults will show the same effect as younger adults (i.e., an interaction between grammar and group (Nonce-only and Control); 2) older adults will be more sensitive to lexical effects and will show a decreased influence of lexical analogy when real words are excluded from the stimulus set; 3) older adults will show no effect of the change in the stimulus set.

It is not clear whether outcomes (1) or (3) would contribute to a more conclusive interpretation of the results. However, I predict that outcome (2) would be the result, which would suggest that the current findings on young adults are inconclusive because the change in the influence of lexical analogy in younger adults is too small to be observable in the statistical analysis. In other words, older adults might be a better population to test Hypothesis 3, because their increased lexical sensitivity allows for measuring smaller changes in the influence of the analogical mechanism.

8.2.4 Replicating Experiment 3 Using the Mid 25 Subset

In Chapter 7, I proposed that the influence of lexical analogy is limited to the items with the mid range of PLog (grammatical probability score) values, referred to as "hard" cases. Knowing this, let us consider the possibility that the presence of real words in the stimulus set results in a shift in the balance between the grammatical and the analogical mechanisms that is too small to be detected for items with extreme PLog values. This suggests that, in testing any changes in the effect of lexical analogy, one should use stimuli that do not contain any items with extreme PLog values, as Bailey and Hahn (2001) did. The logic behind this suggestion is that any fluctuation in the influence of the analogical mechanism could be better monitored in the region where it has been shown to most affect acceptability ratings.

One suggested strategy for future study would be to select stimuli that exclude items with extreme PLog values. The selected items, however, should cover a range of variation with respect to their SIM (analogical score) values. Real words should also be selected that had the same properties with respect to their SIM and PLog values. Using these stimuli, Experiment 1 and 3 could be repeated (as Experiments 1b and 3b). Based on the interaction between grammar and analogy, the hypothesis is that the reduced influence of the analogical mechanism would be reflected in analyzing the acceptability ratings from Experiments 1b and 3b, such that the ratings from Experiment 3b (excluding real words) would show that the influence of lexical analogy is reduced in the absence of real words. The logic behind this suggestion is that "hard" cases are most likely to show an effect of lexical analogy on their acceptability ratings.

8.2.5 Implications for Well-formedness Judgments in Children and Bilinguals

The data from older adults indicate that the analogical mechanism's influence on acceptability ratings increases with age. This is interpreted as the result of having a larger lexicon, extended use and more experience with words. If this interpretation is correct, we would expect to find that acceptability ratings in children show a smaller influence of lexical analogy than in adults because children not only have a smaller lexicon, but they have been using words for a shorter time than adults. Thus, another experimental offshoot of the work in this dissertation is replicating Experiment 1 with much younger subjects.

It has been shown that infants are sensitive to many of the phonotactic properties of their language, and that this sensitivity may be a tool in speech segmentation and acquisition (e.g., Jusczyk et al., 1993; Cutler, 1994). Nine-month-olds are reported to be sensitive to prosodic cues and stress patterns, as well as grammatical probability (Mattys, Jusczyk, Luce and Morgan, 1999). In light of these findings about the early knowledge of phonotactic patterns and grammar, it is not implausible that, by the age of 6-7 years, children would have a grammar that is adult-like and stable. On the other hand, they continue to learn new words. Between the ages of eighteen months and six years, children acquire approximately eight words each day without instruction (Gleitman and Landau, 1994). The estimate is that a 6-year-old child has a vocabulary of

8,000 – 14,000 words (Carey, 1978). This means that children have a relatively large vocabulary. However, if we were to compare their lexicon with young and older adults, we would expect to find differences in size and language experience.

Based on the findings from the experiment with older adults, I predict that children will have the least amount of sensitivity to lexical analogy, compared to young and old adults. The proposed study would replicate Experiment 1 with children of about 6 years of age. Clearly, modifications will need to be made. For example, the rating scale must be reduced in order for it to be usable by young children. Also, children may assume that there are many words they do not know, and this assumption may affect their judgments. The hypothesis is that a decrease in age (i.e., testing young children) would result in a decrease in the influence of the analogical mechanism, but not in the influence of the grammatical mechanism. Such a study would not only enrich the current account of phonotactic well-formedness judgments, but it will also have implications for an analysis of well-formedness judgments that can account for maturational changes over the life span.

The impact of lexicon size and language use on the influence of the analogical mechanism has implications also for bilingual studies. If the influence of lexical analogy is closely tied with language use, lexicon size, and experience with words, one would expect to find these factors to have a similar influence on bilinguals. My predictions for a follow-up experiment with subjects who are bilingual speakers of English (where bilingual refers to speakers who have received input from two languages during the critical period for language acquisition) are that: 1) bilingual adults responding to the full list of stimuli from Experiment 1 (Chapter 4) will show a reduced effect of the analogical mechanism because they split their language use between two different languages (Gollan and Brown, 2006); but 2) bilinguals will show the same degree of influence from the grammatical mechanism that monolinguals do, because splitting their language use between two languages should not impact the grammatical mechanism (i.e., the grammatical mechanism would not be affected by the split use between the two languages).

Appendices

Appendix 1: List of Role Labels

Label	Description
Osi	stressed initial onset
Osf	stressed final onset
Osif	onset of a monosyllable
Owi	unstressed initial onset
Owf	unstressed final onset
Rsi	stressed initial rime
Rsf	stressed final rime
Rsif	rime of a monosyllable
Rwi	unstressed initial rime
Rwf	unstressed final rime

Appendix 2: List of Codas that Appear only in Suffixed Forms

zd

vz

vd

tθ

sts

ʃt

spt

sps

skt

sks

rz

rvz

rvd

rθs

rst

rpt

rnz

rnd

rmz

rmθ

rmd

rlz

rld

rkt

rks

rʤd

rgz

rft

rfs

rdz

rʧt

rbz

rbd

pts

pθ

pst

nzd

ŋz

nθs

ŋθ

nst

nʤd

ndz

ŋd

nʧt

mz

mt

mpst

mft

mfs

md

lz

lvz

lvd

ltst

lts

lst

lpt

lps

lmz

lmd

lkt

lks

lʤd

lft

lfθ

lfs

ldz

lʧt

lbz

kts

ksts

ʤd

gz

gd

fts

fθ

fs

ðz

dθ

dst

ðd

ʧt

bz

bd

Appendix 3: List of Non-word Stimuli

HS: High SIM LS: Low SIM

HP: High PLog LP: Low PLog V: Violation

Stimulus	Condition	Transcription
biss	HSHP	bɪs
cose	HSHP	kowz
dar	HSHP	dɑɹ
eed	HSHP	id
keet	HSHP	kit
kett	HSHP	kɛt
pazz	HSHP	pæz
pell	HSHP	pɛl
proom	HSHP	pɹum
reet	HSHP	ɹit
rinn	HSHP	ɹɪn
sahn	HSHP	sɑn
tood	HSHP	tud
cheb	HSLP	ʧɛb
gavv	HSLP	gæv
geeb	HSLP	gib
krezz	HSLP	kɹɛz
pairse	HSLP	peɪɹs
pelse	HSLP	pɛls
ruck	HSLP	ɹʌk
sess	HSLP	sɛs
sheeb	HSLP	ʃib
thal	HSLP	θæl
thidge	HSLP	θɪʤ
zeck	HSLP	zɛk
zoss	HSLP	zɑs

aksh	HSV	ækʃ
balt	HSV	bælt
dej	HSV	dɛʤ
froive	HSV	fɹoɪv
kloomp	HSV	klump
krih	HSV	kɹɪ
pairthe	HSV	peɪɹð
parr	HSV	pæɹ
petk	HSV	pɛtk
peysh	HSV	peɪʃ
sedl	HSV	sɛdl
soog	HSV	sug
tlee	HSV	tli
clail	LSHP	kleɪl
flist	LSHP	flɪst
jound	LSHP	ʤæwun
lunk	LSHP	lʌŋk
mamp	LSHP	mæmp
nank	LSHP	næŋk
prinks	LSHP	pɹɪŋks
quide	LSHP	kwaɪd
sadge	LSHP	sæʤ
scrize	LSHP	skɹaɪz
spreyt	LSHP	spɹeɪt
streen	LSHP	stɹin
swide	LSHP	swaɪd
byust	LSLP	bjust
dwelm	LSLP	dwɛlm
glarm	LSLP	glɑɹm
scube	LSLP	skjub
smusp	LSLP	smʌsp
splosp	LSLP	splɑsp

sprelse	LSLP	spɹɛls
spugue	LSLP	spjug
squav	LSLP	skwɑz
squemp	LSLP	skwɛmp
strulb	LSLP	stɹʌlb
twemp	LSLP	twɛmp
twimf	LSLP	twɪmf
blousk	LSV	blæwsk
froisp	LSV	fɹoɪsp
glousp	LSV	glæwsp
lafsh	LSV	læfʃ
plithf	LSV	plɪθf
reetk	LSV	ɹitk
skewfsh	LSV	skjufʃ
smaigl	LSV	smaɪgl
snoulp	LSV	snæwlp
splaishf	LSV	spleɪʃf
sprarthe	LSV	spɹæɹð
squailk	LSV	skwaɪlk
stroosk	LSV	stɹusk

Appendix 4: List of Real Word Stimuli

Real Word	Condition
rise	HSHP
mile	HSHP
flight	HSHP
deep	HSHP
beach	HSHP
mist	HSHP
wreath	LSHP
soup	LSHP
cling	LSHP
spy	LSHP
chop	LSHP
spare	LSHP
noise	HSLP
crush	HSLP
twine	HSLP
scrooge	HSLP
yelp	HSLP
twelve	HSLP
skew	LSLP
quad	LSLP
swamp	LSLP
glanced	LSLP
scope	LSLP
dwarf	LSLP

References

Abrams, L., White, K.K. and Eitel, S.L. 2003. Isolating phonological components that increase tip-of-the-tongue resolution. *Memory and Cognition*, **31**, 1153–1162.

Albert, M.S., Heller, H.S. and Milberg, W. 1988. Changes in naming ability with age. *Psychology and Aging*, **3**, 173–188.

Albright, A. 2006. Gradient phonotactic effects: lexical? grammatical? both? neither? Talk presented at the 80th annual meeting of Linguistic Society of America, 5–8 January 2006.

Albright, A. and Hayes, B. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, **90**, 119–161.

Baayen, H. In press. *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge: Cambridge University Press.

Baayen, R.H., Piepenbrock, R., and Gulikers, L. 1995. The CELEX lexical database (Release 2) [cd-rom]. Philadelphia, PA: Linguistics Data Consortium, University of Pennsylvania.

Bailey, T.M. and Hahn, U. 2001. Determinants of wordlikeness: phonotactics or lexical neighborhoods? *Journal of Memory and Language*, **44**, 568–591.

Bailey, T.M. and Hahn, U. 2004. What makes words sound similar? *Cognition*, **97**, 227-267.

Bailey, T.M. and Hahn, U. 2005. Phoneme similarity and confusability. *Journal of Memory and Language*, **52**, 339–362.

Balota, D.A., Cortese, M.J., Sergent-Marshall, S.D., Spieler, D.H. and Yap, M. 2004. Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, **133**, 283–316.

Bates, E. and Goodman, J. 1997. On the inseparability of grammar and the lexicon: evidence from acquisition, aphasia and real-tine processing. *Language and Cognitive Processes*, **12**, 507–584.

Berent, I., Steriade, D., Lennertz, T. and Vaknin, V. In press. What we know about what we have never heard before: Evidence from perceptual illusions. *Cognition*.

Birren, J.E. and Schaie, K. W. (Eds.). 2006. *Handbook of the Psychology of Aging*, 6th edition. San Diego, CA: Academic Press.

Boersma, B. and Weenink, D. Praat: doing phonetics by computer. http://www.praat.org.

Brown, A.S. and Nix, L.A. 1996. Age-related changes in the tip-of-thetongue experience. *American Journal of Psychology*, **109**, 79–91.

Brown, A.S. and Mitchell, D.B. 1991. Age differences in retrieval consistency and response dominance. *Journal of Gerontology: Psychological Processes*, **15**, 113–121.

Burke, D.M., MacKay, D.G., Worthley, J.S. and Wade, E. 1991. On the tip of the tongue: what causes word finding failures in young and older adults? *Journal of Memory and Language*, **30**, 542–579.

Burke, D.M., White, H. and Diaz, D.L. 1987. Semantic priming in young and older adults: Evidence for age constancy in automatic and attentional processes. *Journal of Experimental Psychology: Human Perception and Performance*, **13**, 79–88.

Bybee, J. 2003. *Phonology and Language Use*. Cambridge: Cambridge University Press.

Carey, S. 1978. The child as word learner. In M. Halle, J. Bresnan and G.A. Miller (Eds.), *Linguistic Theory and Psychological Reality*, 264–293. Cambridge, MA: MIT Press.

Carnegie Mellon University dictionary. 2004. A pronunciation dictionary for North American English found at http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Charles-Luce, J. and Luce, P. A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language*, **22**, 727–735.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. and Halle, M. 1965. Some controversial questions in phonological theory. *Journal of Linguistics*, **1**, 97–138.

Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English*. Cambridge, MA: MIT Press.

Clahsen, H. 1999. Lexical entries and rules of language: A multidisciplinary study in German inflection. *Behavioral and Brain Sciences*, **22**, 991–1060.

Clark, H. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, **12**, 335–359.

Clements, G.N. and Keyser, S.J. 1983. *CV phonology: a generative theory of the syllable*. Cambridge, MA: MIT Press.

Coetzee, A. and Pater, J. 2005. Gradient Phonotactics in Muna and Optimality Theory. ms. University of Michigan and University of Massachusetts, Amherst.

Cohen, G. and Faulkner, D. 1986. Memory for proper names: Age differences in retrieval. *British Journal of Developmental Psychology*, **4**, 187–197.

Cohen, J.D., MacWhinney, B., Flatt, M. and Provost, J. 1993. Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments and Computers*, **25**, 257–271.

Coleman, J.S., and Pierrehumbert, J. 1997. Stochastic phonological grammars and acceptability. *Computational Phonology, 3rd Meeting of the ACL special interest group in computational phonology*, 49–56. Somerset, NJ: Association for Computational Linguistics.

Cross, E.S. and Burke, D.M. 2004. Do alternative names block young and older adults' retrieval of proper names? *Brain and Language*, **89**, 174–181.

Cutler, A. 1994. Segmentation problems, rhythmic solutions. In L. Gleitman and B. Landau (Eds.), *The Acquisition of the Lexicon*. Cambridge: MIT Press.

Damasio, A.R. and Damasio, H. 1992. Brain and language. *Scientific American*, **267**, 88–95.

Daugherty, K. and Seidenberg, M. 1994. Beyond rules and exceptions: A Connectionist approach to inflectional morphology. In S.D. Lima, R.L. Corrigan and G.K. Iverson (Eds.), *The Reality of Linguistic Rules*, 353–388. Amsterdam: Benjamins.

Dell, G.S. and Gordon, J.K. 2003. Neighbors in the lexicon: Friends or foes? In N. O. Schiller and A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, 9–39. New York: Mouton de Gruyter.

Derwing, B.L. and Nearey, T.M. 1984. Experimental Phonology at the University of Alberta. In J. J. Ohala and J. J. Jaeger (Eds.), *Experimental Phonology*, 187–210. Orlando, FL: Academic Press.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D. and Plunkett, K. 1996. *Rethinking Innateness: A Connectionist Perspective on Development.* Cambridge, MA: MIT Press.

Fodor, J. 1983. *The modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Forster, K. 1979. Levels of processing and the structure of the language processor. In W. E. Cooper and E. C. Walker (Eds.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.

Friederici, A.D., Schriefers, H. and Lindenberger, U. 1998. Differential age effects on semantic and syntactic priming. *International Journal of Behavioral Development*, **22**, 813-845.

Friedrich, M. and Friederici, A.D. 2005. Phonotactic knowledge and lexical-semantic processing in one-year-olds: Brain responses to words and nonsense words in picture contexts. *Journal of Cognitive Neuroscience*, **17**, 1785–1802.

Frisch, S. 1996. Similarity and frequency in phonology. Ph.D. Thesis, Northwestern University, Evanston, Illinois.

Frisch, S., Broe, M. and Pierrehumbert, J. 1997. Similarity and phonotactics in Arabic. Rutgers Optimality Archive [Online], ROA-223-1097. Available at http://www.webslingerz. com/cgi-bin/oa_list.cgi.

Frisch, S.A., Large, N.R. and Pisoni, D. 2000. Perception of wordlikeness: effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language*, **42**, 481–496.

Gleitman, L. and Landau, B. (Eds.). 1994. *The Acquisition of the Lexicon*. Cambridge: MIT Press.

Goldinger, S.D., Luce, P.A. and Pisoni, D.B. 1989. Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, **28**, 501–518.

Goldsmith, J. 2006. Information theory and phonology. Talk presented at the 80th annual meeting of Linguistic Society of America, 5–8 January 2006.

Gollan, T.H. and Acenas, L.A. 2004. What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **30**, 246–269.

Gollan, T.H. and Brown, A.S. 2006. From tip-of-the-tongue data to theoretical implications in two steps: When more TOT means better retrieval. *Journal of Experimental Psychology*, **135**, 462–480.

Gollan, T.H., Montoya, R.I., Cera, C. and Sandoval, T.C. 2006. More use almost always a means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. Submitted.

Gordon, J.K. 2002. Phonological neighborhood effects in aphasic speech errors: Spontaneous and structured contexts. *Brain and Language*, **82**, 113–145.

Greenberg, J.H. 1965. Some generalizations concerning initial and final consonant sequences. *Linguistics*, **18**, 5–34.

Greenberg, J.H. and Jenkins, J.J. 1964. Studies in the psychological correlates of the sound system of American English. *Word*, **20**, 157–177.

Halle, M. and Mohanon, K.P. 1985. Segmental phonology of modern English. *Linguistic Inquiry*, **16**, 57–116.

Hammond, M. 1999. *The phonology of English*. New York: Oxford University Press.

Hammond, M. 2004. Gradience, phonotactics, and the lexicon in English phonology. *International Journal of English Studies*, **4**, 1–24.

Harley, TA. and Brown, H.E. 1998. What causes a tip-of-the-tongue state? Evidence for lexical neighborhood effects in speech production. *British Journal of Psychology*, **89**, 151–174.

Hay, J., Pierrehumbert, J. and Beckman, M. 2003. Speech perception, well-formedness and the statistics of the lexicon. In J. Local, R. Ogden, and R. Temple (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press.

Hayes, B. and Wilson, C. To appear. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*.

James, L. and Burke, D.M. 2000. Phonological priming effects on word retrieval and tip-of-the-tongue experiences in young and older adults. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **26**, 1378–1391.

Johnson, S.C., Saykin, A.J., Flashman, L.A., McAllister, T.W., O'Jile, J.R., Sparling, M.B., Guerin, S.J., Moritz, C.H. and Mam I amian, A.C. 2001. Similarities and differences in semantic and phonological processing with age: Patterns of functional MRI activation. *Aging Neuropsychology*, **8**, 307–320.

Jones, G.V. 1989. Back to Woodworth: Role of interlopers in the tip of the tongue phenomenon. *Memory and Cognition*, 17, 69–76.

Jones, G.V. and Langford, S. 1987. Phonological blocking in the tip of the tongue state. *Cognition*, **26**, 115–122.

Jurafsky, D. and Martin, J.H. 2000. *An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, NJ: Prentice Hall.

Jusczyk, P.W., Cutler, A. and Redanz, L. 1993. Infants' sensitivity to predominant stress patterns in English. *Child Development*, **64**, 675–687.

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

Kemper, S. 1992. Adults' sentence fragments: Who, what, when, where, and why. *Communication Research*, **19**, 444–458.

Kemper, S., Greiner, L., Marquis, J.G., Prenovost, K. and Mitzner, T. 2001. Language decline across the life span: Findings from the Nun study. *Psychology and Aging*, **16**, 227–239.

Kempler, D. and Zelinski E.M. 1994. Language in dementia and normal aging. In F. A. Huppert, C. Brayne and D. W. O'Connor (Eds.), *Dementia and Normal Aging*, Chapter 16. New York: Cambridge University Press.

Kempler, D., Curtiss, S. and Jackson, C. 1987. Syntactic preservation in Alzheimer's disease. *Journal of Speech and Hearing Research*, **30**, 343–350.

Kessler,B. and Treiman, R. 1997. Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language*, **37**, 295–311.

Kruskal, J.B. 1983. An overview of sequence comparison. In D. Sankoff and J.B. Kruskal (Eds.), *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*, 1-44. Reading, MA: Addison-Wesley.

Laver, G.D. 2000. A speed-accuracy analysis of word recognition in young and older adults. *Psychology and Aging*, **15**, 705–709.

Laver, G.D. and Burke, D.M. 1993. Why do semantic priming effects increase in old age? A meta-analysis. *Psychology and Aging*, **8**, 34–43.

Light, L.L. 1991. Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology*, **42**, 333–376.

Lipinski, J. and Gupta, P. 2003. Separating the effects of duration and neighborhood density in nonword repetition latency. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th annual conference of the cognitive science society*. Hillsdale, NJ: Lawrence Erlbaum.

Lipinski, J. and Gupta, P. 2005. Does neighborhood density influence repetition latency for nonwords? Separating the effects of density and duration. *Journal of Memory and Language*, **52**, 171–192.

Lorch, R. F., Myers, J. 1990. Regression analyese of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 149-157.

Luce, P. 1986. A computational analysis of uniqueness points in auditory word recognition. *Perception and Psychophysics*, **39**, 155–158.

Luce, P.A. and Pisoni, D.B. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, **19**, 1–36.

Luce, P.A., Goldinger, S.D., Auer, E.T. and Vitevitch, M.S. 2000. Phonetic priming, neighborhood activation, and PARSYN. *Perception and Psychophysics*, **62**, 615–625.

Luce, P.A., Pisoni, D.B. and Goldinger, S.D. 1990. Similarity neighborhoods of spoken words. In G. Altmann (Ed.), *Cognitive Models of Speech perception: Psycholinguistic and Computational Perspectives*, 122–147. Cambridge, MA: MIT Press.

MacDonald, M.C., Pearlmutter, N.J. and Seidenberg, M. S. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review*, **101**, 676–703.

MacKay, D.G. and Burke, D.M. 1990. Cognition and aging: A theory of new learning and the use of old connections. In T. M. Hess (Ed.), *Aging and Cognition: Organization and Utilization*, Chapter 5. North-Holland: Elsevier Science Publishers.

Marslen-Wilson, W.D. 1989. Access and integration: Projecting sound onto meaning. In W. D. Marslen-Wilson (Ed.), *Lexical Access and Representation*. Cambridge, MA: Bradford.

Marslen-Wilson, W.D. and Tyler, L.K. 1998. Rules, representations, and the English past tense. *Trends in Cognitive Science*, **11**, 428–435.

Massaro, D.W. and Cohen, M.M. 1983. Phonological constraints in speech perception. *Perception and Psychophysics*, **34**, 338–348.

Mattys, S.L., Jusczyk, P.W., Luce, P.A. and Morgan, J.L. 1999. Phonotactic and Prosodic Effects on Word Segmentation in Infants. *Cognitive Psychology*, **38**, 465–494.

Maylor, E. 1990a. Age, blocking and the tip of the tongue state. *Journal of Psychology*, **81**, 123–133.

Maylor, E. 1990b. Recognizing and naming faces: Aging, memory retrieval, and the tip-of-the-tongue state. *Journal of Gerontology*, **45**, 215–226.

McCullagh, P. and Nelder, J.A. 1989. *Generalized Linear Models*, 2nd edition. London: Chapman and Hall.

McQueen, J.M., Cutler, A., Briscoe, T. and Norris, D. 1995. Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes*, **10**, 309–331.

Meyer, A.S. and Bock, K. 1992. The tip-of-the-tongue phenomenon: Blocking or partial activation? *Memory and Cognition*, **20**, 715–726.

Mitton, R. 1992. A computer-usable dictionary file based on the Oxford Advanced Learner's Dictionary of Current English. ftp://ota.ox.ac.uk/pub/ota/public/dicts/710/text710.dat

Newman, R.S., Sawusch, J.R. and Luce, P.A. 1997. Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance*, **23**, 873–889.

Nofosky, R. 1986. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **14**, 700–708.

Norris, D., McQueen, J.M. and Cutler, A. 2000. Merging information in speech recognition: Feedback is never necessary. *Brain and behavioral Sciences*, **23**, 299–370.

Ohala, J. and Ohala, M. 1986. Testing hypotheses regarding the psychological manifestation of morpheme structure constraints. In J. J. Ohala and J. J. Jaeger (Eds.), *Experimental Phonology*, 239–252. Orlando, FL: Academic Press.

Pertz, D.L. and Bever, T.G. 1975. Sensitivity to phonological universals in children and adolescents. *Language*, **51**, 149–162.

Pinker, S. 1991. Rules of language. *Science*, **253**, 530–534.

Pinker, S. 1994. *The Language Instinct*. New York: William Morrow.

Pinker, S. 1999. *Words and Rules: the Ingredients of Language.* New York: Harper Collins Publishers.

Pinker, S. and Prince, A. 1988. One language and connectionism: Analusis of a paralle distributed processing model of language acquisition. *Cognition*, **48**, 21–69.

Pitt, M.A. 1998. Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception and Psychophysics*, **60**, 941–951.

Pitt, M.A. and McQueen, J.M. 1998. Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, **39**, 347–370.

Pitt, M.A. and Samuel, A.G. 1993. An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance*, **19**, 699–725.

Pitt, M.A. and Samuel, A.G. 1995. Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology*, **29**, 149–188.

Prasada, S. and Pinker, S. 1993. Generalizations of regular and irregular morphological patterns. *Language and Cognitive Processes*, **8**, 1–56.

Pylkkänen, L., Stringfellow, A. and Marantz, A. 2002. Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language*, **81**, 666–678.

R statistical programming environment, version 2.4.1. 2007. http://lib.stat.cmu.edu/R/CRAN/

Raudenbush, S.W. and Bryk, A.S. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd Ed. Thousand Oaks: Sage Publications.

Reason, J.T. and Lucas, D. 1984. Using cognitive diaries to investigate naturally occurring memory blocks. In J. E. Harris and P. E. Morris (Eds.), *Everyday Memory, Actions, and Absentmindedness*, 53–69. Orlando, FL: Academic Press.

Rumelhart, D.E. and McClelland, J.L. 1986. On learning the past tenses of English verbs. In D. E. Rumelhart. J. D. McClelland and the PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: vol.2*, 216-271. Cambridge, MA: MIT Press.

Salthouse, T.A. 2000. Methodological assumptions in cognitive aging research. In F. I. Craik and T. A. Salthouse (Eds.), *The Handbook of Aging and Cognition*, 467-498. Mahwah, N.J.: Lawrence Erlbaum Associates

Samuel, A.G. and Pitt, M.A. 2003. Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, **48**, 416–434.

Schaie, K.W. and Parham, I.A. 1977. Cohort-sequential analyses of adult intellectual development. *Developmental Psychology*, **13**, 649–653.

Scholes, r. 1966. *Phonotactic Grammaticality*. The Hague: Mouton.

Schreuder, R. and Baayen, H. 1995. Modeling morphological processing. In L. Feldman (Ed.), *Morphological Aspects of Language Processing*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Schütze, C.T. 1996. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.

Seidenberg, M. 1992. Connectionism without tears. In S. Davis (Ed.), *Connectionism: Theory and Practice*, 84–122. New York: Oxford UP.

Spieler, D.H. and Balota, D.A. 2000. Factors influencing word naming in younger and older adults. *Psychology and Aging*, **15**, 225–231.

Stevens, James. 1996. *Applied multivariate statistics for the social sciences*, 4th edition. Mahwah, N.J.:Lawrence Erlbaum Associates.

Storkel, H.L., Armbrüster, J. and Hogan, T.P. 2006. Differentiating Phonotactic Probability and Neighborhood Density in Adult Word Learning. *Journal of Speech, Language, and Hearing Research*, **49**, 1175–1192.

Taylor, J.K. and Burke, D.M. 2002. Asymmetric aging effects on semantic and phonological processes: Naming in the picture-word interference task. *Psychology and Aging*, **17**, 662–676.

Thornton, R. and Light, L.L. 2006. Language comprehension and production in normal aging. In J. E. Birren and K. Warner Schaie (Eds.), *Handbook of the Psychology of Aging*, Chapter 12. Burlington, MA: Elsevier Academic Press.

Treiman, R., Kessler, B., Knewasser, S., Tincoff, R. and Bowman, M. 2000. English speaker's sensitivity to phonotactic patterns. In M.B. Broe and J.B. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 269–282. Cambridge, UK: Cambridge UP.

Ullman, M. 2001. The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, **30**, 37–69.

Ullman, M., Corkin, S., Coppola, M., Growdon, J., Koroshez, W. and Pinker, S. 1997. A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience*, **9**, 266–276.

Varhaeghen, P. 2003. Aging and vocabulary scores: A meta-analysis. *Psychology and Aging*, **18**, 332–339.

Vitevitch, M. and Luce, P.A. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, **40**, 374-408.

Vitevitch, M., Luce, P., Pisoni, D. and Auer, E. 1999. Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language*, **68**, 306–311.

Vitevitch, M.S. 1997. The neighborhood characteristics of malapropisms. *Language and Speech*, **40**, 211–288.

Vitevitch, M.S. 2002. The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **28**, 735–747.

Vitevitch, M.S. and Luce, P.A. 1998. When words compete: Levels of processing in perception of spoken words. *Psychological Science*, **9**, 325–329.

Vitevitch, M.S. and Luce, P.A. 2005. Increases in phonotactic probability facilitate spoken non-word repetition. *Journal of Memory and Language*, **52**, 193–204.

Vitevitch, M.S. and Sommers, M.S. 2003. The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory and Cognition*, **31**, 491–504.

Vitevitch, M.S., Luce, P.A., Charles-Luce, J. and Kemmerer, D. 1997. Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, **40**, 47–62.

White, K.K. and Abrams, L. 2002. Does priming specific syllables during tip-of-the-tongue states facilitate word retrieval in older adults? *Psychology and Aging*, **17**, 226-235.

Wilson, C. 2003. Experimental investigation of phonological naturalness. In G. Garding and M. Tsujimura (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, 553-546. Somerville, MA: Cascadilla Press.

Wingfield, A. and Stine-Morrow, E.A. 2000. Language and speech. In F. I. M. Craik and T. A. Salthouse (Eds.), *Handbook of Aging Cognition*, Chapter 6. Mahwah, NJ: Lawrence Erlbaum Associates.

Wurm, L.H. and Samuel, A.G. 1997. Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring. *American Journal of Psychology*, **85**, 425–430.

Zacks, R.T. and Hasher, L. 1997. Cognitive gerontology and attentional inhibition: A reply to Burke and McDowd. *Journal of Gerontology: Psychological Sciences*, **52**, 274-283.

Zamuner, T.S., Gerken, L. and Hammond, M. 2004. Phonotactic probabilities in young children's speech production. *Journal of Child Language*, **31**, 515–536.

Zelinski, E. M., Gilewski, M. J. and Thompson, L. W. 1980. Do laboratory tests relate to self-assessment of memory ability in the young and old? In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Arenberg and L. W. Thompson (Eds.), *New directions in memory and aging: Proceedings of the George A. Talland Memory Conference*, 519–544. Hillsdale, NJ: Lawrence Erlbaum Associates.