COMPLEXITY AND NATURALNESS IN FIRST LANGUAGE AND SECOND LANGUAGE PHONOTACTIC LEARNING

Brandon Prickett

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Linguistics.

Chapel Hill
2015

Approved by:

Elliott Moreton

Jennifer Smith

Jeff Mielke

ABSTRACT

BRANDON PRICKETT: Complexity and Naturalness in First Language and Second Language Phonotactic Learning
(Under the direction of Elliott Moreton)

This thesis sought to test learning biases present in first language (L1) and second language (L2) acquisition. It was predicted that a bias for phonetically grounded, typologically common ("natural") patterns would be present in L1 learning, and that a bias for patterns involving a small number of phonetic features ("simple") would be present in L2 learning. A methodology similar to Hayes and White (2013) was used to test the presence of these biases in the learning of 30 speakers of Canadian English and Canadian French by having the subjects perform grammaticality judgment tasks in both of these languages.

The results suggest biases for simple and natural constraints could be more active in L1 learning than in L2 learning. This falls in line with past studies and raises questions about the nature of L2 acquisition. Future research could explore whether L2 phonological learning has fewer biases in general, or if it has biases other than naturalness and complexity.

ACKNOWLEDGEMENTS

First of all, thank you to the Juanita A. Melchert Graduate Excellence Fund for the money to run and advertise the experiment for this thesis. Thanks to my family and to my fiancée for all of their support during a year when I was often too busy to see them as much as I would have liked to. I was lucky enough to discuss my thesis with Joe Pater and John McCarthy who gave some helpful advice and pointed me toward some great resources of relevance—so thank you to both of them. Several graduate students from other universities offered their assistance, including Sophia Stevenson, Megan Gotowski, and especially Jeff Lamontagne. Thank you to Rachel Broad and Katya Pertsova for their advice and assistance during Sporklab meetings and to the other members of UNC's P-Side group. Chris Wiesen at the Odum Institute has now done the statistical analysis for two of my theses, and for his patience and helpfulness I am very grateful.

A big thanks goes to my thesis committee members, Jennifer Smith and Jeff Mielke, who not only advised me during the process of making the thesis, but were invaluable when I was recruiting subjects for my experiment. And the biggest thanks of all goes to Elliott Moreton for both advising this project and encouraging me to get my master's degree in the first place. This research—and any other phonological research I do—is a direct result of the "Intro to Phonology" class that he taught when I was an undergraduate.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

Recently a "surfeit of the stimulus" has been discovered in language learning[1]. That is, speakers do not seem to be learning all of the sound patterns that exist in the words of their language (Becker et al. 2011). It appears that humans, when learning their first language, use certain analytical biases when deciding which of these patterns to internalize into their mental grammars. Two biases have been proposed: a bias towards phonetically grounded (sometimes called substantive) patterns (hereafter referred to as a "naturalness" bias) and a bias towards patterns that are less logically complex (what "complex" means in this context will be described more in Section 2). So far, artificial language studies have primarily found an effect of complexity, but no effect of naturalness (for instance, Skoruppa and Peperkamp 2011; for a review of studies like this, see Moreton and Pater 2011: Parts I and II). Conversely, studies performed on speakers' native languages (hereafter L1) have found a strong apparent effect of naturalness (Becker et al. 2011; Hayes and White 2013; Prickett 2014) but no effect of complexity (Prickett 2014).

One explanation for this phenomenon is that speakers go through a period during which naturalness has a strong effect on their phonological development (Hayes 1999). If there is a critical period for phonology, any task designed to test phonological learning after this critical

[1] For the purposes of this paper, the terms "learning" and "acquisition" will be used interchangeably to mean any process by which a person discovers and attempts to copy the phonological patterns within a language.

period (such as an artificial language study) would be testing a different process than speakers use to learn their L1. This could explain why there seems to be different learning biases in artificial language studies than in studies that involve speakers' L1. If a critical period is the cause of this distinction, second language (hereafter L2) acquisition would likely behave in the same manner, and have the same biases as artificial language learning since they would both be happening outside of the critical period. This means that pattern complexity would be expected to have an effect on L2 learning—as in artificial languages. If the same biases that exist in artificial language learning could be shown to exist in L2 acquisition, this would suggest that L1 learning could be a unique experience at the beginning of a person's life.

The idea that children learn language differently than adults is not a new one (see, for instance, Lenneberg 1967), but it remains an important and widely debated theory in the field of Linguistics (Long 2005 gives a review of some of the controversy still surrounding this issue). If it was found that L1 and L2 learning are similar, this could suggest that at least some adult acquisition is similar to that of children and could support findings that show L2 learning to be unaffected by a speaker's first language (Trapman and Kager 2009).

If it was found that L1 speakers and L2 speakers learn phonological patterns differently, then this could help support and explain the concept of a critical period. This could also help with the interpretation of artificial language studies, since these would be shown to be more accurate simulations of L2 learning than of L1 learning. Furthermore, if artificial language learning is more analogous to L2 learning than to L1 learning, this would mean that the findings of an effect of naturalness in studies that test L1 learning such as Hayes and White (2013) do not contradict the findings of artificial language studies that find a stronger effect of complexity. This could provide support for the idea that L1 phonological development is strongly influenced by phonetic groundedness. This would help explain the surfeit of the stimulus, improve models of phonological learning, and help to explain why typologically common phonological processes occur in many of the world's languages. The results of this study found significant biases in L1 learning that supported the findings of studies like Hayes and White (2013), but did not find similar results in L2 learning.

2. BACKGROUND

2.1 Testing for a Naturalness Bias in English Phonological Learning

One of the biases that was tested in this study was a bias for natural phonological patterns. Hayes and White (2013) tested for this bias in the L1 learning of American English speakers. Using the UCLA Phonotacic Learner (Hayes and Wilson 2008), Hayes and White produced a list of phonotactic patterns (in the form of weighted constraints) apparent in English[2]. Ten of these constraints that were well attested among languages and were phonetically logical—such as voicing assimilation—were called “natural” for the purposes of the study. Another ten constraints that seemed to be “accidentally true”, as in the case of some the Wargamay constraints found by Hayes and Wilson (2008), were chosen as well. Although these constraints did represent patterns from the English data, they did not seem to be phonetically motivated, nor were they typologically common and were called “unnatural” by Hayes and White.

Novel words were created in order to test each constraint’s effect on speaker judgments. If words that violated a constraint were judged to be ungrammatical by a speaker, it was assumed that the violated constraint existed in the speaker’s grammar. Twenty-nine participants were presented with novel words that each violated one of the constraints and a control group of words that did not violate any constraints. The control group words were as similar to the violating

[2] A modified form of the CMU pronouncing dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict) was used as training data for the learner. The words were filtered for inflected forms, compounds, and “highly transparent processes of derivation” (Hayes and White 2013:7).

group words as possible without actually violating the constraints. The participants in the study were asked to perform magnitude estimations of how “good” the words sounded to them. They expressed this judgment both numerically and through drawing lines of different lengths (the better sounding words received higher numerical ratings and longer lines). The line lengths and numerical ratings of violating words were compared with those of their control group partners to determine how much each constraint affected the speakers’ judgments.

If natural constraints had more of an effect on speaker judgments than unnatural constraints, then the difference between judgments of words that violated natural constraints and their control group partners would be larger than the difference for words that violated unnatural constraints and their control group partners. For example, if a voicing assimilation constraint had a large affect on subjects’ judgments, this would mean that the subjects would rate words that violated voicing assimilation as very low, causing a large difference between these words and their control group partners (which would have a high rating since they did not violate any constraints). Alternatively, if a constraint against diphthongs before [g] did not affect judgments very much, ratings of words violating this constraint would be about the same as the ratings of their control group partners.

Hayes and White found that on average, the difference in subject ratings for words that violated unnatural constraints and ratings for the words’ control group partners were not significantly different—meaning that people did not seem to be affected by the unnatural constraints. Words that violated the natural constraints, on the other hand, were given significantly lower ratings, on average, than their control group partners. Figure 1 shows both of these results. These findings suggest that natural constraints were being learned by speakers while unnatural constraints were not, or at the very least that unnatural constraints were not being learned as well as natural constraints. Hayes and White propose that this bias could arise from a "phonetically driven" phonology (2013:24) which would mean that speakers are more prone to learning constraints that aid in the production and perception of speech sounds.

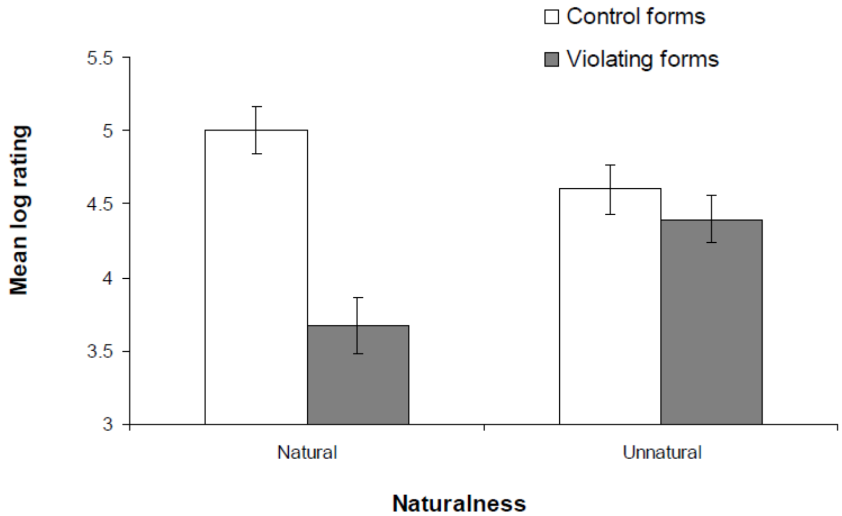

Figure 1. The mean log ratings of words that violate natural and unnatural constraints and their control group partners. The larger the difference between the white bar and the gray bar, the larger effect that group of constraints had on native speaker ratings. Taken from Hayes and White (2013).

2.2 Phonological Learning Biases in Artificial Languages

There have been numerous studies that used artificial language to explore the subject of learning biases in phonology (see Moreton and Pater 2012 parts I and II for an extensive review). Skoruppa and Peperkamp (2011) performed one of these studies using artificial accents in French. French speakers listened to stories read in an artificial accent and were then tested on how well they had learned the patterns that the accent used. Specifically, the accents differed from French in the harmonization processes that their vowels went through. In their first experiment, some subjects were exposed to an accent that harmonized vowels (representing a natural pattern), while others were exposed to an accent that underwent vowel disharmony (representing an unnatural pattern). Both of these accents were learned by the participants equally well, which suggested that there was no naturalness bias in their learning. In Skoruppa and Peperkamp's (2011) second experiment, French speakers were exposed to an artificial accent that harmonized some vowels while disharmonizing others (presenting a higher level of logical complexity than the patterns in the first experiment). The more complex pattern was more difficult for the speakers to adapt to, suggesting that complexity had an effect on speaker learning.

This kind of complexity has been shown to affect other forms of learning, such as visual pattern learning (Shepard et al.1961). It involves the idea that the easier it is to generalize a category, the simpler the pattern that matches it is. So in Skoruppa and Peperkamp (2011), for example, subjects who were presented with an accent that harmonized all vowels ("harmonize *x*" where *x* is a single, generalized group of all vowels) were learning a simpler pattern than the subjects that were tasked with learning an accent that only harmonized some of the vowels ("harmonize *x* vowels, but disharmonize *y* vowels" where *x* and *y* are two different types of vowels). Skoruppa and Peperkamp propose that rather than naturalness, this form of complexity seems to be the predominant bias in phonological learning. In order to test this claim, complexity was also one of the biases that this study looked for.

2.3. Naturalness vs. Complexity

Prickett (2014) sought to explore the seemingly contradictory findings of studies on natural languages (such as Hayes and White 2013) that find a bias towards naturalness in phonotactic learning, and studies that use artificial languages (such as Skoruppa and Peperkamp 2011) that find an effect of complexity in phonotactic learning. In order to do this, a methodology similar to Hayes and White (2013) was used to test both the effect of naturalness and of complexity on the learning of constraints in American English.

As in Hayes and White (2013), constraints that represented phonological patterns found in English by the Hayes and Wilson (2008) learner were used. Constraints were then chosen that fell into one of the following categories: Natural Simple, Natural Complex, Unnatural Simple, and Unnatural Complex. Constraints were considered natural if they could represent typologically common processes and if they seemed phonetically grounded. Constraints were considered complex if they used a large number of features to describe the patterns they banned. Chomsky and Halle (1968) describe this kind of complexity when discussing phonological rules, and it has been shown to affect non-linguistic learning (Bulgarella and Archer 1962). For example, "*[+round,+high][-consonantal,-sonorant]" would be more complex than "*[+voice][-voice]".

Prickett (2014) presented 77 subjects with novel words that violated constraints in each of the above-mentioned categories and with novel words that did not violate any of the constraints (to act as a control group as in Hayes and White 2013). Subjects were recruited using Amazon.com's Mechanical Turk website, which allows workers to be anonymously paid for performing short tasks. Participants were asked to numerically rate words on how "good" they sounded. Words were presented using orthography and an audio recording.

The results of this experiment were very similar to those in Hayes and White (2013). Like in Hayes and White (2013), words that violated natural constraints were given significantly lower ratings than their control group partners, and words that violated unnatural constraints did not have significantly lower rating than the control group. Complexity did not seem to have the same kind of an effect on word ratings. See Figure 2 for an illustration of these results. This further enforces the seeming divide between learning bias studies that have been performed using natural languages (which tend to show an effect of naturalness) and studies that have been performed using artificial languages (which tend to show a complexity effect). Prickett (2014) suggests that this could be caused by a difference in the way speakers learn the phonology of their L1 and the process that a person would go through to learn an artificial language in an experiment.

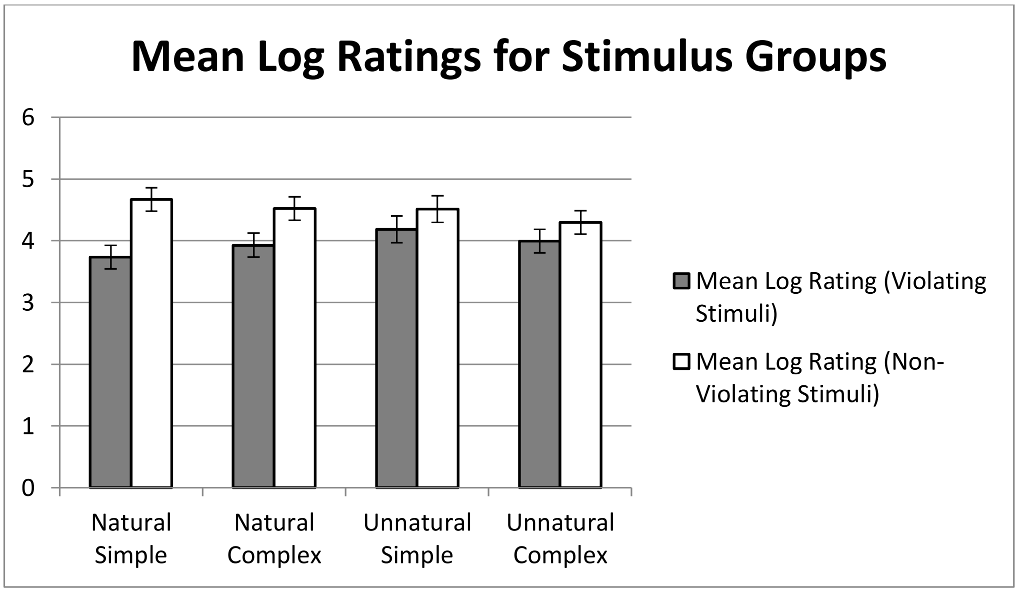

Figure 2. A comparison between the mean log ratings of each group. Control groups (stimuli that did not violate any constraints) are in white and experimental groups (those that did violate constraints) are in gray. Taken from Prickett (2014).

2.4. Learning L1 Phonology

One possible reason why the learning of one's native language might be different than the learning of an artificial language in an experiment is that the original phonology used in a person's L1 is created using a process that is unique to early childhood. Hayes (1999) presented the idea, called *inductive grounding*, that speakers could be born with the knowledge of how to derive natural constraints. Children, Hayes said, experiment with the sounds and sequences of sounds that occur in the language they hear being spoken around them. They then calculate which of the sound sequences require more or less effort by articulating the sounds themselves. After finishing this test drive of their vocal tract, the children create a set of phonological constraints that strike a balance between being efficient on a phonetic level and being logically simple. The children then proceed to rank these constraints according to observations about how the adults around them are speaking until the children have constructed a grammar identical to that of the adults they were learning from.

This theory allows children to construct grammars based on information that is readily available to them—such as their own physiology. Since human beings have a generally universal physiology, the constraints that arise cross-linguistically would be basically the same; with the constraint ranking being what differed between languages, as in classical Optimality Theory (Prince and Smolensky 1993). The Hayes (1999) approach could account for speakers' preference for natural constraints in their L1 (as shown in studies such as Hayes and White 2013 and Prickett 2014), since these would be the original constraints speakers induced as children. Perhaps unnatural constraints are learned on a subconscious level as well as people grow older (or in artificial language studies), but these would not be internalized into the grammar through the same process and could be subject to different learning biases (such as a bias for simple constraints).

If this theory is correct, and if it accounts for the different learning biases found in artificial and natural language studies, then one would also expect any phonology learned after the critical period of natural constraint induction to have effects more like artificial language studies. Since L2 learning would also occur after a child has already created all of their natural constraints, L2 grammaticality judgments would be expected to show learning biases more similar to artificial language learning—that is less of an effect of naturalness and more of an effect of complexity. The current study compared L1 and L2 learning biases to test this idea that artificial language learning could be more analogous to the learning of a second language.

3. EXPERIMENT DESIGN

3.1 Predictions

If Hayes' (1999) theory of phonetic inductive grounding is a correct explanation for why artificial language learning studies (such as Skoruppa and Peperkamp 2011) seem to find a greater effect of complexity in phonotactic learning, and studies on natural language (such as Prickett 2014) seem to find a larger effect of naturalness in phonotactic learning, then one would expect L2 learning to have the same biases as artificial language learning. Specifically, one would expect L2 learning to show effects of complexity and not of naturalness. The effects of these variables on learning can be measured as in Prickett (2014), and as described below in §4.4.

3.2 Choosing Languages to Test

In order to test L2 learning, it was necessary to first choose two languages to use for testing. Using two languages provided two different L2 learning experiences to test (since there would be participants who learned each language as their L2), and provided L1 acquisition information from both languages for comparison (since participants would speak the other relevant language as their L1).

English and French were chosen because of the unique environment the two languages share in Canada. Two kinds of participants were needed for this study: those who learned English as their first language and French as their second, and those who learned French as their first language and English as their second—both of which can be found in abundance in Canada.

3.3 Finding Patterns to Test

In order to test for biases in L1 and L2 phonological learning, it was necessary to find phonological patterns (in the form of phonotactic constraints) that are apparent English as well as patterns (also in the form of constraints) apparent in French. These constraints needed to fall into four categories: Natural Simple, Natural Complex, Unnatural Simple, and Unnatural Complex.

To be considered *natural*, constraints had to have an obvious phonetic motivation and resemble a phonological pattern seen in at least one other language in the P-Base database (Mielke 2008). This was designed to ensure that these constraints were both phonetically natural (Hayes 1999) and typologically natural. For example, voicing assimilation would be considered very natural, since it allows for easier pronunciation of consonant clusters and is seen in many of the world's languages. *Unnatural* constraints were those that seemed phonetically irrelevant and were not well attested typologically. An example of an unnatural constraint is the tendency in English for round, tense vowels to not be followed by [g]. This does not aid in production or perception of words and is not typologically well attested. The same process for determining the naturalness of a constraints was used in Prickett (2014), and similar logic was applied in Hayes and White (2013).

Complexity was defined as the number of variables needed to describe a pattern. This was the kind of complexity tested for in L1 learning by Prickett (2014) and was described by Chomsky and Halle (1968). It has also been shown to affect non-linguistic learning (Bulgarella and Archer 1962). For the purposes of this study, any constraint with less than four variables (in the form of features[3]) used in its description was considered *simple*, and any constraint with four or more features was considered *complex*. For example, a constraint like "*[+LABIAL][+LABIAL]" would be considered simple, since it only requires two features to describe (the two instances of "Labial"). However, a constraint like "*[+round,+high][-high,-low,-back,+tense]" would be considered complex, since it requires six feature values to describe (round, the two instances of high, low, back, and tense).

The constraints also needed to control for L1 interference. L1 interference has been shown to affect L2 acquisition in a variety of ways (for a review of many of these, see Kilpatrick 2009: Section 1.2). While the extent to which an L1 affects the L2 is debated, it was decided that for the purposes of this study, any possible L1 effects on L2 learning should be avoided. In order to do this, constraints needed to be chosen that were representative of patterns only in the language they were found in. In order to assure this, all of the English constraints chosen were violated by French words, and all of the French constraints chosen were violated by English words.

Prickett (2014) and Hayes and White (2013) relied on the UCLA Phonotactic Learner (Hayes and Wilson 2008) for finding the constraints that they tested. Unfortunately, due to the fact that this study required constraints that existed solely in either English or French, the constraints induced by the learner for these two languages were not sufficient. While some of the constraints used in the study *were* induced by the UCLA Phonotactic Learner, the majority were not. Those that were induced by the learner for English were taken from the

[3] Features such as [rhyme] and [word_boundary] were included in this count, since the UCLA phonotactic learner treats these in the same way as their more phonetically relevant counterparts (such as features like [voice] and [DORSAL]).

list of 160 constraints used by Hayes and White (2013) based on how well they met the criteria for this study discussed above[4].

Those that were induced by the Learner for French used the Lexique database (New et al. 2001) as training data. The pronunciations from over 350,000 words were taken from Lexique and converted to a transcription style that was similar to the one used for the training data in Hayes and White (2013). Any differences in the transcription styles were a result of differences in the phonemic inventory of English and French. Once the words in Lexique were converted, they were fed into the UCLA phonotactic learner to produce constraints apparent in the French data.

The rest of the patterns were found by looking at a phonological grammar of English (Hammond 1999), a grammar of Canadian French (Walker 1984), a pattern I found myself by searching for possible violating words in the CMU pronouncing dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict), and a pattern mentioned by Albright 2008 and confirmed by a search in the CMU dictionary). The All of the patterns were then converted into constraints in the style of the UCLA phonotactic learner (Hayes and Wilson 2008). For instance, the English tendency to avoid words beginning with the sound [z] (Albright 2008:32) would be written as [+word_boundary][+voice,+anterior,+continuant].

Example (1) shows the constraints found in both languages, the method or source in which the constraint was found, and phonetic and typological justifications for the natural constraints.

---

[4] The 160 constraints induced for Hayes and White (2013) can be downloaded from Bruce Hayes' website at: http://www.linguistics.ucla.edu/people/hayes/PhonologicalNaturalness/Grammar.txt .

(1) Constraints Used

CANADIAN FRENCH

Natural Simple	Written with Segments[5]	Source	Phonetic Justification	Typological Justification[6]
*[+LABIAL][+consonantal,+LABIAL]	*[p,b,f,v,m,w][p,b,f,v,m]	UCLA Learner	dissimilation	Xhosa
*[-sonorant][+rhyme]	*[coda obstruent][coda]	Walker 1984:93-94	cluster reduction	Nigerian English
Natural Complex				
*[-continuant,+anterior][+high,-back]	*[t,d][i,I,y,Y]	Walker 1984:90-92	assibilation	Central Ojibwa
*[+round,+high][+round,+high]	*[u,w][u,w]	UCLA Learner	dissimilation	Kinyarwanda
Unnatural Simple				
*[+nasal,-anterior][+word_boundary]	*[ɲ]#	Walker 1984:98		
*[+sonorant,+LABIAL][-anterior]	*[m,w][tʃ,dʒ,ʃ,ʒ]	UCLA Learner		
Unnatural Complex				
*[+high,+tense][+voice,-continuant]	*[u,y,i][b,d,dʒ,g]	Walker 1984:53		
*[+round,+high][-high,-low,-back,+tense]	*[w,u][e]	UCLA Learner		

[5] Whenever a vowel class is listed for the French constraints, I only list the [–nasal] vowels in that class. Nasalized counterparts can be assumed.

[6] The languages listed in the "Typological Justification" column are from P-Base (Mielke 2008) unless otherwise noted and have phonological rules that represent similar patterns to the constraint of interest.

CANADIAN ENGLISH

Natural Simple	Written with Segments	Source	Phonetic Justification	Typ. Justification
*[+anterior][-back,-syllabic]	*[t,d,s,z][j]	Hammond 1999:52	assimilation	Ndyuka
*[-continuant,-rhyme][+continuant]	*[p,b,t,d,tʃ,dʒ,k,g][s,z,ʃ,ʒ,h]	Hammond 1999:51	Onset cluster ordering	Hindi (Morelli 1999 :42)
Natural Complex				
*[+consonantal][-back,-syllabic][-back]	*C[j][front vowels]	Hammond 1999:52	dissimilation	Afar
*[+round,-tense][-sonorant,+LABIAL,-rhyme]	*[ʊ][b,p in onset]	UCLA Learner	dissimilation	Kilivila (Kiriwana)
Unnatural Simple				
*[+consonantal,-anterior][-sonorant]	*[ʃ,ʒ,tʃ,dʒ][obstruents]	UCLA Learner		
*[+voice, -anterior, +continuant]	*[ʒ]	Hammond 1999:49		
Unnatural Complex				
*[+round,+tense][+voice,+DORSAL]	*[u,o][g]	CMU Dictionary		
*[+word_boundary][+voice,+anterior,+continuant]	*#[z]	Albright 2008:32[7]		

[7] Albright (2008) does not imply that English never has words beginning with [z], only that such words are rare. This rarity was enough for the UCLA Phonotactic learner to assign this constraint a weight of 2.615 which made it one of the more highly weighted English constraints used in this study. In addition to this, French has a much larger amount of words beginning with [z] (New et al. 2001), meaning that this rarity is not present in French.

4. METHODOLOGY

4.1 Stimuli

As in Prickett (2014) and Hayes and White (2013), novel words were created to test how well each of the constraints discussed in §3.3 were learned by speakers. These words were made to violate only one of the constraints that the UCLA Phonotactic learner found in the language they were designed for. An example of a violating word for the constraint *[+voice, -anterior, +continuant] is the word [ʒa] (which violates the constraint because of the consonant it begins with).

In addition to the violating words (which will be called *experimental* for the remainder of the paper), was a group of *control* words. Each of the control words had an experimental partner which it resembled closely—if possible these only differed by a single feature in a single segment within the words. For example, the control word partner for the example above was [dʒa], which only differs from [ʒa] in the value of the feature [continuant] in their initial segment. The control words were designed to not violate any constraints found by the UCLA Learner in the language they were made for.

The stimuli were presented to participants using orthography and audio recordings. The audio was recorded by a linguist who is a speaker of both Canadian English and Canadian French. The stimuli were presented to him in the form of IPA transcriptions (which can be seen in Example (2) below) and were divided into French and English stimuli. Other than the language the words were made for, no other knowledge of the experiment was communicated to the person making the recordings before the recordings were made. Praat was used to remove a small amount of background noise from a minority of the recordings (using the default settings on the “Remove noise…” function) and to scale the volume of all of the stimuli. The orthographic representations of the words were created by the same speaker of French and English that created the recordings (with some minor alterations based on my own judgments and feedback from my committee).

Example (2) shows the stimuli used in the experiment, according to which constraint each stimulus violates (or, in the case of the control words, which category of constraint its partner violates).

(2) Stimuli Used in the Experiment (transcribed in IPA)

FRENCH

Constraint	**Stimulus Type**			
Natural Simple	**Experimental**	**Control**	**Experimental**	**Control**
*[+LABIAL][+consonantal,+LABIAL]	tʊmp	tʊmt	kImp	kIlp
*[-sonorant][+rhyme]	gast	galt	vɛkt	vɛk
Natural Complex				
*[-continuant,+anterior][+high,-back]	tiv	tʃiv	dIf	dɛf
*[+round,+high][+round,+high]	wʊk	jʊk	wuʒ	woʒ
Unnatural Simple				
*[+nasal,-anterior][+word_boundary]	Iɲ	Iŋ	kʊɲ	kʊɲa
*[+sonorant,+LABIAL][-anterior]	komʃãn	komsãn	bemʃIn	benʃIn

Unnatural Complex				
*[+high,+tense][+voice,-continuant]	kug	kʊg	belyd	belyz
*[+round,+high][-high,-low,-back,+tense]	kwes	kjes	swes	swas

ENGLISH

Natural Simple	Experimental	Control	Experimental	Control
*[+anterior][-back,-syllabic]	sjak	swak	sjup	kjup
*[-continuant,-rhyme][+continuant]	psit	spit	ksot	skot
Natural Complex				
*[+consonantal][-back,-syllabic][-back]	vjim	vjum	kjeb	kweb
*[+round,-tense][-sonorant,+LABIAL,-rhyme]	tʊpən	tupən	kʊbət	kubət
Unnatural Simple				
*[+consonantal,-anterior][-sonorant]	ʃpig	ʃɹig	ʃkod	skod
*[+voice, -anterior, +continuant]	ʒa	dʒa	miʒ	miʃ
Unnatural Complex				
*[+round,+tense][+voice,+DORSAL]	pug	pud	tog	tag
*[+word_boundary][+voice,+anterior,+continuant]	zuk	suk	zib	dib

4.2 Materials

As in Prickett (2014), the experiment for this study was run online. A Perl CGI script was used with HTML in order to present participants with each stimulus, ask for and record participant ratings of each stimulus (see §4.4 for more details about this), and ask for and record demographic information for each participant. When presenting each stimulus, the audio recording of the word was played automatically when each page loaded and could be replayed by the participant if needed. Figure 3 and Figure 4 show screenshots of the webpages used to present stimuli.

Figure 3. Screenshot from the English section of the experiment.

Figure 4. Screenshot from the French section of the experiment.

There were two sections of the experiment: English and French (see §4.4 for a more detailed discussion of the two sections). Every participant took both sections, with the software counterbalancing the order of the sections so that the number of subjects that took the English section first was equal to the number of subjects that took the French section first. Within the two sections, the order of the stimuli was randomized using Perl’s “shuffle” function.

4.3 Participants

There were 30 participants who participated in the experiment and whose data could be used. Out of these participants, 10 spoke French as their first language and 20 spoke English as their first language. There were also 7 people who took the experiment and reported learning both French and English as their first language. Since this study was specifically looking at L2 learning, the data from these participants were not analyzed.

Out of the 30 participants with data that could be analyzed, three were recruited using Amazon's Mechanical Turk. Mechanical Turk has been shown to be a viable means of acquiring linguistic data (Sprouse 2011). In fact, the use of Mechanical Turk in Prickett (2014) garnered results that were not different in a significant way to those of Hayes and White (2013)—an experiment that was performed with the subjects and experimenter meeting in person in the conventional manner (see Prickett 2014:20 for a direction comparison of the values these two studies shared). Because of the limited amount of Canadians on Mechanical Turk, the rest of the experiment participants were recruited with emails and advertisements on social media.

4.4 Procedure

The first task that participants in the study performed was a verification of their knowledge of French. The verification page consisted of three audio recordings of French questions which participants had to listen to and then answer. The recordings were made by translating three English questions into French using Google Translate (www.google.com/translate) and then recording the synthesized production of the questions' French translation (which the Google translating tool automatically creates). This translation was rough, and was edited slightly based on recommendations from various French speakers.

If participants answered all three of the French questions correctly, they were directed to a page that gave them directions for the experiment. These directions were:

> "Languages have rules that determine how well words sound in that language. For example, in Canadian English "bzarshk" would sound very odd but "biss" would sound fine, even though neither of them are actual words. The following survey will contain two sections: a section for Canadian English and a section for Canadian French. For the made-up words in each section, rate numerically how acceptable they would be for that section's language. Each section will provide you with a baseline word to use as a standard of measurement. Please be sure to listen to the audio clips and read the written representations so that you know how each made-up word is supposed to be pronounced."[8]

After continuing, participants then began one of two sections: one in which participants were asked to rate the stimuli made for English and one in which they were asked to rate the stimuli made for French. To ensure that the participants were using the correct language in their judgments, the instructions for each section were written in that section's language (see Figure 3 and Figure 4 for the instructions given on each page). As mentioned above, the order that the two sections were presented in was counterbalanced across participants.

In each of the sections, participants were asked to judge the novel words described in §4.1 on how "good" they seemed for the language they were made for. The judgments were in the form of a numerical rating (or *magnitude estimation*). Magnitude estimation grammaticality judgments were used by Hayes and White (2013) and Prickett (2014), and have been shown to accurately

---

[8] All participants saw these initial instructions in English, however a summary of the instructions was given at the beginning of each section in that section's language.

record linguistic judgments (Bard et al.1996). Asking for L2 speakers to judge how acceptable a word would be in their L2 language has been used before in the literature (Pavlovskaya et al. 2013, for instance), but, to the best of my knowledge, has not been performed using magnitude estimation before this study.

In addition to the two groups of stimuli discussed in §4.1, there were three *diagnostic* stimuli for each of the sections in the experiment that were designed to ensure subjects were judging words based on the correct language. The diagnostic words violated many constraints in their section's language, but no constraints in the other section's language. For example, the word [pysjuʁ] was one of the diagnostic words in the English section. This word violates many constraints in English and should receive a very low rating from the participants. The diagnostic stimuli were presented randomly with the other stimuli, with no indicator that they were special in any way. The average rating for French diagnostic stimuli was 48.51 and the average rating for the English diagnostic stimuli was 25.46. Both of these were well below 100, which was the baseline value for words given in the experiment instructions, suggesting that participants did judge words according to the correct language.

After rating the stimuli in both sections, participants were asked information about when they learned French and English, along with standard demographic information such as gender and age.

5. RESULTS

5.1 Stimulus Ratings and Partnership Differences

As in Hayes and White (2013) and Prickett (2014), the natural log of the ratings given for the stimuli were averaged across participants. In Figure 5 and Figure 6, these mean log ratings for the individual stimuli are given.

In both Figure 5 and Figure 6, the lowest rated control stimulus is *ing* from the French portion of the experiment (meaning that it was perceived as being less grammatical than the other control stimuli). Since control stimuli do not violate any known constraint in the language they're made for, its low rating was not expected. The rating could have been a result of English loans in French often ending in the letters "ing" (Walker 1984). This orthographic tendency could have caused subjects to have viewed *ing* as sounding too foreign to receive a high rating. Orthography was also an issue in Prickett (2014) and is likely an unavoidable symptom of this particular methodology—however it is important to note that this problem is particularly apparent in this French stimulus.

Thirteen of the experimental stimuli in Figure 5 and twenty stimuli in Figure 6 have mean log ratings above 4—making them a relatively highly ranked (i.e. perceived as being highly grammatical) group. Most of these violate unnatural constraints, so given the results of Hayes and White (2013), these high ratings can be expected. However, the Natural Simple French Experimental stimuli have high ratings from both groups of participants, which was not predicted (since Natural Experimental stimuli violate natural constraints—a category of constraints that should have strong effects based on Hayes and White 2013). In addition, the L1-English speakers rated most of the Natural Complex French Experimental stimuli higher than many of the control stimuli. These high ratings for stimuli that violate Natural French constraints was not predicted and could represent problems with the constraints chosen or with the stimuli that were made for the French section of the experiment.

Figure 7 and Figure 8 show the mean log ratings of the experimental stimuli subtracted from the mean log ratings of their control stimuli partners. These values are expected to be positive, since the control stimuli should have been rated higher than those that were experimental. The Natural English categories tend to have large, positive differences (which were expected, based on Hayes and White 2013). The Natural French categories' differences are not as high as expected. This is especially apparent in Figure 8, where the L1-English speakers give a majority of the French Experimental stimuli higher ratings than their control group partners.

Many of the largest positive differences in these graphs are in the English partnerships, for both groups of speakers. Higher differences in Figure 7 and Figure 8 should mean better constraint learning—however, it is not likely that L1-French speakers know English constraints better than those of their native language. This further suggests that there could have been issues with the French constraints or stimuli. Possible sources for these issues are discussed in §5.4.

5.2 Constraint Effects

In Examples (1) and (2) the effects of individual constraints are shown. These effects represent the average of the partnership differences discussed above for each constraint. Like the partnership differences, the constraint effects were expected to be positive (since this would mean that the control stimuli were rated higher than the experimental stimuli). The French constraints, especially for the L1-English speakers, have a large amount of negative constraint effects (meaning that the experimental stimuli were perceived as more grammatical than the control ones). Of these negative French constraint effects, three are for natural constraints—which would be expected to be positive based on Hayes and White (2013). Only one of the English natural constraint effects is negative—and it's only negative for the L1-French speakers. These negative values in the French constraints are a result of the unexpectedly low ratings given to French experimental stimuli discussed in §5.1. In order to compare these constraint effects to one another, and to test the broader predictions about second language learning biases presented in §3.1, a more sophisticated analysis will need to be applied. This analysis will be discussed in §(1).

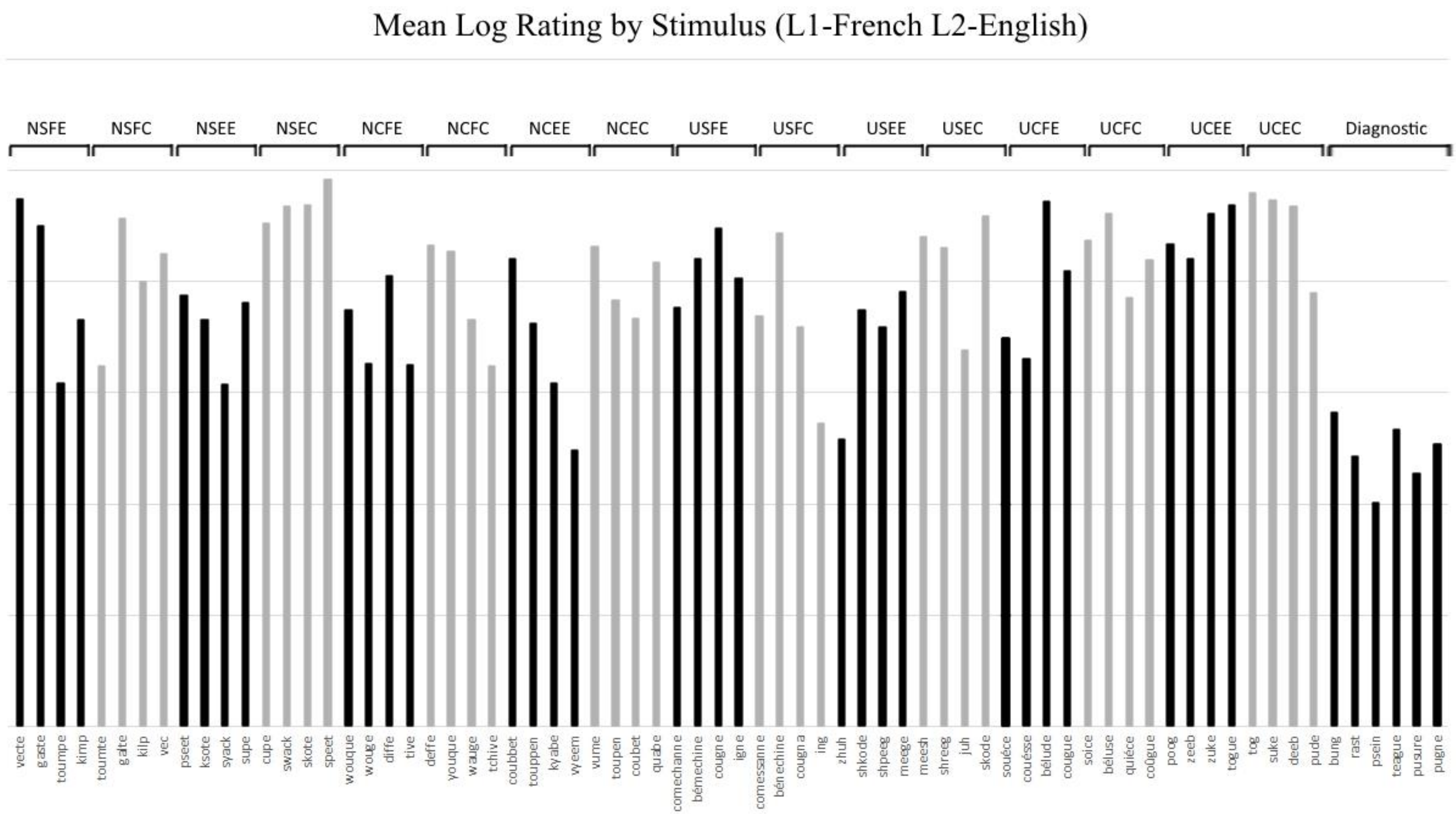

Figure 5. *Mean log ratings of individual stimuli given by L1-French L2-English speakers. Stimuli are grouped according to the category of constraint that either they or their partner violates, and then by whether they are control or experimental. Categories are labeled as follows: NSFE (Natural Simple French Experimental), NSFC (Natural Simple French Control), NSEE (Natural Simple English Experimental), NSEC (Natural Simple English Control), NCFE (Natural Complex French Experimental), NCFC (Natural Complex French Control), NCEE (Natural Complex English Experimental), NCEC (Natural Complex English Control), USFE (Unnatural Simple French Experimental), USFC (Unnatural Simple French Control), USEE (Unnatural Simple English Experimental), USEC (Unnatural Simple English Control), UCFE (Unnatural Complex French Experimental), UCFC (Unnatural Complex French Control), UCEE (Unnatural Complex English Experimental), and UCEC (Unnatural Complex English Control). In addition to these, the ratings for diagnostic stimuli (discussed in §4.4) are given.*

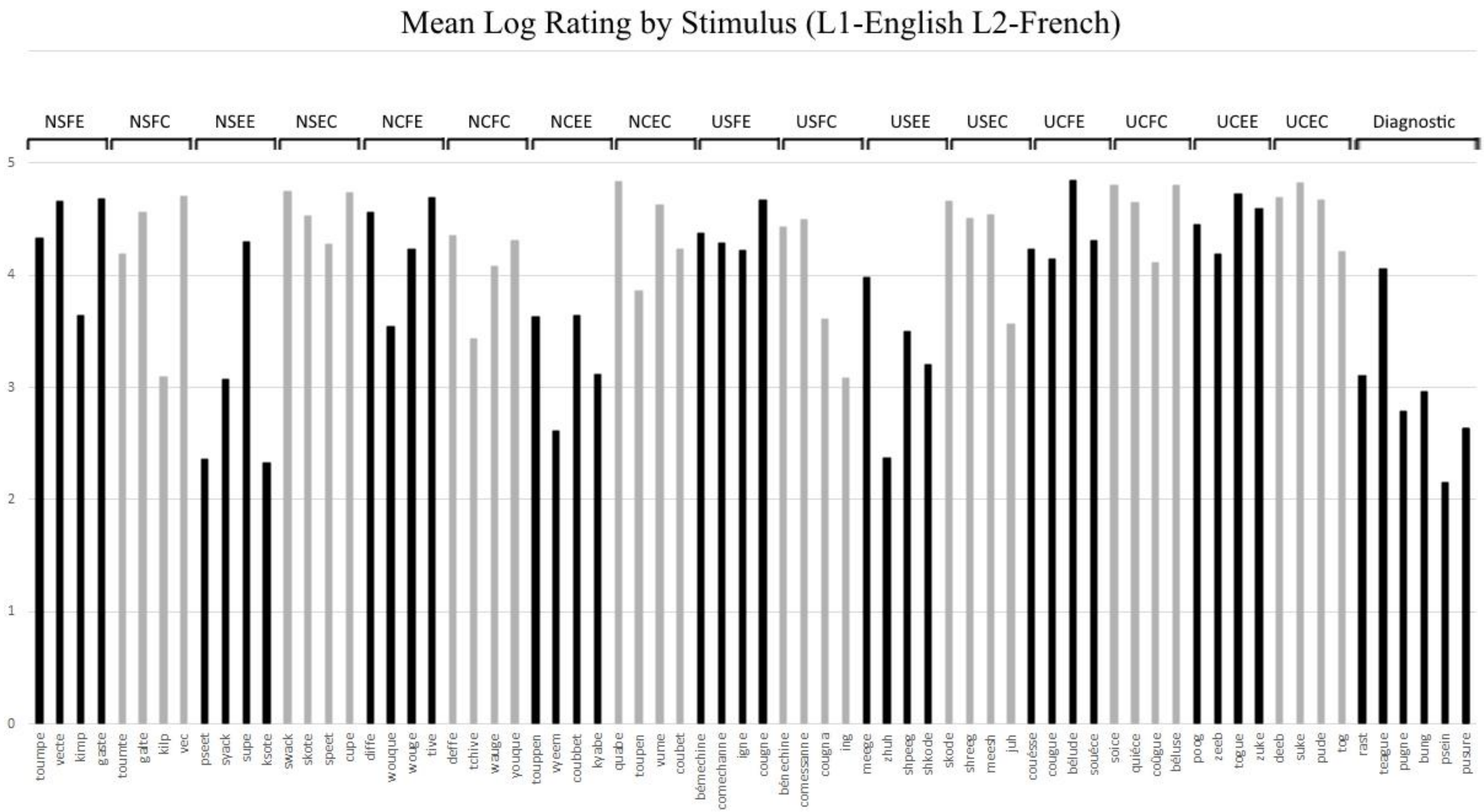

Figure 6. Mean log ratings of individual stimuli given by L2-French L1-English speakers. Stimuli are grouped according to the category of constraint that either they or their partner violates, and then by whether they are control or experimental. Categories are labeled as follows: NSFE (Natural Simple French Experimental), NSFC (Natural Simple French Control), NSEE (Natural Simple English Experimental), NSEC (Natural Simple English Control), NCFE (Natural Complex French Experimental), NCFC (Natural Complex French Control), NCEE (Natural Complex English Experimental), NCEC (Natural Complex English Control), USFE (Unnatural Simple French Experimental), USFC (Unnatural Simple French Control), USEE (Unnatural Simple English Experimental), USEC (Unnatural Simple English Control), UCFE (Unnatural Complex French Experimental), UCFC (Unnatural Complex French Control), UCEE (Unnatural Complex English Experimental), and UCEC (Unnatural Complex English Control). In addition to these, the ratings for diagnostic stimuli (discussed in §4.4) are given.

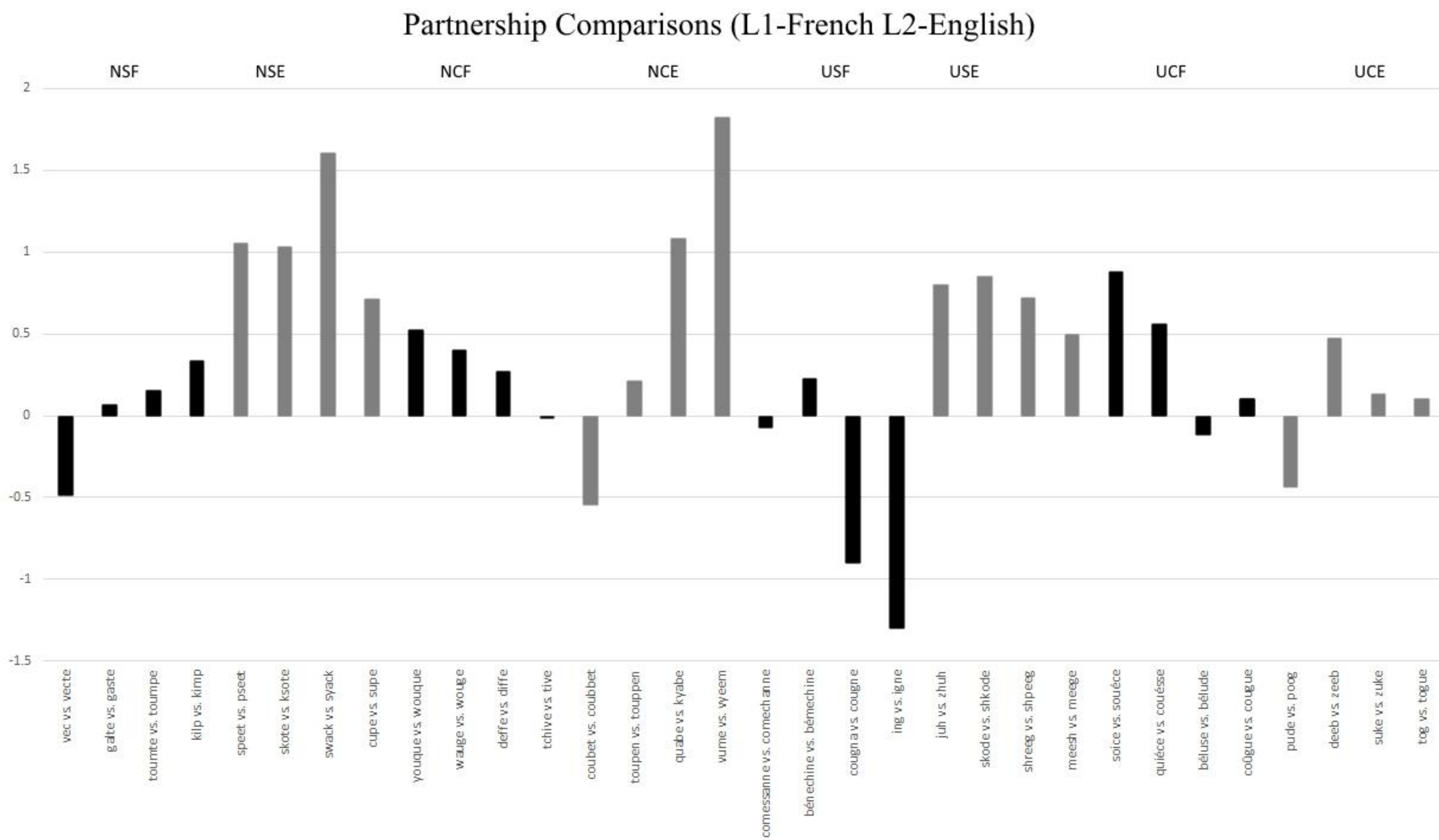

Figure 7. *Differences between the mean log ratings given by L1-French L2-English speakers of control and experimental partner stimuli. The category of each partnership is shown at the top of the graph. Abbreviations as follows: NSF (Natural Simple French), NSE (Natural Simple English), NCF (Natural Complex French), NCE (Natural Complex English), USF (Unnatural Simple French), USE (Unnatural Simple English), UCF (Unnatural Complex French), and UCE (Unnatural Complex English).*

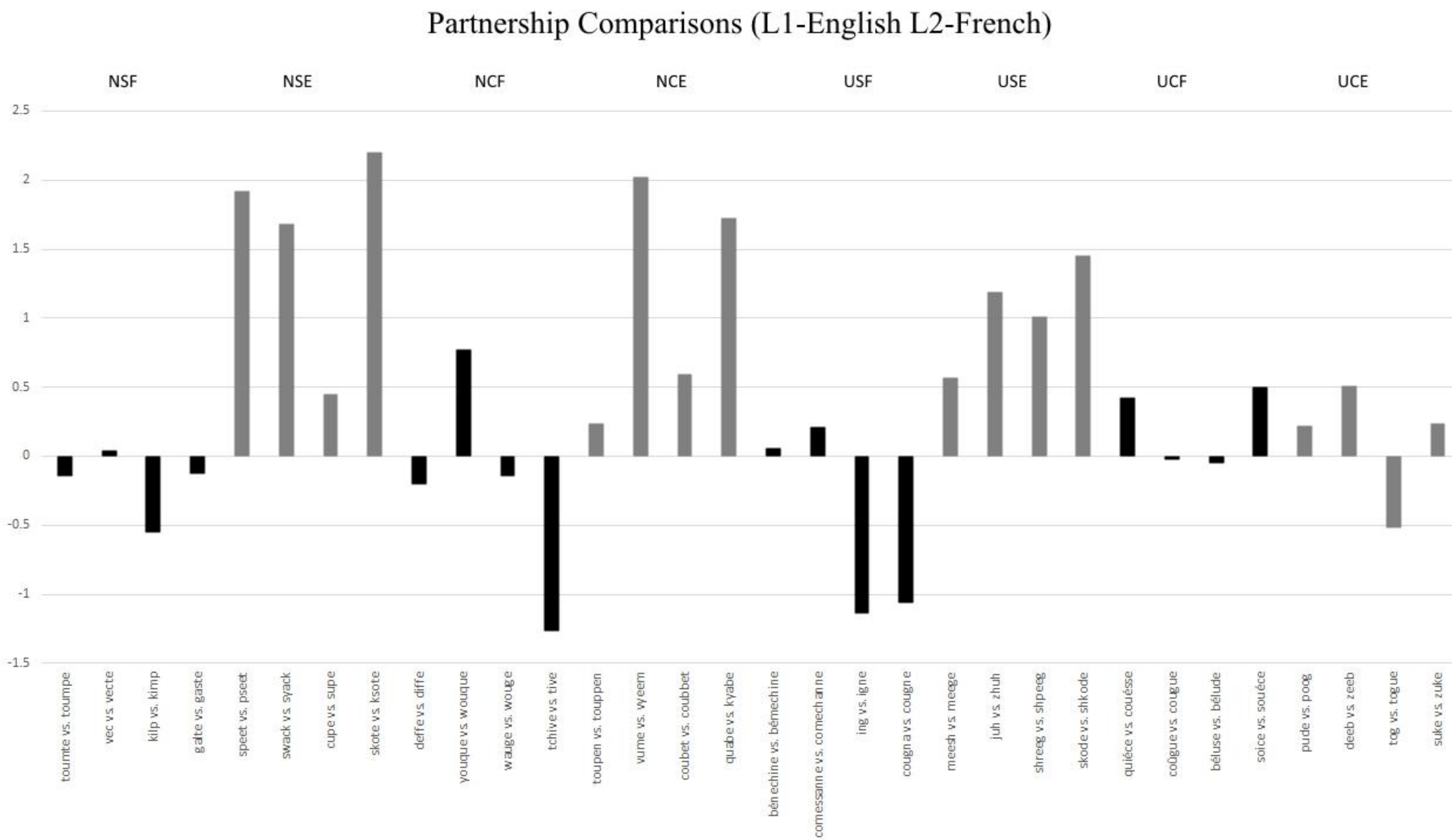

Figure 8. *Differences between the mean log ratings given by L1-English L2-French speakers of control and experimental partner stimuli. The category of each partnership is shown at the top of the graph. Abbreviations as follows: NSF (Natural Simple French), NSE (Natural Simple English), NCF (Natural Complex French), NCE (Natural Complex English), USF (Unnatural Simple French), USE (Unnatural Simple English), UCF (Unnatural Complex French), and UCE (Unnatural Complex English).*

(1) L1-French L2-English Constraint Effect Sizes

CANADIAN FRENCH

Constraint	Effect Size
Natural Simple	
*[+LABIAL][+consonantal,+LABIAL]	0.24486533
*[-sonorant][+rhyme]	-0.212243274
Natural Complex	
*[-continuant,+anterior][+high,-back]	0.126160716
*[+round,+high][+round,+high]	0.4609784
Unnatural Simple	
*[+nasal,-anterior][+word_boundary]	-1.098865949
*[+sonorant,+LABIAL][-anterior]	0.074683611
Unnatural Complex	
*[+high,+tense][+voice,-continuant]	-0.005735764
*[+round,+high][-high,-low,-back,+tense]	0.719352746

CANADIAN ENGLISH

Constraint	Effect Size
Natural Simple	
*[+anterior][-back,-syllabic]	1.158573908
*[-continuant,-rhyme][+continuant]	1.04105884
Natural Complex	
*[+consonantal][-back,-syllabic][-back]	1.451928466
*[+round,-tense][-sonorant,+LABIAL,-rhyme]	-0.165517078
Unnatural Simple	
*[+consonantal,-anterior][-sonorant]	0.782798316
*[+voice, -anterior, +continuant]	0.646389805
Unnatural Complex	
*[+round,+tense][+voice,+DORSAL]	-0.166837649
*[+word_boundary][+voice,+anterior,+continuant]	0.301102052

(2) L1-English L2-French Constraint Effect Sizes

CANADIAN FRENCH

Constraint	Effect Size
Natural Simple	
*[+LABIAL][+consonantal,+LABIAL]	-0.34255
*[-sonorant][+rhyme]	-0.04125
Natural Complex	
*[-continuant,+anterior][+high,-back]	-0.73229
*[+round,+high][+round,+high]	0.314489
Unnatural Simple	
*[+nasal,-anterior][+word_boundary]	-1.09655
*[+sonorant,+LABIAL][-anterior]	0.13018
Unnatural Complex	
*[+high,+tense][+voice,-continuant]	-0.03751
*[+round,+high][-high,-low,-back,+tense]	0.456005

CANADIAN ENGLISH

Constraint	Effect Size
Natural Simple	
*[+anterior][-back,-syllabic]	1.062756
*[-continuant,-rhyme][+continuant]	2.058179
Natural Complex	
*[+consonantal][-back,-syllabic][-back]	1.871569
*[+round,-tense][-sonorant,+LABIAL,-rhyme]	0.409579
Unnatural Simple	
*[+consonantal,-anterior][-sonorant]	1.226938
*[+voice, -anterior, +continuant]	0.876576
Unnatural Complex	
*[+round,+tense][+voice,+DORSAL]	-0.14873
*[+word_boundary][+voice,+anterior,+continuant]	0.366199

5.3 Statistical Analysis

In Examples (3) and 0, an analysis of the experimental data is given. The values in these tables are from a linear model for the mean log ratings of the stimuli, with fixed effects for Natural versus Unnatural, Simple versus Complex, French versus English, and Experimental versus Control. Standard errors were adjusted for multiple observations within subjects[9].

Values (a-h) in these examples represent the effect sizes of each of the constraint categories. The category effect sizes for this analysis were found by subtracting the mean log ratings of the control stimuli in each category from their experimental group partners (see §4.1 for more on these two groups). This can be written mathematically as:

Category Effect = (experimental - control)

The difference in grammaticality ratings between the experimental and control words should indicate how well speakers learned the constraints (when they were learning English and French) that the experimental stimuli violated. Specifically the difference should be negative for this calculation, since the experimental stimuli should have been rated lower than the control stimuli. The remainder of each table shows comparisons between these effect sizes.

The next step in the analysis was measuring the difference between L1 and L2 constraint effects. This can be seen in the values labeled (i-l) in each example. These values represent the effect of a constraint being in a speaker's first language, as opposed to a constraint being in the

[9] In order to perform these calculations, the procedure "GENMOD" from the software SAS was used with the "REPEATED" statement.

speaker's second language (which will be called the constraint category's *L1 Effect*). This can be represented as[10]:

$$L1\ Effect = (experimental - control)_{English} - (experimental - control)_{French}$$

This L1 effect could then be compared across the four different categories of constraints (Natural Simple, Natural Complex, Unnatural Simple, Unnatural Complex) to find the effects of the variables of interest in gauging the accuracy of the predictions discussed in §3.1. In that section, it was predicted that naturalness would effect L1 learning the most, while complexity would effect L2 learning the most. This means that the L1 effect of naturalness should go in the opposite direction, when compared across categories (e.g. the categories Natural Complex and Unnatural Complex), of the L1 effect of complexity. The values labeled (m-p) show these comparisons. They can be illustrated by:

$$\begin{aligned} Variable\ Effect\ Across\ Categories = &((experimental - control)_{English} - (experimental - control)_{French})_{Category\ 1} \\ &- ((experimental - control)_{English} - (experimental - control)_{French})_{Category\ 2} \end{aligned}$$

By specifically looking at the values in (o) and (p), one can begin to see the different effects of naturalness and complexity in L1 and L2 learning. In (o), the L1 effect of the Unnatural Simple category is compared to the L1 effect of the Unnatural Complex category. By comparing the two unnatural categories, any effects of naturalness are avoided (which might have been expected inflate the effect sizes, based on Hayes and White 2013 and Prickett 2014). Instead, the effect of complexity is isolated. For French-L1 English-L2 speakers, (Example (3)), a significantly negative value for (o) would indicate that complexity has a stronger effect in L2

[10] It's important to note that the ordering of the languages stayed consistent in all calculations. Meaning that the L1 effects in French ought to go in the opposite direction of the L1 effects in English.

learning (the predicted outcome). A significantly positive value for (o) in Example (3) would indicate that complexity has a stronger effect in L1 learning for those speakers (which is not the predicted outcome). The opposite is true for English-L1 French-L2 speakers (Example 0), since the L1 and L2 values are switched for those comparisons in the effect calculations. So a positive value for (0o) would be indicative of a higher effect of complexity in L2 learning (which is predicted) and a negative value for (o) in Example 0 would mean a higher effect of complexity on L1 learning (which is not predicted).

The opposite process can be used to isolate naturalness by examining the value in (p). This value is the difference between the L1 effect of the Natural Complex category and the L1 effect of the Unnatural Complex category. By comparing across the complex groups, effects of complexity are avoided (which might have been expected to inflate effect sizes, based on Skoruppa and Peperkamp 2011). Instead, we see an isolated effect of naturalness. For French-L1 English-L2 (Example (3)), a significantly positive value for (p) in would indicate that naturalness has a stronger effect in L1 learning for those speakers (which is the predicted outcome). A significantly negative value for ((3)p) would indicate that naturalness has a stronger effect in L2 learning (which is not the predicted outcome). And the opposite is again true for English-L1 French-L2 speakers (Example 0). So a negative value for (p) in Example 0 would mean a higher effect of naturalness on L1 learning (which is predicted) and a positive value for (0t) would be indicative of a higher effect of naturalness in L2 learning (which is not predicted).

In (3), the values for (o) and (p) are shown for L1-French L2-English speakers. Since the value of (3) is significantly below zero, it seems that complexity *does* have a stronger effect for these speakers in their L2, which was predicted. However, (3) does not go in the predicted direction. This value is also significantly below zero, meaning that naturalness *also* has a stronger effect in the speakers' L2.

Example 0 shows the effect values for the L1-English L2-French speakers. Since the value of 0 is significantly below zero, it seems that complexity has more of an effect in these subjects' L1, which was not predicted. The value of 0, does go in the predicted direction. This value is also significantly below zero, meaning that naturalness *does* have a stronger effect in the speakers' L1 (which was the predicted outcome for that variable).

5.4 Interpretation

The data discussed in §(1) does not seem to lend itself to any hypothesis, since the data for the L1-French L2-English speakers (which showed predicted values for complexity, but not for naturalness) goes in the opposite direction of the data for the L1-English L2-French speakers (which showed predicted values for naturalness, but not for complexity).

Part of the problem could have been issues in the French section of the experiment, since both groups of speakers seem to have stronger biases for the English constraints. It was evident when looking at the individual French stimuli ratings (discussed in §5.1) that generally the French stimuli were rated in unexpected ways. This tendency then caused unpredicted French partnership differences and unpredicted French constraint effects (discussed in §5.2).

One factor causing these issues could have been that the French experiment was made to specifically test Canadian French. While the experiment was only run on Canadians, there still could have been some subjects who were influenced by other dialects of the language. Many French speakers in Canada use a dialect more similar to Parisian French (Walker 1984) and often these pronunciation differences are influenced by extralinguistic factors such as socioeconomic status (Brent 1970). One of the constraints that had a negative effect size in both speaker groups was "*[+nasal,-anterior][+word_boundary]". This constraint represents Canadian French speakers' tendency to velarize word final palatal nasals—a pattern that is not found in Parisian French. The constraint "*[-continuant,+anterior][+high,-back]" (another pattern only found in Canadian French) also had a negative effect, but only in the L1-English group. One last piece of evidence suggesting dialect interference was the fact that the diagnostic stimulus *teague*, which would be ungrammatical only in Canadian French (due in part to the constraint "*[-continuant,+anterior][+high,-back]" shown above; see Walker 1984), had one of the highest mean ratings of all the diagnostic stimuli in both groups[11].

Other possible reasons for the French constraints being less effective seem unlikely. For instance, constraint weight is a measure by the UCLA phonotactic learner of how likely a constraint is to affect a speaker's judgments or productions (Hayes and Wilson 2008). Constraints with higher weights would be expected to have stronger effect sizes. However, the average weight that the UCLA learner assigned to the English constraints tested in this experiment[12] was 2.43, while the average weight for the French constraints was 4.55. This would suggest that the English constraints ought to have smaller effect sizes and does not explain the weaker French constraints.

Another possibility could be a difference in the type of features used in the English and French constraints. The UCLA Phonotactic Learner treats prosodic features like "[word_boundary]" and "[rhyme]" as if they were the same as phonetic features like "[voice]"

[11] In fact, the French diagnostic stimuli had the highest average ratings in general. See §4.4 for the mean ratings of the two groups of diagnostic stimuli.

[12] It's important to note that, as stated in §3.3, not all of the constraints were found using the UCLA Phonotactic Learner. However, after the constraints were all found, they were weighted together by the learner.

and “[anterior]”. An imbalance of prosodic or phonetic features in one of the languages might have caused a difference in constraint effects—but this seems unlikely since there is only one more prosodic feature used in the English constraints than there is in the French constraints.

If the results for the French portion of the experiment are ignored, it seems that the speakers in this experiment have more of both biases in their L1—regardless of whether the bias is one for simple phonotactic patterns or for natural ones. This could mean that when learning a second language, speakers learn most patterns equally, without much bias causing them to learn certain patterns better. The exception to this seems to be constraints in the Unnatural Complex category. This category’s L1 Effect was low for both L1-French L2-English and L1-English L2-French speakers, meaning that for everyone in the experiment there wasn’t much effect of these constraints in either the L1 or the L2 grammaticality judgments. It seems that the patterns that are most ignored in the L2 are those that are both complex and unnatural.

(3) Table Showing Results of L1-French L2-English Speakers[13]

	Calculation	Estimate	Mean Lower CL	Mean Upper CL	Std Err	ChiSq	Prob ChiSq	Description	Predictions
a	NS-L1 (Experimental – Control)	0.0912	-0.0243	0.2067	0.0589	2.39	0.1219	Effect of NSF constraints	-
b	NS-L2 (Experimental – Control)	-0.4708	-0.6795	-0.2622	0.1064	19.56	<.0001	Effect of NSE constraints	-
c	NC-L1 (Experimental – Control)	0.1182	-0.1384	0.3748	0.1309	0.82	0.3666	Effect of NCF constraints	-
d	NC-L2 (Experimental – Control)	-0.3756	-0.5349	-0.2163	0.0813	21.35	<.0001	Effect of NCE constraints	-
e	US-L1 (Experimental – Control)	0.1024	-0.1598	0.3646	0.1338	0.59	0.4441	Effect of USF constraints	-
f	US-L2 (Experimental – Control)	-0.3243	-0.5096	-0.1389	0.0946	11.76	0.0006	Effect of USE constraints	-
g	UC-L1 (Experimental – Control)	-0.1497	-0.3312	0.0317	0.0926	2.62	0.1058	Effect of UCF constraints	-
h	UC-L2 (Experimental – Control)	-0.1832	-0.3425	-0.0238	0.0813	5.07	0.0243	Effect of UCE constraints	-
i	NS (L2 – L1)	-0.5620	-0.7933	-0.3307	0.1180	22.68	<.0001	Effect of L2 vs. L1 on NS Effect	
j	NC (L2 – L1)	-0.4938	-0.8180	-0.1695	0.1654	8.91	0.0028	Effect of L2 vs. L1 on NS Effect	+
k	US (L2 – L1)	-0.4266	-0.8071	-0.0462	0.1941	4.83	0.0280	Effect of L2 vs. L1 on US Effect	-
l	UC (L2 – L1)	-0.0334	-0.2465	0.1796	0.1087	0.09	0.7584	Effect of L2 vs. L1 on UC Effect	
m	NS (L2 – L1) - US (L2 – L1)	-0.1353	-0.5149	0.2442	0.1936	0.49	0.4846	Effect of naturalness on L1 Effect (Simple)	
n	NS (L2 – L1) – NC (L2 – L1)	-0.0682	-0.4301	0.2937	0.1846	0.14	0.7119	Effect of complexity on L1 Effect (Natural)	
o	US (L2 – L1) – UC (L2 – L1)	-0.3932	-0.7565	-0.0299	0.1854	4.50	0.0339	Effect of complexity on L1 Effect (Unnatural)	+
p	NC (L2 – L1) – UC (L2 – L1)	-0.4603	-0.9206	-0.0001	0.2348	3.84	0.0499	Effect of naturalness on L1 Effect (Complex)	-

[13] N = *natural*, U = *unnatural*, S = *simple*, C = *complex*, F = *French*, E = *English*. In the column labeled “Predictions”, a (-) denotes a value that was predicted to be significantly below zero, while a (+) denotes a value that was predicted to be significantly above zero.

Table Showing Results of L1-English L2-French Speakers[14]

	Calculation	Estimate	Mean Lower CL	Mean Upper CL	Std Err	ChiSq	Prob ChiSq	Description	Predictions
a	NS-L2 (Experimental – Control)	0.0858	-0.0009	0.1725	0.0442	3.77	0.0523	Effect of NSF constraints	-
b	NS-L1 (Experimental – Control)	-1.0384	-1.3440	-0.7328	0.1559	44.36	<.0001	Effect of NSE constraints	-
c	NC-L2 (Experimental – Control)	0.2115	0.0332	0.3898	0.0910	5.41	0.0201	Effect of NCF constraints	-
d	NC-L1 (Experimental – Control)	-0.8844	-1.1295	-0.6393	0.1251	50.02	<.0001	Effect of NCE constraints	-
e	US-L2 (Experimental – Control)	0.3831	0.2179	0.5483	0.0843	20.66	<.0001	Effect of USF constraints	-
f	US-L1 (Experimental – Control)	-0.7998	-1.0254	-0.5743	0.1151	48.29	<.0001	Effect of USE constraints	-
g	UC-L2 (Experimental – Control)	-0.1525	-0.2520	-0.0531	0.0508	9.03	0.0027	Effect of UCF constraints	-
h	UC-L1 (Experimental – Control)	-0.0518	-0.1553	0.0517	0.0528	0.96	0.3265	Effect of UCE constraints	-
i	NS (L1 – L2)	-1.1242	-1.4801	-0.7684	0.1816	38.34	<.0001	Effect of L1 vs. L2 on NS Effect	
j	NC (L1 – L2)	-1.0960	-1.3805	-0.8115	0.1452	57.00	<.0001	Effect of L1 vs. L2 on NS Effect	-
k	US (L1 – L2)	-1.1829	-1.5140	-0.8518	0.1689	49.03	<.0001	Effect of L1 vs. L2 on US Effect	+
l	UC (L1 – L2)	0.1007	-0.0385	0.2400	0.0711	2.01	0.1563	Effect of L1 vs. L2 on UC Effect	
m	NS (L1 – L2) - US (L1 – L2)	0.0587	-0.2889	0.4063	0.1773	0.11	0.7406	Effect of naturalness on L1 Effect (Simple)	
n	NS (L1 – L2) – NC (L1 – L2)	-0.0282	-0.4395	0.3830	0.2098	0.02	0.8929	Effect of complexity on L1 Effect (Natural)	
o	US (L1 – L2) – UC (L1 – L2)	-1.2837	-1.6712	-0.8961	0.1977	42.15	<.0001	Effect of complexity on L1 Effect (Unnatural)	-
p	NC (L1 – L2) – UC (L1 – L2)	-1.1967	-1.5122	-0.8812	0.1610	55.27	<.0001	Effect of naturalness on L1 Effect (Complex)	+

[14] N = *natural*, U = *unnatural*, S = *simple*, C = *complex*, F = *French*, E = *English*. In the column labeled "Predictions", a (-) denotes a value that was predicted to be significantly below zero, while a (+) denotes a value that was predicted to be significantly above zero.

6. DISCUSSION

6.1 Comparison to Past Work

The L1-English speakers' results for the English portion of this experiment could be viewed as a kind of replication of the experiments performed by Hayes and White (2013) and Prickett (2014) with different participants and with a slightly different procedure. If a comparison like this is made, many of the conclusions of those past studies are reinforced.

First of all, the Natural English constraint categories both had effect sizes significantly below zero (see Example 0b and 0d). This means that the difference between the mean log ratings of the Natural Experimental stimuli and the Natural Control stimuli for English was significant. This kind of difference between control and experimental stimuli was used by Hayes and White (2013) and Prickett (2014) to demonstrate a bias for natural constraints and the results from this study suggest the same kind of bias.

However, this study doesn't exactly match the findings of Prickett (2014). The L1-English speakers also had a significant effect for the Unnatural Simple category of constraints. Prickett (2014) did not find a significant difference between the control and experimental stimuli in this category and concluded that this was indicative of a lack of a complexity bias in phonological learning. This experiment's effect size for Unnatural Simple English constraints, seen in 0, is significantly below zero—suggesting that this may not be this case. This category's effect is weaker than the natural categories, but not significantly so.

Overall, the similarity to past work seems to support the idea that the English results are the more accurate ones in this study, and that the French results could be skewed by factors that were not controlled for in the experiment.

6.2 Future Work

Future work could help decode some of the issues in the data that this experiment found. For instance, looking specifically at L1 interference on L2 phonotactic learning could help to discover whether controlling for interference in this experiment was even necessary. If interference was not an issue, than many of the variables that might have confounded the results of this experiment could have been avoided. For example, the constraints used in the experiment could have limited to only those found by the UCLA Phonotactic Learner (Hayes and Wilson 2008). Having all of the constraints be induced by the learner could have allowed for more control over the constraints' weight (as mentioned in §5.4; see Prickett 2014 for an example of how constraint weight can be more carefully controlled for if there are less requirements of the constraints in an experiment).

Another way to avoid the issues that come with controlling for interference would be to pick languages with very dissimilar phonologies. French and English were chosen for this experiment because speakers of these languages would be easy to find. But if more resources for finding and surveying speakers were available, there might be a better choice of languages to test. When choosing languages, it could also be helpful to choose a language without the same dialectal variety of French. As mentioned above, the differences between "standard" French and Canadian French could have caused some issues with the French results of the experiment.

Another part of the experiment that could be changed in future research is the use of magnitude estimation for grammaticality judgments. Although this method worked for other studies (such as Hayes and White 2013 and Prickett 2014), it might have been too fragile of a metric for this experiment. With as many variables that were at play (i.e. L1 interference, constraint quality, stimuli quality, etc.), a form of surveying that did not allow participants as much variability could be beneficial. In addition to this, several participants remarked on the difficulty of a magnitude estimation judgment and this difficulty likely caused many to abandon the experiment before finishing it. A simpler task—such as a sliding scale from 1 to 10—might encourage more participation from possible participants.

6.3 Conclusions

In §3.1, it was predicted that L1 learning would have more of a naturalness bias than L2 learning and that L2 learning was more likely to have a bias for complexity. The data found in this study suggests that there is less bias in L2 learning generally, and that L1 learning has more of a bias for naturalness and for simplicity in the patterns being learned by speakers.

These results do not contradict the idea that L1 learning is affected by naturalness because of phonetic grounding (Hayes 1999), but it does not shed light on why studies like Skoruppa and Peperkamp (2011) have failed to find this naturalness bias. Recent research has begun to explore this issue, suggesting that subjects not being able to sleep in between learning sessions could be a factor making artificial language different than natural language learning (Peperkamp and Martin 2015).

More work will need to be done on second language phonotactic learning before any solid conclusions can be made about differences between L1 and L2 learning biases. But in the meantime, this study's findings and exploration of methodological techniques creates a solid foundation for future exploration into the subject.

REFERENCES

Bard, Ellen Gurman, Dan Robertson, and Antonella Sorace (1996). Magnitude estimation of linguistic acceptability. *Language*, 32-68.

Becker, Michael, Nihan Ketrez, and Andrew Nevins (2011). The Surfeit of the Stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87, 84–125.

Brent, E. (1971). *Canadian French: a synthesis*. Phd Dissertation. Cornell University.

Bulgarella, Rosaria, and E. James Archer (1962). Concept identification of auditory stimuli as a function of amount of relevant and irrelevant information. *Journal of Experimental Psychology* 63.3: 254.

Chomsky, Noam, and Morris Halle (1968). *The Sound Pattern of English*. New York: Harper & Row.

Hayes, Bruce (1999). Phonetically driven Phonology. *Functionalism and Formalism in Linguistics* 1: 243-285.

Hayes, Bruce, and James White (2013). Phonological Naturalness and Phonotactic Learning. *Linguistic Inquiry* 44.1: 45-75.

Hayes, Bruce, and Colin Wilson (2008). A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry* 39.3: 379-440.

Kilpatrick, Cynthia D (2009). The acquisition of ungrammaticality: learning a subset in L2 phonotactics.

Lenneberg, Eric H. (1967). *Biological Foundations of Language* (Vol. 68). New York: Wiley.

Long, Mike (2005). Problems with supposed counter-evidence to the Critical Period Hypothesis. *International Review of Applied Linguistics in Language Teaching*, 43(4), 287-317.

Morelli, Frida (1999). The phonotactics and phonology of obstruent clusters in Optimality Theory (Doctoral dissertation).

Moreton, Elliott, and Joe Pater (2012). Structure and substance in artificial-phonology learning. Part I, Structure. Part II, Substance. *Language and Linguistics Compass* 6 (11): 686-701 and 702-718.

New, Boris, Christophe Pallier, Ludovic Ferrand, and Rafael Matos (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE, L'Année Psychologique, 101, 447-462. http://www.lexique.org

Pavlovskaya, Natalia, Samawal Jarad, Alex Ho-Cheong Leung, and Martha Young-Scholten (2013). What do adult L2 learners know about phonology after minimal exposure? *Opening New Lines of Communication in Applied Linguistics*: 393.

Peperkamp, Sharon and Alexander Martin (2015). Phonological rule learning and consolidation with sleep. The 37th Annual Meeting of the Deutsche Gesellschaft für Sprachwissenschaft (DGfS).

Prickett (2014). The Effect of Complexity versus the Effect of Naturalness on Phonological Learning. Carolina Digital Repository.

Prince, Alan, and Paul Smolensky (1993/2004). Optimality Theory: Constraint interaction in generative grammar. Technical report, Rutgers University, New Brunswick, NJ, and University of Colorado, Boulder. [ROA-537] Revised version, Oxford: Blackwell (2004).

Shepard, Roger N., Carl I. Hovland, and Herbert M. Jenkins (1961). Learning and memorization of classifications. *Psychological Monographs*: 75 (13, Whole No. 517).

Skoruppa, Katrin, and Sharon Peperkamp (2011). Adaptation to novel accents: feature-based learning of context-sensitive phonological regularities. *Cognitive Science*: 35, 348-366.

Sprouse, Jon (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior research methods*: 43(1): 155-167.

Trapman, Mirjam, and René Kager (2009). The acquisition of subset and superset phonotactic knowledge in a second language. *Language Acquisition*: 16(3), 178-221.

Walker, Douglas C. (1984). *The pronunciation of Canadian French*. Ottawa, ON: University of Ottawa Press.

Weber, Andrea, and Anne Cutler (2006). First-language phonotactics in second-language listening. *The Journal of the Acoustical Society of America*: 119(1), 597-607.