UNIVERSITY OF CALIFORNIA

Los Angeles

Vowel and Consonant Lengthening

in Finnish Loanword Adaptation

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Arts

in Linguistics

by

Margaret Ilona Kroll

2014

ABSTRACT OF THE THESIS

Vowel and Consonant Lengthening

in Finnish Loanword Adaptation

by

Margaret Ilona Kroll

Master of Arts in Linguistics

University of California, Los Angeles 2014

Professor Bruce Hayes, Chair

This project looks at loanword adaptation in Finnish, focusing on the role of perception in the adaptation of voiceless stops as geminates. In the Finnish loanword process, non-geminate donor voiceless stops are often adapted in geminated form. This particular adaptation strategy is interesting because it serves a purpose other than to correct a phonological ill-formedness in the borrower language, meaning that it cannot be accounted for by appealing to native Finnish structural constraints. This project will argue that, while there is evidence that word-structure constraints play a role in the adaptation process, gemination is also influenced by perceptual cues. Specifically, I will show using on-line adaptation data that geminates are adapted as such because of two perceptual factors: stress placement in the donor and loan word and the existence of subsequent stops in the donor word.

The thesis of Margaret Ilona Kroll is approved.

Robert Daland

Kie Zuraw

Bruce Hayes, Committee Chair

University of California, Los Angeles

2014

Table of Contents

List of Tables

Preamble. Loanword adaptation is particularly interesting to the linguist because it offers insights into language not accessible within native phonologies. Many loanword adaptations can be categorized as adaptations that exist to correct a phonological ill-formedness in the borrower language. These adaptations can correct, for example, consonant clusters from the donor language that are illegal in the borrower language. Interestingly, these adaptations do not always use the same repair strategies as in native words. This project, though, will look at adaptations that fall into a different camp, namely, adaptations that serve some purpose other than to repair phonological forms. For example, in Korean, English loanwords ending in voiceless stops are often adapted by aspirating the stop and epenthesizing a vowel, e.g. 'bat' → [pæt^{h}ɨ]; however, native Korean words that end in voiceless stops are well formed, as we see in [pat] 'field' (Peperkamp 2005, Kang 2003). Studying this type of adaptation gives the researcher insight into how speakers' native phonology and phonetics color their perception of other languages, how speakers form "best matches" between other languages and their own, and how they translate these intuitions into phonetic and phonological renderings.

A large debate among researchers is whether loanword phonology can be completely explained in terms of native grammars, or whether loanwords need their own strata. While an interesting pursuit, this project does not attempt to address this theoretic question. Instead, it focuses on a point in the study of loanword adaptation that forces the researcher to address the complex nest of issues that arise at the intersection of phonology, phonetics, and perception. A complete theory of loanword adaptation will draw from each of these subfields in search of a unified theory. This project will cover a small piece of ground in that spirit, with a particular interest in the role that perception plays in the adaptation process.[1]

1 Thank you to my advisor, Bruce Hayes; my committee members Robert Daland and Kie Zuraw; my Finnish consultant, Tuomo Tiisala; and Jason Bishop for their help on this project.

0. Introduction

This project will look at loanword adaptation in Finnish, focusing on the gemination of voiceless stops. In the Finnish loanword process, non-geminate donor voiceless stops are often adapted in geminated form. This gemination cannot be accounted for by appealing to any native Finnish phonological ill-formedness constraint. Previous theories have hypothesized that the gemination is explainable using phonological principles. Kiparsky (2003) posits a phonological theory in which the gemination of stops is motivated by faithfulness to the donor word, while Karvonen (1998, 2005) posits a phonological theory of gemination based on word-structure constraints. I will argue that, while there is evidence that word-structure constraints play a role in the gemination adaptation process, gemination must also be influenced by perceptual cues. Specifically, I will show that the data support a theory in which geminates are adapted as such because of two perceptual factors: stress placement in the donor and loan word ("stress effects"), and the existence of subsequent stops in the donor word ("stop effects").

Data used in this project are on-line adaptations by a native speaker of Finnish. On-line adaptations are used instead of existing corpora data in order to control for confounding factors in the corpora data, namely uncertainty regarding the origin language of donor words and the times at which the donor words were adapted into Finnish. Two separate on-line adaptation corpora are used here. The first consists of adaptations of English words and the second of adaptations of "English" nonce words. In addition to controlling for the confounding factors mentioned, using on-line adaptations also allows the researcher to target specific segments and sequences of segments in the adaptation process. Because this project is interested in the gemination of voiceless stops, the donor English words and nonce words chosen are forms that contain occurrences of voiceless stops.

Section 1 will provide an overview of Finnish phonology, including a description of the phonological rules Quantitative Consonant Gradation and Stop Deletion. Section 2 will provide an overview of Kiparsky and Karvonen's theories of gemination in Finnish loanword adaptation and provide data that are problematic for the two theories. Section 3 will present on-line adaptations of English donor words. I will argue based on these data that there is evidence of both structural rules and perceptual effects in play in Finnish loanword gemination. Section 4 will present and discuss wug-test nonce words adaptation data. Section 5 will construct a grammar in the maxent framework to account for the adaptation of the nonce wug forms. The grammar will show that the two effects mentioned, the stress effect and stop effect, are actively involved in gemination during the adaptation of the nonce words. Section 6 will further explore the stop effect. I will show that while the stop effect pattern exists in the loanword data, no strong evidence supports its existence in the native lexicon. I will argue that this result strongly suggests that the effect is a perceptual and not structural effect. Additionally, this section will provide evidence that a main Finnish phonological constraint, Quantitative Consonant Gradation, is in its current formulation an overgeneralization that obscures two distinct constraints violated at different rates in the native Finnish lexicon.

1. Finnish Phonology and Orthography

Finnish is a Finno-Ugric language with about 5 million speakers and is, with Swedish, the official language of Finland. Finnish is a morphologically rich language, with about 15 cases,[2] and written and spoken phonemes always correspond. The present analysis will concentrate on stems, though the reader should be aware that the rules discussed also have extensive interaction with morphology that won't be discussed here.

1.1 Phonological Inventory

Vowels

The native Finnish vowel inventory includes eight monophthongs, given below in their

2 Nominative, partitive, accusative, genitive, inessive, elative, illative, adessive, ablative, allative, essive, abessive, comitative, instructive, and translative (Karlsson, 2008).

orthographic form:

Table 1.1.1 Native Finnish Vowel Inventory

	Front	Back
High	i y	u
Mid	e ö	o
Low	ä	a

Finnish spelling is phonemic. The segment [ö] is pronounced as IPA /ø/ and [ä] is pronounced as IPA /æ/. Finnish spelling will be used in this paper. All monophthongs appear contrastively as long vowels, which are spelled as double vowels. Finnish also has 16 common diphthongs (Karlsson 2008):

Table 1.1.2 Native Finnish Diphthong Inventory

Ending in /i/	Ending in /u/	Ending in /y/	Opening
ei	au	äy	ie
äi	ou	öy	yö
ui	eu		uo
ai	iu		
oi			
öi			
yi			

Finnish has vowel harmony, for which the vowels are split into three classes (Suomi et al. 2008):

Table 1.1.3 Finnish Vowel Harmony Classes

Front Harmonic	Back Harmonic	Neutral
y ö ä	u o a	i e

Within non-compounded words, vowels from the front harmonic and back harmonic groups cannot co-occur. Neutral vowels can co-occur with either group.

Consonants

The Finnish consonant inventory is given in Table 1.1.4 below. Loanword inventory is in bold.

Table 1.1.4 Finnish Consonant Inventory

	Bilabial	Labio-dental	Dental	Palatal	Palato-alveolar	Velar	Glottal
Stops	p **b**		t d[3]			k **g**	
Flaps							
Fricatives		**f**	s		**ʃ**		h
Nasals	m		n			ŋ[4]	
Trills			r				
Laterals			l				
Glides		v		j			

The loanword segments [b], [g], [f], and [ʃ] are not native Finnish phonemes, but are used freely in loanwords by younger speakers. All native segments except for [v], [h], and [j] appear contrastively as geminates.[5] Phonotactically, complex onsets and codas are banned in native Finnish words. However, both are now common in loanwords, e.g. *presidentti*. CCC clusters can appear in native words, but they are restricted to a nasal or liquid followed by obstruents, e.g. *tarkka*. Any CCC cluster must have a syllable break between the second and third consonant. Word-final consonants are legal, but in native words only the set of coronal consonants can appear word-finally.

1.2 Syllabification

Syllabification can, in general, be determined by applying the following rule: place a syllable boundary before every sequence of a single consonant followed by a (long or short) vowel (Karlsson 2008) and between a sequence of two internal consonants (Suomi et al. 2008). Minimal word structure in Finnish is (C)VV or (C)VCV (Hanson and Kiparsky 1996, Karvonen 2005). Notably, a CVC structure is not a minimal word in Finnish. Table 1.2 gives the ten basic syllable structures with their relative proportion (Suomi et al. 2008 and citations within):

3 I will analyze [d] featurally as the flap [ɾ]. See Karvonen (2005) and sources cited within.
4 [ŋ] appears only word medially in the combination *ng* [ŋŋ] or *nk* [ŋk] (Karvonen 2005). The combination *ng* is also the only place that the segment *g* appears in native words.
5 See Suomi et. al. (2008) pg. 41 for a discussion of particular dialects in which the [h] [j] [v] consonants also geminate.

Table 1.2 Ten Basic Syllable Types in Finnish

Syllable Structure	Proportion
CV	40.40%
CVC	27.50%
CVV	12.70%
CVVC	9.60%
VC	3.90%
V	3.90%
VV	1.20%
CVCC	0.60%
VVC	0.30%
VCC	0.10%

1.3 Stress

Primary stress in Finnish always falls on the first syllable of a word, with secondary stress in stems generally[6] falling on odd-number syllables starting from the left, with the requirement that the last syllable of a word is never stressed (Kiparsky 2003). Stress in loanwords adheres to the rules of native stress. This is predicted by Peperkamp's (2004) theory in which lexical exceptions to phonological rules occur in a language only if speakers of that language can perceive the exceptions *as* exceptions. Because primary stress in native words is fully surface observable (exception-less) in Finnish, speakers do not encode primary stress in the phonological representation of words in their mental lexicon. Speakers of Finnish are therefore unable to recognize or store exceptional stress patterns (Peperkamp 2004 and experimental studies on Finnish speakers cited within), and thus cannot store irregular stress patterns in loanwords. Loanwords are therefore adopted with native stress patterns. Peperkamp leaves open the question of secondary stress, which in a language such as Finnish is not surface observable—that is, not exceptionless—and therefore must be learned lexically by speakers, even while these speakers have difficulty perceiving main stress. She suggests the possibility that primary and secondary stress differ in their acoustic cues and are processed differently. While an interesting question, this project will leave discussions of secondary stress aside.

6 For an extensive discussion of stress and exceptions to this generalization, see Kiparsky (2003) and Karvonen (2005).

1.4 Phonological Rules

Finnish phonology is characterized by two main processes affecting voiceless stops. Both of these processes will be relevant in our discussion of voiceless stop gemination. The first process is *Quantitative Consonant Gradation*, in which a voiceless stop geminate degeminates in the onset of a heavy syllable (Kiparsky 2003). Minimal heavy syllable structures in Finnish are (C)VV and (C)VC. The process can be formulated in a rule as follows:

QUANTITATIVE CONSONANT GRADATION
[-cont, -voice, α place] → ∅ / ____ $]_\sigma$ [[-cont, -voice, α place] μμ$]_\sigma$

Quantitative Consonant Gradation is an active alternative, for example:

[ma**t.t**o]	'flower'	+	[-lla]	=	[ma.**t**ol.la]	*[mat.tol.la]	'of the flower'

The second process is *Stop Deletion*, in which an intervocalic [t] is deleted when the first vowel is unstressed and short (Kiparsky 2003). The process can be formulated in a rule as:

STOP DELETION
t → ∅ / [V, -stress, short] ____ V

Stop Deletion is also an active alternation, for example:

[lakko]	'strike'	+	[-ta]	=	[lak.ko.a]	*[lak.ko.ta]	'about a/the strike'

Although Stop Deletion is actively seen operating only on [t], the rule generalizes as a phonotactic constraint for [k] and [p] in native Finnish words (Kiparsky 2003), i.e. you do not find [k] and [p] segments in the environment after a short, unstressed vowel and before a vowel in native words. Both Quantitative Consonant Gradation and Stop Deletion hold in stem as well as derived environments and in loanwords, though not without exceptions.

2. Gemination in Finnish Loanword Phonology

2.1 A History

Finnish is a language particularly rich in loanwords. To understand why, it helps to be familiar with the broad strokes of its history. Historically, Finland was an agriculture-based society with the arguable misfortune of being located between the Kingdom of Sweden and the Russian Empire. Finland was first intregrated into Europe during the 13th century, when populations from the Kingdom of Sweden settled on the land. Soon after, in the 14th century, Sweden incorporated the area as part of its kingdom. Finland stayed a part of Sweden until 1809, when it became an autonomous Grand Duchy within the Russian Empire following the Finnish War. Finland remained part of Russia until it declared independence at the onset of the Russian Revolution in 1917. It has been a sovereign state since.

Finland's language history generally follows a similar arc as its political history. Records of Finnish as a written language start in the 1540s, when the first Finnish language book was published (ABC by Mikael Agricola). The period between 1540-1810 is considered the era of Old Literary Finnish. During this time, most Finnish texts were translations of religious or legal documents from other languages; therefore, many sentence structures and orthographic rules were copied from languages such as Latin and Swedish. Swedish was a dominant language and was spoken by those in the upper classes. 1810 marks the beginning of 19th Century Finnish. A new orthography was established when it was used in a bible translation in the early 1850s. Following this, the vocabulary of the language expanded greatly and the written syntax was revised. By 1880, the era of 19th Century Finnish was over and Modern Finnish had taken root.[7]

The importance of this history on the language should not be overlooked. Although Finnish is not related to the languages with which it has had most contact, the close and frequent contact of these languages over such a long period, and while Finnish was developing as a written language, has resulted in a vast number of loanwords entering the language. Although many older loans are likely

7 Historical information from the Institute for the Languages of Finland (www.kotus.fi).

Swedish in origin, many are now from English. Especially among younger Finns, English words are extremely trendy and seem to be borrowed quite freely.

The exact number of loanwords in Finnish is difficult to pin down precisely. A 2003 study commissioned by the Nordic Language Council[8] found that the average proportion of modern loanwords, defined as loans that entered the language after 1945, in Finnish newspaper texts from 1975 and 2000 was .2%—actually quite low compared to other Nordic languages. For example, loans at a rate of 1% were found in Danish. Running a check on the corpora used in this project (detailed in Appendix A) found that 7% of the ~14,000 types in the corpus could be identified as recent loanwords[9] —a significantly greater percentage. The greater frequency of loans found in my corpus is likely attributable to the fact that my corpora newspapers are all from the 1990s, and would therefore be expected to have a greater number of loans than newspapers from the 1970s.

2.2 Current Theories of Gemination in Finnish Loanword Adaptation

There are two theories of gemination in Finnish loanword adaptation currently in the literature. The first is from Kiparsky (2003), in which voiceless stops geminate to avoid being the target of Stop Deletion. The second is from Karvonen (1998, 2005), in which voiceless stops obligatorily geminate word-finally. Both theories draw from the existing vocabulary of Finnish loanwords. I will briefly summarize each position and then present data that I find to be problematic for each theory. Because data that cannot be explained by either theory were easily found, I will conclude that the theories need to be modified to have greater descriptive power.

Gemination for Stop Deletion

Kiparsky (2003) proposes that in loanword adaptation voiceless stops are geminated to avoid being the target of Stop Deletion. Gemination of this type is seen in (1).[10] In the adaptation process, the

8 Anne-Line Graedler. Modern Loanwords in the Nordic Countries: Presentation of a Project. *Nordic Journal of English Studies*. Special issue. 3(2).
9 Meaning that they exhibit the phonological characteristics of modern loanwords, including foreign segments and clusters. These will be explained in greater detail later in this section.
10 All non-experimental data used from Karvonen (1998), Kiparsky (2003), and Campbell (1998).

voiceless stops [k] and [t] are both underlyingly subject to Stop Deletion as they are in an intervocalic environment in which the first vowel is short and unstressed. If these forms were left unchanged, the stops would delete and the resulting forms would surface as the starred forms listed below. Presumably, these starred forms can be ruled out via a highly ranked faithfulness constraint that regulates the input/output pairs of each segment, such as *MaxDonorC*. To avoid deleting an input consonant, then, the stops geminate and, as geminates are not subject to Stop Deletion, surface as shown.

(1)	'picnic'	→	/pík.ni.**k**i/	→	[pík.ni**k.k**i]	*[pík.nii]
	'mammoth'	→	/mám.mu.**t**i/	→	[mám.mu**t.t**i]	*[mám.mui]

However, this analysis cannot be extended to additional data:

(2)	'prophet'	→	/pró.f**ee**.ta/	→	[pró.feet.ta]	*[pró.fee.ta]
	'satellite'	→	/sá.tel.l**ii**.ti/	→	[sá.tel.liit.ti]	*[sá.tel.lii.ti]
	'idiot'	→	/í.di.**oo**.ti/	→	[í.di.oot.ti]	*[í.di.oo.ti]

(3)	'tape'	→	/t**éi**.pi/	→	[téip.pi]	*[téi.pi]
	'rock'	→	/r**ó**.ki/	→	[rók.ki]	*[ró.ki]

The data in (2) and (3) are examples in which voiceless stops have been adapted as geminates; however, the stops in these examples are not underlyingly in the environment for Stop Deletion. In the examples in (2), the environment for Stop Deletion is not met because the vowel preceding the geminate voiceless stop is long (vowel in bold). The examples in (3) do not meet the environment for Stop Deletion because the vowel preceding the geminate voiceless stop is stressed (vowel in bold).[11] If the criterion for gemination is to avoid the deletion of a voiceless stop through Stop Deletion, then none of the forms in (2) and (3) should have undergone gemination. These examples clearly demonstrate that

11 I am assuming that forms are subjected to Stop Deletion once stress has already been assigned, since stress is one of the criteria in Stop Deletion's environment.

the hypothesis that voiceless stops geminate to avoid being deleted through the process of Stop Deletion is not explanatorily sufficient.[12]

Obligatory Gemination

Karvonen (1998, 2005) predicts that stem-final voiceless stops will geminate in Finnish loanwords except when the donor word ends in a vowel. He provides the following data as support for this hypothesis:

Table 2.2 Karvonen (1998, 2005) Data

Geminating Forms		Non-Geminating Forms	
English Donor Word	Finnish Loanword	English Donor Word	Finnish Loanword
satellite	satelliitti	city	siti
tape	teippi	khaki	kaki
jeep	jeeppi	happy	häpi
rock	rokki		

Again, additional data are problematic for the analysis. For example, the data given in (4) do not conform to the theory, as each adapted word contains voiceless stop gemination that does not target a stem-final stop:

(4) 'senator' → [se.naat.to.ri] *[se.naa.to.ri]

'capital' → [ka.pit.te.li] *[ka.pi.te.li]

Karvonen's theory does not predict gemination occurring in non-stem-final positions. These examples cast doubt on the completeness of the analysis.

I also have two concerns about the data that Karvonen provides in support of his theory. The first is that native speakers in the Finnish (Helsinki) language community[13] judge the forms [siti],

12 One notable characteristic of these examples is that vowels are epenthesized to the end of the words. This is not done to avoid a bad word-final coda, as even words with coronal final-consonants get epenthesis. I will adopt Kiparsky's (2003) STEM CONSTRAINT to explain this behavior. The constraint requires that stems end in a vowel at the point of affixation; loanwords adapt to this constraint by epenthesis (146-147). Kiparsky argues that i-epenthesis in loanwords reflects an "allomorphy between an uninflectable -C stem which preserves the donor language's word form, and an inflectable -V stem which is adapted so as to conform to the Stem Constraint" (148). That is, i-epenthesis is a way for a noun to prepare itself to receive an overt case ending. For example, in Finnish, the loanword 'nylon' [nailon] is allomorphic with [nailoni], the latter of which is an older form of the word; but, only the latter form can appear when inflected, e.g. as in the generative singular for [nailonin] (Kiparsky 2003).

13 Communication with my native speaker consultant.

[kaki], and [häpi] as being not in standard spoken/written usage. Non-standard forms are not optimal data to base an analysis on. The second is that the immediate donor language(s) of the geminating forms is not known for sure (Karvonen does not cite the source of the data). For example, the loanwords [idiotti] and [satelliitti] are credited as coming from the English ['ɪdiət] and ['satəlaɪt], though an equally likely direct source of the loanwords is from the Swedish loanword adaptations [idi'oːt] and [sɑte'liːt].[14] Because the stress and pronunciation of the words are different in English and in Swedish, it is crucial to know which language the word was directly borrowed from when forming an explanatory theory. In light of the discussion in this section, I conclude that neither theory satisfactorily explains all the available data and that, therefore, a modified theory is needed.

3. On-line Adaptations of English Loanwords

In Section 2, I concluded that neither Kiparsky nor Karvonen's theory of loanword adaptation is sufficient to explain the existing Finnish loanword data. In this section, I will test their theories against a corpus of on-line adaptations and demonstrate that, again, the two theories fail to provide a comprehensive explanation for gemination adaptation. There are two reasons why I believe that using experimental data in this case is preferable to using existing data. The first is the concern mentioned in the previous section, which is that of gathering clean data from a determined source. The phonological adaptation of Finnish loanwords has changed over time, resulting in classes of loanwords that have been adapted differently into the language. For example, historically, loanwords were adapted to fit native phonological patterns and so are almost indistinguishable from native words, e.g. *ranta* from Swedish *strand*. However, newer adaptations are easily identifiable because they retain non-native segments and clusters, e.g. *presidentti* from English *president* (Suomi et al. 2008). These discrepancies mean that different loanwords have been adapted into Finnish using different principles and thus

14 Similarly, the loanword [prof'eetta] in (2) is credited as coming from the English ['prɑfet] (Kiparsky), though another likely direct source is the Swedish [pro'feːt].

occupy different “strata”; clearly, different strata should not be analyzed under a single grammar. Relatedly, it is hard to pin down the direct borrowing language of loanwords. As mentioned in the last section, various words have possibly been borrowed from English into Swedish and then into Finnish.

The second reason that I believe that using experimental data is preferable to using existing data is that the experimental data can be controlled to allow targeted study of particular phenomena. For example, this study focuses on geminates; experimental data allow me to create data points to test these specific segments. In this case, given the subject of interest is geminate stops, words with sequences of voiceless stops were chosen.

3.1 Corpus Data

The corpus of on-line adaptations consists of 126 forms adapted from 118 English donor words (some forms were adapted in two ways). The English words were chosen with the goal of maximizing the presence and interaction of voiceless stops. A few “dummy” words were also included in the data in order to mask the purpose of the experiment from my consultant. The donor words were randomized into nine groups. A slide show, with each word placed on a separate blank screen, was created for each group. Three groups at a time were shown to the consultant on three different days. The consultant was instructed to view each slide and to listen carefully to my simultaneous pronunciation of the word, then to adapt the word into Finnish using the orthography and pronunciation he felt appropriate. Out of the 126 adaptation forms, 38 forms had segments adapted as geminates (all voiceless stops). Adaptations that underwent gemination are listed in Table 3.1.

Table 3.1 On-Line Adaptation Forms with Gemination

English Word	Finnish Adaptation
absurdity	absurditeetti
batik	batiikki
celibate	selibaatti
citation	saitaatti
combatant	kombatantti
commentator	kommentaattori
committal	komittaali
democratic	demokraattinen
detente	detaantti
exorbitant	eksorbitantti
gecko	gekko
grandiloquent	grandilokuentti
hematite	hematiitti
hepatitis	hepatiitti
hermaphroditic	hermafrodiittinen
hierarchical	hierarkkinen
hippopotamus	hippopotamus
hypocritical	hippokritikaalinen
infertility	infertiliteetti
institute	instituutti
levant	levantti
mastoiditis	mastoidiitti
narcotic	narkoottinen
omnipotent	omnipotentti
ottoman	ottomani
peripatetic	peripateettinen
perpetrator	perpetraattori
petticoat	pedikoutti
petticoat	petikoutti
pocket	poketti
potent	potentti
potentate	potentaatti
predicate	predikaatti
rebuttal	rebuttali
socratic	sokraattinen
streptococcus	streptokokki
title	titteli
tittle	titteli

3.2 Kiparsky's Stop Deletion Theory

Out of the 38 gemination forms, only one form can be explained by Kiparsky's theory:

(5) 'pocket' [pɑkət] → /po.ke.ti/ → [po.ket.ti]

Here, the [t] is in the environment for Stop Deletion because it is intervocalic and the first vowel is short and unstressed. However, predicting one gemination out of 38 does not give the theory much predictive power, and I again conclude that the theory does not sufficiently explain why segments are adapted as geminates.

3.3 Karvonen's Word-Final Theory

Out of the 38 geminates, 19 can be analyzed as having word-final gemination. For example,

(6) 'batik' [bəˈtiːk] → [ba.tiik.ki]

'potent' [ˈpoʊtnt] → [po.tent.ti]

These 19 forms include data that are exceptions to Karvonen's claim that vowel-final forms do not geminate. For example,

(7) 'absurdity' [əbˈsəɹdəti] → [ab.sur.di.teet.ti]

In example (7), a vowel-final donor word undergoes gemination of the preceding voiceless stop, counter to Karvonen's claim. These data could maybe be accounted for by modifying the theory.

However, the theory still does not account for the other 19 geminating forms. Some of these forms are those that have word-final gemination in the adaptation form but in which the segment geminated was not word-final in the donor form. For example,

(8) 'streptococcus' [ˌstɹɛptəˈkɑkəs] → [strep.to.kok.ki]

In example (8), the [k] is adapted as a word-final geminate in the Finnish loanword, but is word-medial in the donor form. Other forms undergo gemination that is neither word-final in the donor form nor in the loanword form. For example,

(9) 'perpetrator' [ˈpəɹpəˌtɹeɪɾəɹ] → [per.pet.raat.to.ri]

In example (9), the geminated [t] is neither word-final in the donor word not does it end up in a word-final position in the loanword. In conclusion, Karvonen's account has fairly good predictive power in the corpus data. Forms such as (7) and (8) could possibly be accounted for under a reformulation of the theory; however, forms such as (9) appear to pose a greater challenge. Because of these exceptions to the theory, I again conclude that it does not have the predictive power to account for all the observed data.

3.4 Stress

When looking at the corpus data, another pattern emerges. Of the 38 geminations, 17 followed the primary-stressed donor syllable and four followed a secondary-stressed donor syllable. For example:

(10) 'title' [ˈtaɪtəl] → [tit.te.li] 'hierarchical' [ˌhɑɪəɹˈɑɹkɪkəl] → [hie.rark.ki.nen]

Five of these 21 forms overlap with the group in section 3.3, i.e. the gemination in these five forms follows stress and is also word final. For example:

(11) 'gecko' [ˈgɛkoʊ] → [gek.ko]

This observation suggests that stress is a predictive factor in voiceless stops being adapted as geminates.[15] Together, Karvonen's word-final theory and the effect of stress explain all the gemination observed in the corpus. Section 4 will explore this stress effect further.

3.5 Morphology and Orthography

Two other effects tangentially related to gemination are observable in the data. Neither of these effects will play a role in the rest of the paper; however, they are worth a brief discussion as they are of interest to loanword adaptation more generally.

15 Constraints on this process include the formation of an illegal consonant cluster, e.g. [pkk], pronunciation as a voiceless stop phoneme, e.g. the [t] in [-tion] is not a candidate for gemination because it is pronounced as [ʃ] in the donor word, and an apparent dispreference for gemination that also includes an orthographic change, e.g. [c] → [k] → [kk].

Morphology

The data support the hypothesis that morphology plays a role in how English words are adapted into Finnish. It is reasonable to hypothesize that naïve adapters, speakers who have had no or very limited exposure to a donor langauge, will rely entirely upon phonetics and native phonology when adapting loanwords. That is, a speaker hears a stream of sounds, filters them through their own phonetic and phonological system, and produces a form that is a mix of faithfulness to the original word form and faithfulness to their native phonetic/phonological system. Non-naïve adapters, on the other hand, will have access to categorical information about the donor language and will utilize this knowledge in the adaptation process. Evidence for this second principle is seen in the adaptations of my consultant, who is also a fluent English speaker. While adapting English words, he adapted particular English forms categorically into their Finnish equivalent form. For example, -*ic* endings were regularly adapted as -*inen*, an adjectival ending in Finnish, e.g. 'socratic' → [sokraattinen]. This adaptation is clearly not based on phonetic input or native phonological rules, but is a categorical translation from an English morpheme into the equivalent Finnish one.

Orthography

The data also support the hypothesis that orthography plays a role in adaptation, including the presence of geminates in adaptation forms. For example, 'hippopotamus' → [hippopotamus] and 'ottoman' → [ottomani] are examples of gemination following donor stress, but they are also cases where the geminate exists orthographically in the donor word. A clearer case is the following: 'commentator' → [kommentaattori]. This example contains the only non-voiceless stop geminate in an adapted form; most likely it is a faithful rendering of the orthography of the donor word.

The rest of this project will not discuss either of these two effects. However, it is worth keeping in mind that a complete grammar of loanword adaptation would need to capture both the phonetic/phonological and categorical features of adaptation.

4. Wug Testing

In the previous section, I concluded that Karvonen's structural account—word-final gemination—together with the effect of donor stress on gemination accounted for all the gemination observed in the on-line adaptation English corpus. This section will further explore the stress effect. Because stress as we are interested in it is not a structural constraint, meaning it is independent of word meaning, structure, or well-formedness, this section will abstract away from actual words and will instead use on-line adaptations of “English” nonce words. Using these simple, experimental nonce forms allows us to observe the effects of stress while minimizing the interference of other features, such as word meaning and morphology. The forms used include sequences of voiceless stops, as target data, and sequences of voiced stops, as control data.

4.1 Experimental Data

The data set consists of nonsense words created specifically to test the influence of stress on adapted segments. The nonce words were formed from nine initial tokens: *apapapa*, *abababa*, *apatata*, *akatapa*, *atataka*, *apapa*, *ababa*, *atapa*, and *akata*. Each of these nine tokens were then used to create a pattern that included the original token and the words formed by substitution of each voiceless stop (respectively) in the original token with the segment [b]. Because [b] does not geminate, the inclusion of these data allows for comparison cases between forms that have multiple voiceless stop segments and those that do not.[16] Forms representing each possible combination of primary and secondary stress were included, excepting those with adjacent stress.

After elicitation of the adapted forms, no consistent variation was found to exist between adaptations of voiceless stop segments. All voiceless stop segments appeared to be treated the same, meaning that there was no evidence that forms with the bilabial voiceless stop [p] were adapted any differently than forms with the velar voiceless stop [k]. Therefore, the voiceless stop patterns were

16 In my maxent analysis, I adopt the simplifying assumption that gen does not produce geminate [b]s. I adopted this simplification because the phoneme /bb/ does not occur to my knowledge even in loanwords, and the sequence was never produced by my consultant. This seems to constitute evidence against Suomi et al.'s (2008) claim that dialects that have /b/ should also have /bb/.

condensed and treated as variation on a single input form. For example, input/output pairs such as 'akata' → [aakkata] and 'apapa' → [aapapa] are treated as **one** input, 'apapa', with **two** surface forms, [aappapa] and [aapapa]. The segment [p] is used throughout the data to represent a generic voiceless stop segment. Condensing the data in this way led to a total of 54 inputs and 199 surface forms. Because the data has been condensed, each four syllable input has four output forms, each of which originally represented a form with a different voiceless stop pattern. Condensed input forms are listed in Tables 4.1.1 and 4.1.2 (again, [p] represents a generic voiceless stop).

Table 4.1.1 Four syllable wug donor forms

Stop Sequence → Syllable Stress ↓	/p/ /p/ /p/	/b/ /b/ /p/	/b/ /p/ /b/	/b/ /p/ /p/	/p/ /b/ /b/	/p/ /b/ /p/	/p/ /p/ /b/
Primary: 2nd Secondary: 4th	a.ˈpa.pa.ˌpa	a.ˈba.ba.ˌpa	a.ˈba.pa.ˌba	a.ˈba.pa.ˌpa	a.ˈpa.ba.ˌba	a.ˈpa.ba.ˌpa	a.ˈpa.pa.ˌba
Primary: 4th Secondary: 2nd	a.ˌpa.pa.ˈpa	a.ˌba.ba.ˈpa	a.ˌba.pa.ˈba	a.ˌba.pa.ˈpa	a.ˌpa.ba.ˈba	a.ˌpa.ba.ˈpa	a.ˌpa.pa.ˈba
Primary: 1st Secondary: 3rd	ˈa.pa.ˌpa.pa	ˈa.ba.ba.ˌpa	ˈa.ba.pa.ˌba	ˈa.ba.pa.ˌpa	ˈa.pa.ba.ˌba	ˈa.pa.ba.ˌpa	ˈa.pa.pa.ˌba
Primary: 4th Secondary: 1st	ˌa.pa.pa.ˈpa	ˈa.ba.ˌba.pa	ˈa.ba.ˌpa.ba	ˈa.ba.ˌpa.pa	ˈa.pa.ˌba.ba	ˈa.pa.ˌba.pa	ˈa.pa.ˌpa.ba
Primary: 3rd Secondary:1st	ˌa.pa.ˈpa.pa	ˌa.ba.ba.ˈpa	ˌa.ba.pa.ˈba	ˌa.ba.pa.ˈpa	ˌa.pa.ba.ˈba	ˌa.pa.ba.ˈpa	ˌa.pa.pa.ˈba
Primary: 1st Secondary: 4th	ˈa.pa.pa.ˌpa	ˌa.ba.ˈba.pa	ˌa.ba.ˈpa.ba	ˌa.ba.ˈpa.pa	ˌa.pa.ˈba.ba	ˌa.pa.ˈba.pa	ˌa.pa.pˈa.ba

Table 4.1.2 Three syllable wug donor forms

Stop Sequence → Syllable Stress ↓	/b/ /b/	/b/ /p/	/p/ /b/	/p/ /p/
Primary: 2nd	a.ˈba.ba	a.ˈba.pa	a.ˈpa.ba	a.ˈpa.pa
Primary: 1st Secondary: 3rd	ˈa.ba.ˌba	ˈa.ba.ˌpa	ˈa.pa.ˌba	ˈa.pa.ˌpa
Primary: 3rd Secondary: 1st	ˌa.ba.ˈba	ˌa.ba.ˈpa	ˌa.pa.ˈba	ˌa.pa.ˈpa

4.2 Elicitations

The consultant tested is a native Finnish speaker, Helsinki dialect, who is also fluent in English.[17] Adaptations were elicited over three sessions. In each session, a section of the “English” nonce words were presented to the consultant in spoken form (no written form was provided). The

17 The data here is from a single speaker and as so is being used only to gather preliminary hypotheses; all claims here are therefore made under this caveat and will obviously need to be tested against a larger data set.

consultant was asked to listen to each word and to adapt it in writing into a Finnish loanword. Each session was recorded. A table of the input and output nonce forms is given in Appendix B.

4.3 Data Results

Out of 199 adapted forms, 31 forms underwent gemination. The wug-test was designed to test the hypothesis that there is a perceptual effect in which the stress of the donor word is relevant to gemination during the adaptation process. As this section will show, the wug-test data support this hypothesis. As well as supporting the observation that donor stress effects gemination in the adapted forms, the data also show that donor stress effects vowel lengthening in the adapted forms. Therefore, I will include both vowel and consonant lengthening as part of the stress effect. In addition to supporting the stress effect hypothesis, the wug-test data also suggest another perceptual effect at work during the adaptation process. This effect I will preliminarily call the “stop effect.” The stop effect is a preference in the data for a voiceless stop to geminate if it is found before a voiced stop as opposed to if it is found before another voiceless stop. The following two sub-sections will discuss both the stress and stop effects further.

Stress and Gemination

As expected vis-à-vis the discussion in section 1.3, primary stress in all the nonce loanwords was adapted as word initial, as in native Finnish, regardless of donor stress position. As in the English corpus, the effect of donor stress on gemination is apparent in the data. Post-tonic voiceless stops were the only voiceless stops to be adapted as geminates; voiceless stops not immediately following stress were never adapted as geminates. This is very strong evidence in favor of a perceptual effect of stress on the adaptation of voiceless stops as geminates. In addition to this evidence, there is also evidence for an effect not observed in the English corpus. The data suggest that, in addition to the effect of donor stress, the stress of the adapted form also plays a role in gemination. Ninety percent of the voiceless stops adapted as geminates are not only post-tonic, but also the first consonant of the vowel-initial word. As already discussed, Finnish primary stress is always word initial; therefore, 90% of the geminate forms are those in which English primary or secondary stress coincides with Finnish primary stress. The three exceptions to this pattern include two forms in which the second consonant was geminated (both voiceless stops were still post-tonic). The third exception is a form that is a strong outlier to all the other data forms. I suspect that the form was misheard by the consultant. Because I believe that the form was misheard by the consultant, I will not discuss it as a relevant outlier to the observations here; the form is however included in all data charts and analysis.

Donor Stress and Vowel Length

The data also suggest another perceptual effect of stress. Specifically, that stressed donor vowels are perceived as long vowels and so are rendered as such in the adaptation form. One hundred percent of primary stressed donor vowels were rendered as long vowels, and 86% of secondary stressed donor vowels were rendered as long vowels.[18] The most plausible explanation of this result is that Finnish speakers, who use long and short vowels contrastively, interpret the stressed English vowel, which is longer than an English unstressed vowel (Lehiste 1970), as corresponding to a Finnish long vowel. This also explains the discrepancy between primary and secondary stressed vowels. Primary stressed vowels are longer than secondary stressed vowels, and are therefore more likely to be perceived by a Finn as corresponding to what they interpret as a long vowel.

Stop Effect

The wug test data suggest another perceptual effect relevant to gemination, other than that related to stress. This effect is caused by the interaction of voiceless and voiced stops. Specifically, a voiceless stop was more likely to geminate, holding other factors such as stress constant, if it was followed by a voiced stop instead of another voiceless stop, i.e. there is a preference for [p.pa.b] forms over [p.pa.p] forms. In the data set, if a voiceless stop is followed by a [b], the voiceless stop geminates 58% of the time. However, if a voiceless stop is followed by another voiceless stop, then it geminates

18 The only non-stressed vowels adapted as long vowels are word final vowels in the donor form. See Section 6 for discussion.

only 16% of the time. This effect will be discussed extensively in Section 6, in which I will provide evidence that the effect is indeed a perceptual one, and not phonotactic in origin.

Summary

In this section, I have argued that the wug-test data support the tentative hypothesis from section 3 that donor stress can cause gemination in the adaptation process. In addition, I showed that the data in this section support three additional hypotheses. The first is that, in addition to donor stress causing gemination in adaptation forms, it also causes vowel lengthening in adaptation forms. The second is that *Finnish* stress also affects gemination, with gemination prefering to align not only with donor stress but with Finnish stress as well. The last hypothesis supported by the data in this section is the stop effect. Namely, there is a preference for gemination to occur before a [b] segment instead of before a [p] segment. In conclusion, it appears that, in the absence of any structural or meaning cues, the only two factors that affect the rendering of voiceless stops as geminates are stress and exposure to subsequent voiced/voiceless stops.

5. Grammar

In section 4, I presented a corpus of on-line elicitations of nonce English words into Finnish loanwords. I also discussed the main patterns found in the data, including the presence of the stress effect and the stop effect. This section will construct a grammar based on these observations. The framework chosen is the maximum entropy OT model (Smolensky 1986, Smolensky et al. 2002, Goldwater et al. 2003, Wilson 2006) and is implemented using the maxent grammar tool. The maxent grammar model is ideal for adaptation because it accounts for variations in surface form by assigning weights instead of inviolable rankings to constraints.

5.1 Input

A GEN (candidate set) was created containing all possible combinations of vowel lengthening and consonant gemination for each input form. For example, the input [ˈa.pa.ˌpa.pa] has 128 candidates, each representing a possible vowel lengthening and gemination combination of the form. Setting up the grammar in this way resulted in 2,712 output candidates.

5.2 Constraints

<u>Stress Effect</u>

As we saw in section 4.3, donor stress appears to effect gemination and vowel lengthening in the adaptation forms. To model this effect in the grammar, a basic stress to weight principle is utilized. The principle holds that the stress value of a syllable in a donor form corresponds to the weight of the syllable in the adapted form. As observed, primary stress and secondary stress did not act identically in the data, for while 100% of primary stressed donor vowels were rendered as long vowels in the adapted forms, only 86% of secondary stressed donor vowels were rendered as long vowels in the adapted forms (no difference was found between primary and secondary stress for gemination). Therefore, the stress to weight constraint used must be sensitive to these differences. To ensure this, the basic stress to weight principle was broken down into four separate constraints. Each of these four constraints targets, respectively, syllables that carry no stress, primary stress, secondary stress, or unspecified (primary *or* secondary) stress. For example, *If Donor Primary Stressed, Loan Heavy* is violated in the input/outpair 'ˈa.pa.pa' → *[ˈa.pa.pa].

Each of these four constraints was then broken down into two constraints each, to account for weight-to-stress effects and stress-to-weight effects, respectively. For example, *If Donor Primary Stressed, Loan Heavy* breaks down into the following two constraints:

If Donor Primary Stressed, Loan Heavy: 'ˈa.pa.pa' → [aa.paa.pa]

If Loan Heavy, Donor Primary Stressed: 'ˈa.pa.pa' → *[aa.paa.pa]

Dividing the logical space in this manner leads to a total of eight constraints. Finally, each of these eight constraints was again broken down to target specifically vowel or consonant lengthening. For example, *If Donor Primary Stressed, Loan Heavy* corresponds to the following two constraints and adaptations:

If Donor Primary Stressed, Long Vowel:	'a.ˈpa.pa'	→	[a.paa.pa]
If Donor Primary Stressed, Long Consonant:	'a.ˈpa.pa'	→	[a.pap.pa]

The resulting set of 16 constraints symmetrically accounts for all the possibilities of weight-to-stress and stress-to-weight translations. All constraints are listed in Table 5.2.1; each constraint is listed by supra-categories with constraint name (in bold) and constraint formulation (on right). Forming the constraint set in this way accounts for all possibilities within the logical space; the grammar will then decide systematically which constraints are relevant in the data set.

Table 5.2.1 Sixteen WtS and StW Candidate Constraints

No Stress to Weight/Weight to No Stress		
IF DONOR UNSTRESSED, LOAN LIGHT.	**If donor unstressed, loan short V**: If the donor syllable is unstressed, then the loan syllable has a short vowel.	IF(English unstressed ∧ ¬ Finnish V), 1, 0[19]
	If donor unstressed, loan singleton C: If the donor syllable is unstressed, then the loan syllable has singleton consonant.	IF(English unstressed ∧ ¬ Finnish C), 1, 0
IF LOAN LIGHT, DONOR UNSTRESSED.	**If loan short V, donor unstressed**: If the loan syllable has a short vowel, then the donor syllable is unstressed.	IF(Finnish V ∧ ¬ English unstressed), 1, 0
	If loan singleton C, donor unstressed: If the loan syllable has a singleton consonant, then the donor syllable is unstressed.	IF(Finnish C ∧ ¬ English unstressed), 1, 0

Primary Stress to Weight/Weight to Primary Stress		
IF DONOR PRIMARY STRESS, LOAN HEAVY.	**If donor primary stress, loan VV**: If the donor syllable carries primary stress, then the loan syllable has a long vowel.	IF(English primary stress ∧ ¬ Finnish VV), 1, 0
	If donor primary stress, loan CC: If the donor syllable carries primary stress, then the loan syllable has a geminate C.	IF(English primary stress ∧ ¬ Finnish CC), 1, 0
IF LOAN HEAVY, DONOR PRIMARY STRESS.	**If loan VV, donor primary stress**: If the loan syllable has a long vowel, then the donor syllable carries primary stress.	IF(Finnish VV ∧ ¬ English primary stress), 1, 0
	If loan CC, donor primary stress: If the loan syllable has a geminate consonant, then the donor syllable carries primary stress.	IF(Finnish CC ∧ ¬ English primary stress), 1, 0

19 Should be read as "If an English syllable carries no stress and is not represented as a short vowel in the Finnish adapted form, assign one violation. If one or both of these conditions fail to hold, then assign no violations."

Secondary Stress to Weight/Weight to Secondary Stress		
IF DONOR SECONDARY STRESS, LOAN HEAVY.	**If donor secondary stress, loan VV**: If the donor syllable carries secondary stress, then the loan syllable has a long vowel.	IF(English secondary stress ∧ ¬ Finnish VV), 1, 0
	If donor secondary stress, loan CC: If the donor syllable carries secondary stress, then the loan syllable has a geminate C.	IF(English secondary stress ∧ ¬ Finnish CC), 1, 0
IF LOAN HEAVY, DONOR SECONDARY STRESS.	**If loan VV, donor secondary stress**: If the loan syllable has a long vowel, then the donor syllable carries secondary stress.	IF(Finnish VV ∧ ¬ English secondary stress), 1, 0
	If loan CC, donor secondary stress: If the loan syllable has a geminate consonant, then the donor syllable carries secondary stress.	IF(Finnish CC ∧ ¬ English secondary stress), 1, 0

Stress to Weight/Weight to Stress		
IF DONOR STRESS, LOAN HEAVY.	**If donor stress, loan VV**: If the donor syllable carries primary or secondary stress, then the loan syllable has a long vowel.	IF(English stress ∧ ¬ Finnish VV), 1, 0
	If donor stress, loan CC: If the donor syllable carries primary or secondary stress, then the loan syllable has a geminate consonant.	IF(English stress ∧ ¬ Finnish CC), 1, 0
IF LOAN HEAVY, DONOR STRESS.	**If loan VV, donor stress**: If the loan syllable has a long vowel, then the donor syllable carries primary or secondary stress.	IF(Finnish VV ∧ ¬ English stress), 1, 0
	If loan CC, donor stress: If the loan syllable has a geminate consonant, then the donor syllable carries primary or secondary stress.	IF(Finnish CC ∧ ¬ English stress), 1, 0

Finnish Stress to Finnish Weight

As we also saw in section 4.3, Finnish primary stress, as well as English stress, is a factor in determining gemination. Two constraints are included in the grammar to account for this, *Finnish Stress to Finnish Vowel Length* and *Finnish Stress to Finnish Gemination*. These constraints cover the two possible ways to make the word-initial syllable heavy, i.e. by forming a long vowel or by forming a geminate that straddles the first and second syllables. Although Finnish stress was not noticed to effect vowel lengthening in the adapted forms, Finnish Stress to Finnish Vowel Length is included in the grammar in the same spirit as the previous section: to account for all possibilities and then allow data-fitting to identify the unneeded constraints. Finnish Stress to Finnish Vowel Length is violated in all surface forms in which the first vowel is short, e.g. [ap.pa.pa]. Finnish Stress to Finnish Gemination is violated in all surface forms in which the first consonant is not geminated, e.g. [aa.pa.pa]. Both constraints are violated in forms with a light initial syllable, i.e. [a.pa.pa]. These constraints will work in concert with the constraints in the previous section to account for the observation that 90% of gemination in the data occurred when English stress and Finnish primary stress coincided.

Partial Consonant Gradation

In section 4.3, we observed that two variables effect gemination in the adaptation process: donor stress and Finnish stress and the stop effect. The last two subsections account for the effects of donor stress and Finnish stress. This section proposes a constraint to account for the stop effect. Again, the stop effect is a preference for gemination to occur before a voiced stop as opposed to before another voiceless stop. This tendency is expressed in the constraint *Partial Consonant Gradation*. Partial Consonant Gradation punishes a geminate voiceless stop when the following consonant is also a voiceless stop. It is violated in all surface forms with the sequence [(a)ap.pa.pa(a)]. The constraint does not include voiceless stops separated by long vowels, as these forms are already punished by Quantitative Consonant Gradation. The name of the constraint is meant to evoke Quantitative Consonant Gradation because the environments for the constraints are so similar.

Additional Constraints

The constraint **HH* is added to the grammar to account for the dispreference for adjacent heavy syllables, a dispreference in Finnish that is one motivating factor for constraints like Quantitative Consonant Gradation (Anttila 1997). This constraint is violated in candidates like [a.paa.paa.pa] and [ap.paa(p).pa.pa]. The constraint **V#*—don't have a short word-final vowel—is added to account for the preference in the data for word-final vowels to be adapted as long. This tendency is likely due to translation of phonetic final-vowel lengthening in the spoken English nonce words, as English word-final vowels are generally longer than word-medial vowels (Wightman et al. 1992). It is violated in all candidates with a short final vowel, e.g. [a.pa.pa.pa]. The constraint **Super Heavy Syllable* punishes candidates that contain a super heavy syllable, in this case created by a long vowel followed by a geminate (Riad 1992).[20] The constraint is violated, for example, in the forms [**aap**.pa.pa.pa] and

20 See also Kiparsky 2008 for a discussion of Fenno-Swedish dispreference for super heavy syllables consisting of long vowels and geminates.

[a.**paap**.pa.pa]. Both bolded syllables are super heavy in these examples because they contain three moras: two from the long vowel and one from the first segment of the geminate.

Summary

So far in this section, I have taken our observations from the data in section 4 and translated them into constraints. I have included weight-to-stress and stress-to-weight constraints to account for the observations about donor stress and consonant and vowel lengthening, and the constraints *Finnish Stress to Finnish Vowel Length* and *Finnish Stress to Finnish Gemination* to account for the effect of Finnish primary stress. Lastly, the constraint *Partial Consonant Gradation* is included in the grammar in order to account for the observed stop effect. The grammar here is an initial grammar covering all logical possibilities for constraints given the data. The next section will systematically trim the grammar to those constraints that receive statistically significant weights. The preliminary grammar with 25 constraints is given below in Table 5.2.2.

Table 5.2.2 Preliminary Maxent Grammar

Constraint	**Weight**
Ident(V Length)	0
Donor Stress to Donor Vowel Length	0
If loan VV, donor secondary stress	0
If donor secondary stress, loan VV	0
*HH	0.14
If loan singleton C, donor unstressed	0.17
If donor stress, loan CC	0.17
If donor secondary stress, loan CC	0.43
If loan CC, donor secondary stress	0.45
If donor unstressed, loan short V	0.53
If loan VV, donor stress	0.53
If donor unstressed, loan singleton C	0.63
If loan CC, donor stress	0.63
If donor primary stress, loan CC	0.78
If loan CC, donor primary stress	0.79
Ident(C length)	1.11
Partial Consonant Gradation	1.81
*Super Heavy Syllable	3.11

*V#	3.11
If loan short V, donor unstressed	3.71
If donor stress, loan VV	3.71
Donor Stress to Donor Geminate	3.83
If loan VV, donor primary stress	4.45
If donor primary stress, loan VV	7.63
Quantitative Consonant Gradation	12.08

5.3 Finalizing the Grammar

Testing Significance of Constraints

In section 5.2, a grammar with 25 constraints was created to account for the nonce adaptation data. These constraints account for all logical possibilities of the grammar. Weights were fitting using the default settings of the maxent grammar tool. To pare the grammar down to the significant constraints, a likelihood ratio test was performed on each constraint. A likelihood ratio test is used to compare the fit of two nested models, in this case the maxent grammar that includes the constraint being tested and the maxent grammar that does not include the constraint being tested. The likelihood ration test is used to determine whether the fit of the grammar with the greater number of constraints is significantly better than the fit of the grammar with fewer constraints. If so, then the grammar with the greater number of constraints is the preferred grammar to use. I used here the standard statistical significance cut-off point of $p<.05$. To test the constraints, I began with the lowest weighted constraint and tested its significance. If it was significant, the constraint was left in the grammar. If it was not significant, the constraint was removed. The next constraint of lowest weight was then tested on the newly weighted grammar. This method continued until all constraints had been tested. All chi-squared distribution values can be found in Appendix C.

Final Grammar

The final grammar with constraints and weights is given in Table 5.3. The plog (log probability) of the grammar with all 26 constraints was -123.91, while the plog of the final grammar is -125.53. The minor worsening of the plog, -1.62, demonstrates that the removed constraints contributed little to the grammar as a whole.

Table 5.3 Wug-Test Maxent Grammar

Constraint	Weight
Partial Consonant Gradation	1.89
Finnish Stress to Finnish Geminate	3.19
*Super Heavy Syllable	3.21
*V#	3.24
If Loan CC, Donor Stress	3.30
If Loan VV, Donor Primary Stress	5.71
If Donor Stress, Loan VV	8.86
QuantitativeConsonantGradation	12.00

QUANTITATIVE CONSONANT GRADATION

As expected, *Quantitative Consonant Gradation* reaches significance. This result confirms that the constraint has a cognitive reality for Finnish speakers. Indeed, the constraint receives the high weight of 12 because it is unviolated in the data.

PARTIAL CONSONANT GRADATION

The constraint Partial Consonant Gradation also reaches significance, receiving a weight of 1.8. As discussed in section 5.2, this constraint was created to account for the observation that voiceless stops preferred to geminate before a [b] rather than before another voiceless stop, the "stop effect." The survival of this constraint confirms that this observation reflects a reality in the data.

STRESS EFFECT CONSTRAINTS

Of the 16 candidate weight-to-stress/stress-to-weight constraints, three survive significance testing. The *If Loan CC, Donor Stress* constraint indicates the preference for geminates in output forms to be represented by primary or secondary stressed syllables in the input form. As we saw in section 4, there are no geminates in the input forms; therefore, all geminates in surface forms were rendered as such in the adaptation process from singletons. It makes intuitive sense for this constraint to reach significance considering the data; all geminates in the output forms are adapted from post-tonic consonants in the input, but not all input post-tonic consonants are geminated in the adaptation process. The process of input post-tonic consonants being represented as geminates in the output was represented by the constraint *If Donor Stress, Loan CC*, which did not reach significance.

The two additional weight-to-stress/stress-to-weight constraints in the final grammar represent the rendering of vowels. *If Loan VV, Donor Primary Stress* represents the preference for long vowels in the output forms to stem from primary stressed vowels in the input forms. *If Donor Stress, Loan VV* represents the preference for primary or secondary stressed vowels in the input forms to be adapted as long vowels in the output forms. As we saw in section 4, all primary stressed input vowels are adapted as long vowels, and 86% of secondary stressed vowels are adapted as long vowels. These two constraints work together to account for this behavior.

The constraint *Finnish Stress to Finnish Geminate* also reaches significance. As discussed in section 5.2, this constraint was included to account for the preference for gemination to occur when Finnish stress and donor stress coincide. The constraint works in concert with *If Loan CC, Donor Stress* to account for the fact that 90% of geminates occur when donor stress and Finnish stress coincide, or, on the boundary between the first and second syllables. The idea that there is something special about the boundary between the first and second syllable is not unprecedented. Karlsson (2005) claims that 84% of initial syllables in native disyllabic nouns are polymoraic, the most common sequence being CVC.CV and the second CVV.CV. And CCC sequences are most commonly found between the first and second syllables (Suomi et al. 2008).

ADDITIONAL CONSTRAINTS

The constraint *Super Heavy Syllable also reaches significance. This accounts for the dispreference in the data of the super heavy syllables [.aap.] and [.paap.]. Lastly, the constraint *V# reached significance, accounting for a preference for final long-vowels.

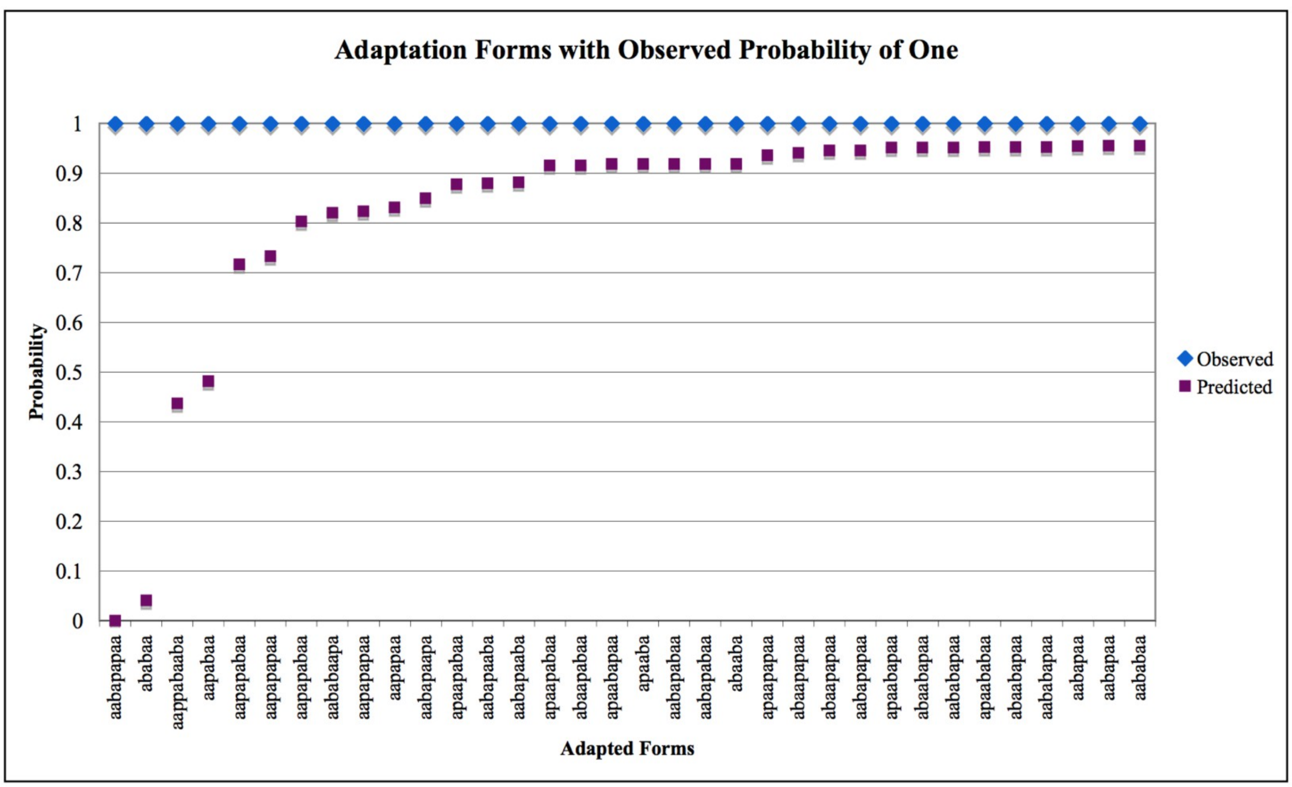
5.4 Evaluating the Grammar

This section will evaluate the predictive power of the grammar described in section 5.3. To best evaluate the predictions of the grammar, I will break the grammar up into sections based on the observed probability of the adapted forms. We will see that the grammar more accurately predicts forms that have an observed probability of 1 or 0, that is, when a donor word is either always or never rendered as a particular loan candidate form (candidate forms are GEN inputs, see section 5.1). The grammar performs worse in predicting adaptation variation, that is, when a donor form can be rendered as more than one possible loan form.

<u>Observed Probability of One</u>

A total of 34 forms in the data have an observed probability of one. In these forms, the grammar has a fairly strong predictive power. Of these 34 forms, 29 forms (85%) have a prediction rate of over .8 and 21 forms (62%) have a predicted probability of over .9. Graph 5.4.1 plots each adapted (loan) form with its observed and predicted probability.

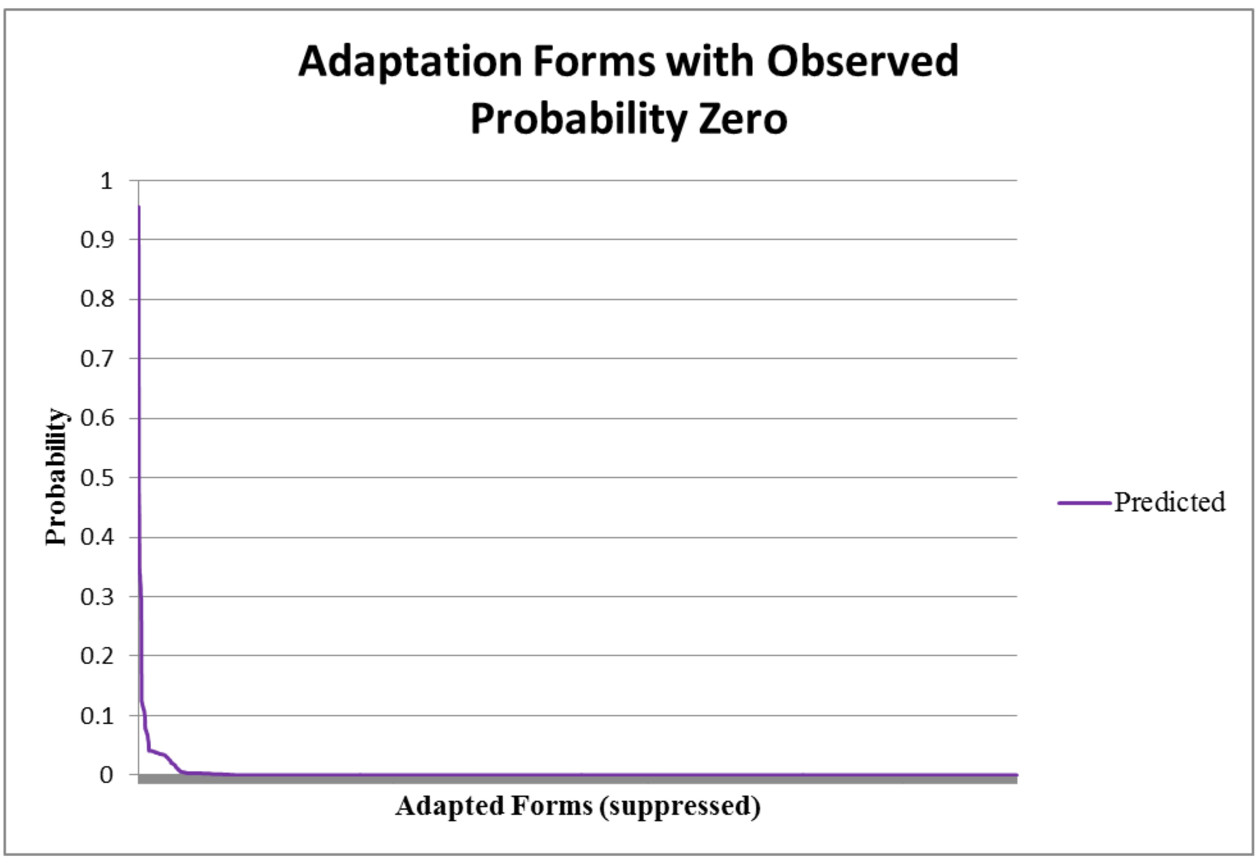
Graph 5.4.1 Wug-Tested Forms with Probability of 1



Observed Probability of Zero

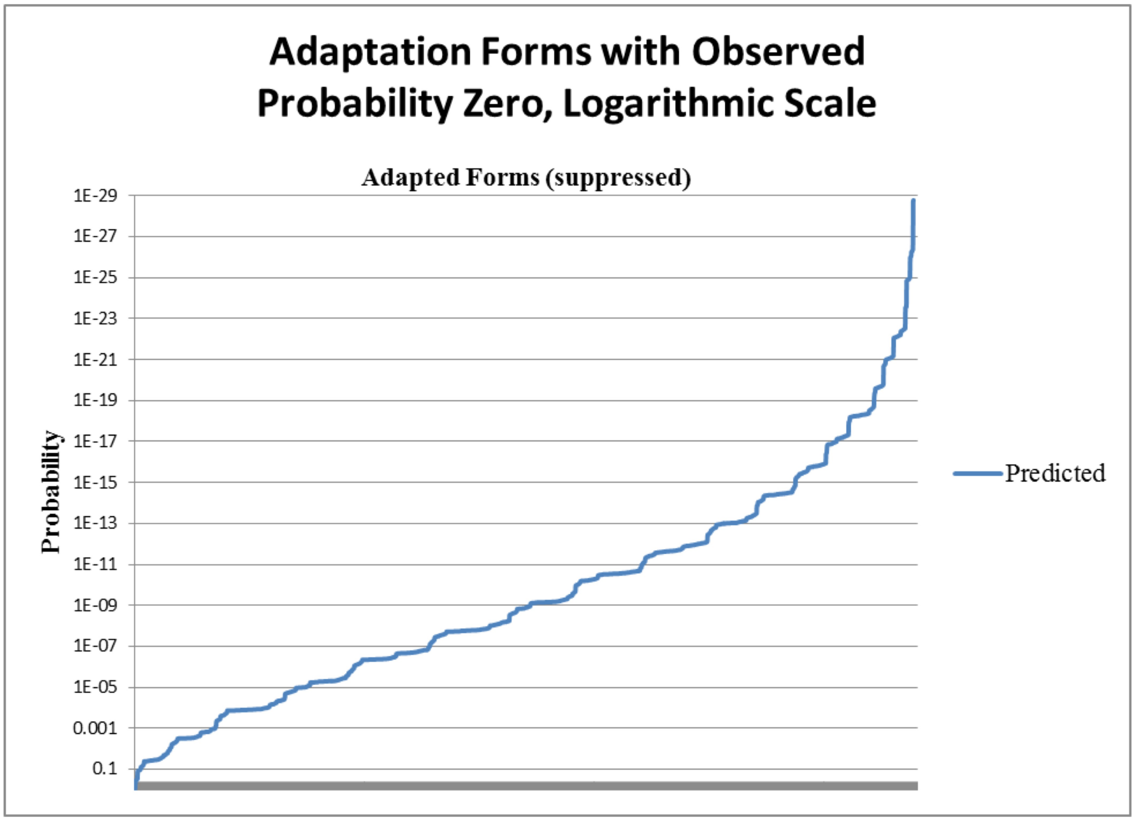
The grammar also did well in predicting forms with an observed probability of zero. There are a total of 2,635 forms such forms in the data. Only 31 of these forms have a predicted probability of greater than .05; only nine have a predicted probability of over .15. All forms are plotted in graphs 5.4.2 and 5.4.3. Because the forms are condensed around the near-equivalents of zero, Graph 5.4.3 shows the forms plotted using a logarithmic scale with base ten. Graph 5.4.2 is included to give a more easily observable big-picture sense of the distribution of the data. The form labels (x-axis) are suppressed on these graphs for the sake of readability.[21]

Graph 5.4.2 Forms with Observed Probability of Zero



21 Also for readability, a line graph is used here instead of a scatter plot; a plot with 2,635 individual data points is illegible on this scale.

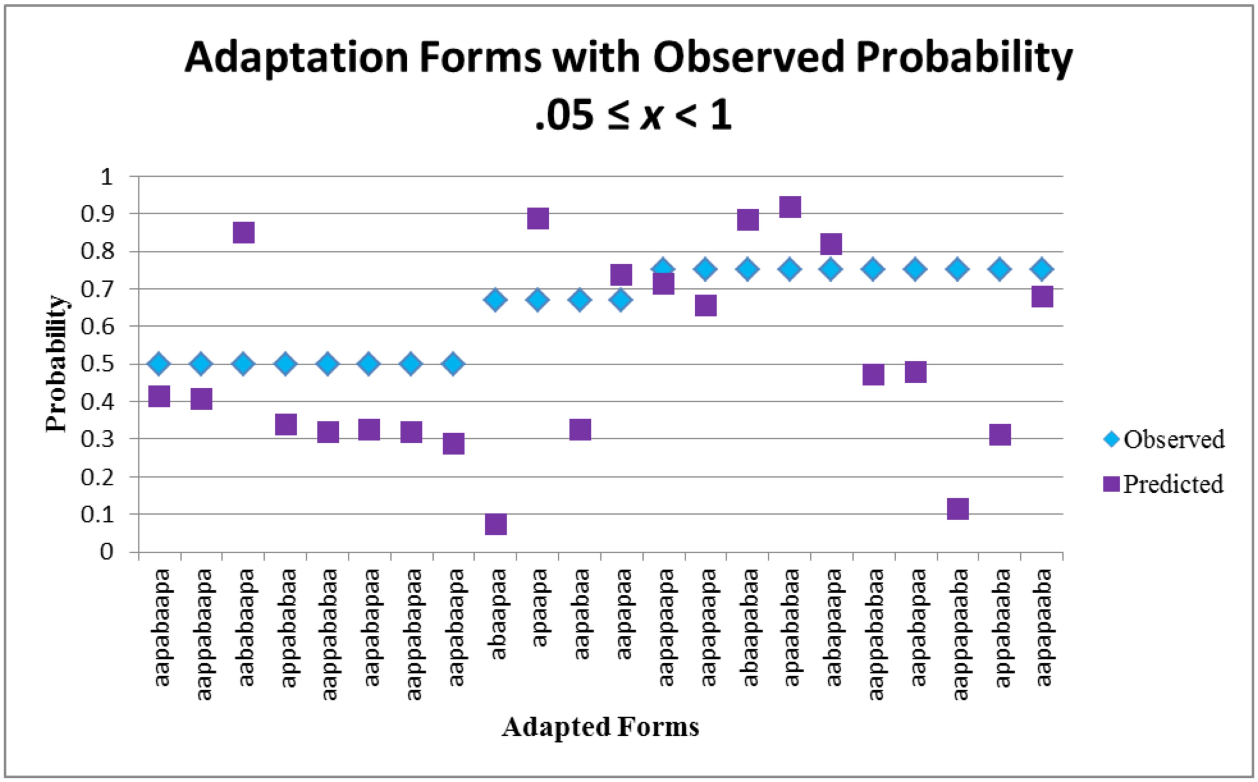
Graph 5.4.3 Forms with Observed Probability of Zero, Logarithmic Scale



Variation in Adaptation Forms

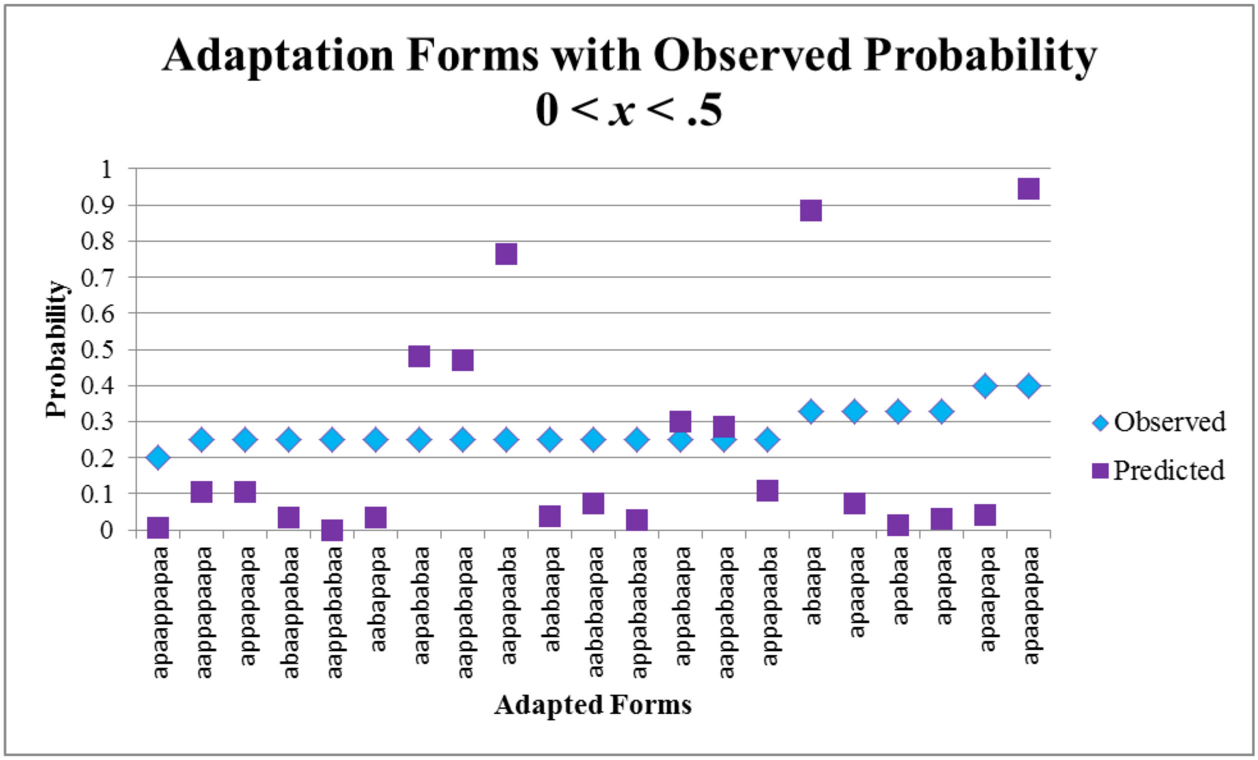
The grammar does less well in predicting forms that show variation in how they are adapted. These are forms in which a single input pattern was adapted as different output forms. For example, the input loan form [a.ˈpa.pa.ˌpa] was adapted twice as the loan form [a.paa.pa.pa] and twice as the loan form [a.paa.pa.paa], for an observed probability of .5 and .5, respectively. Graph 5.4.4 shows one section of this group of data. Specifically, it shows all surface forms *x* such that *x* has the observed probability $.05 \leq x < 1$.

Graph 5.4.4 Wug-Tested Forms with Probability $.05 \leq x < 1$



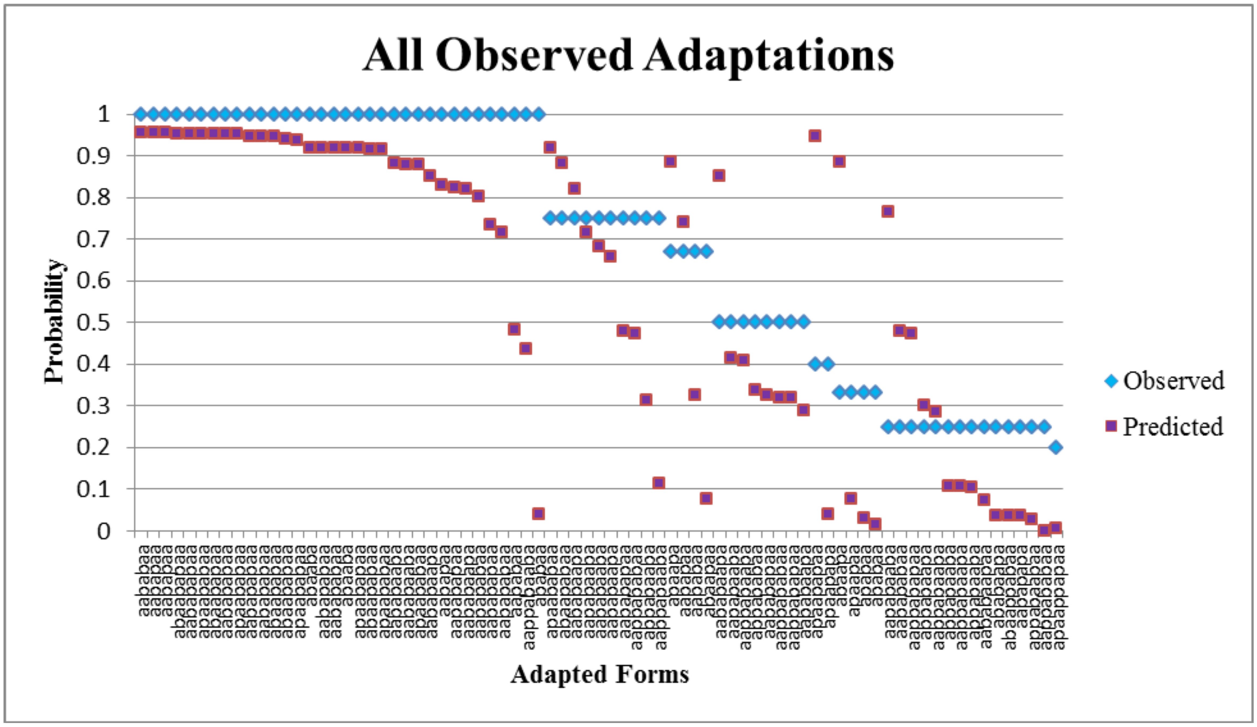
The data that are worst predicted by the grammar are those forms that have an observed probability of greater than 0 but less than .5, i.e. forms that reflect adaptation strategies that surfaced but were uncommon. Graph 5.4.5 shows all surface forms x such that x has the observed probability $0 < x < .5$.

Graph 5.4.5 Wug-Tested Forms with Probability $0 < x < .5$



The ability of the grammar to more accurately predict those surface forms that are observed 100% or 0% of the time is not unexpected given the paucity of the data for variation forms (i.e. the most times any adaptation form could be realized was 4). It is expected that when more data is added into the model the variation numbers would smooth out, i.e. the more data that one has the greater the chance of the set providing an accurate sample of adaptation strategies. This larger data set would provide the grammar with a better chance of having greater predictive power. A complete graph of the observed adaptation data is shown in Graph 5.4.6.

Graph 5.4.6 All Observed Adaptation Forms



6. A Closer Look at Partial Consonant Gradation

In section 4, we observed that, in the wug-tested nonce data, there appeared to be a preference for gemination before [b] segments as opposed to before [p] segments. In fact, holding other factors constant, a voiceless stop followed by a [b] had a 58% chance of geminating during the adaptation process; however, if the voiceless stop is followed by another voiceless stop, then it had only a 16% chance of geminating during the adaptation process. This observation was preliminarily called the “stop effect.” In section 5, the stop effect was formalized into the Partial Consonant Gradation (PCG) constraint and received a statistically significant weight in the maxent grammar. This result reinforces the original observation and confirms that the PCG constraint is operative in the data set. In this section, I will explore the origins of the constraint. In determining the origin, there are two possibilities. The first is that the constraint is a phonotactic constraint of native Finnish that has gone unnoticed. The second is that the constraint results from the perceptual process whereby Finnish speakers adapt

English strings into Finnish strings. Both possibilities are explored here and lead to the conclusion that the second possibility is the most likely explanation.

6.1 Partial Consonant Gradation as a Phonotactic Constraint in Nonce Dataset

The first step in determining whether the PCG constraint is a phonotactic constraint of native Finnish is to test whether the PCG constraint is a phonotactic constraint in the nonce data. This step is crucial because it will serve as a comparison case when testing whether the constraint is operative in the native grammar. Because the constraint was operative in the nonce maxent grammar, we expect it to also be operative in the nonce phonotactic grammar. The phonotactic grammar for the nonce dataset, as well as for the native Finnish dataset in section 6.2, was created using the UCLA Phonotactic Learner (hereafter Learner) (Capodieci et al. 2008).

Learning Data

The Learner is fed a data corpus, "learning data," and formulates the phonotactic constraints that are implicit in the corpus. The learning data here is the corpus of adapted nonce forms.

Feature Chart

The Learner also requires a feature chart file that defines data segments according to their phonetic properties. The feature chart used here is a simple chart containing only the phonemes in the nonce data. Geminates are treated as separate inventory entries specified with the feature "geminate."

Constraints

Once fed the learning data and feature chart, the Learner outputs a grammar containing the phonotactic constraints implicit in the data. The constraint we are particularly interested in, the PCG constraint, is formulated as the phonotactic constraint in (12):

(12) *[+geminate,+stop][-long_vowel][-geminate,+stop]

The phonotactic grammar for the nonce data is shown in Table 6.1. The PCG constraint, bolded in the table, receives a weight of 1.3. This finding confirms the hypothesis that the PCG constraint is phonotactically real in the nonce data.

Table 6.1 UCLA Phonotactic Learner Grammar: Wug-Test Data

Constraint	Weight
*[+consonantal][+word_boundary]	8.72
*[+consonantal][+consonantal]	8.49
*[+word_boundary][+consonantal]	8.47
*[-consonantal][-consonantal][-consonantal]	7.71
*[+geminate][-word_boundary][-consonantal]	6.45
*[+geminate][-word_boundary][+word_boundary]	5.36
*[+consonantal][-word_boundary][+geminate]	5.27
*[+word_boundary][-word_boundary][+geminate]	1.49
***[+geminate][-word_boundary][+voiceless_stop]**	**1.3**
*[+consonantal][-word_boundary][+word_boundary]	1.27
*[+word_boundary][-word_boundary][+consonantal,-voiceless_stop]	1.19
*[+word_boundary][-word_boundary][+voiceless_stop,-geminate]	0.92
*[+consonantal,-voiceless_stop][-word_boundary][+word_boundary]	0.36
*[+consonantal,-voiceless_stop][-word_boundary][+consonantal,-voiceless_stop]	0.07
*[+geminate]	0.03

6.2 Partial Consonant Gradation as a Phonotactic Constraint in Native Finnish

The next step after determining that the PCG constraint is operative in the nonce data is to determine whether the constraint is also operative in the native Finnish lexicon.

<u>Learning Data</u>

The learning data is a corpus of native Finnish words pulled from corpora at Finland's CSC Language Bank (accessed at www.csc.fi/kielipankki). Corpora and search criteria are listed in Appendix A. Since the corpora are newspapers, many loanwords were present in the initial data. The loanwords were weeded out by systematic searches of loanword-specific segments (such as [b] and complex onsets) as well as manual “eyeballing” of the data. The final corpus contains 12,911 types.

<u>Feature Chart</u>

As Finnish spelling is phonemic, only slight changes were made for segmentation in the data. The segment A is used for /ä/ due to complications with font. O is used for /ö/ for the same reason. The segment [d] is used for IPA /ɾ/[22] and [v] for IPA labiodental glide /ʋ/. The cluster [ng] is treated as a single segment and is phonetically realized as IPA /ŋ/.[23] All geminates are treated as separate inventory entries with the feature "+ geminate." Long vowels are also treated as separate inventory entries with the feature "+long." Consonants are not specified (value=0) for [long] and vowels are not specified for [geminate].

Constraints

One feature of the Learner is that it allows you to provide it with pre-constructed constraints to which it will assign weights according to the constraints' fit with the learning data provided. The phonotactic constraints that I created used native Finnish phonotactic information from Suomi et. al (2008).[24] Seventy-five constraints, not including the PCG constraint, were included in the grammar. Constraints were created using natural classes that were assigned to segments by the Learner. One helpful output file of the Learner is a list of nonce words that are created using the constraint grammar. The list gives the user an idea of how good the current grammar is at constructing a language like that the user is trying to emulate. For creating a grammar that will emulate native Finnish, test grammars were refined until the Learner produced a list that emulated possible words of Finnish (as judged by me and a native Finnish consultant). The final grammar, feature chart, and output nonce words file are attached in Appendix D.

Constraint Results

The phonotactic grammar of native Finnish yielded two interesting results. The first is that there is no strong evidence that the PCG constraint is operative in the native Finnish grammar. The second is that there is evidence that Quantitative Consonant Gradation, in its current formulation, is an overgeneralization that obscures two distinct constraints which are violated at different rates in the native Finnish lexicon.

22 [d] is analyzed as IPA /ɾ/ per Karvonen (2005), Suomi (1980), and Karlsson (1983).
23 See footnote 3.
24 Native constraints were developed independently of the constraints generated for the nonce data set.

PARTIAL CONSONANT GRADATION

Once the phonotactic grammar of native Finnish had been created, the PCG constraint was added. If the constraint received weight, then the result would constitute evidence that the PCG constraint is operative in the native Finnish grammar as well as in the loanword grammar.[25] If the constraint didn't receive weight, then the result would constitute evidence that the PCG constraint is not operative in the native Finnish grammar, unlike in the loanword grammar. As Table 6.2.1 shows, the constraint received the small weight of .13, which is likely not statistically significant. Compared to the constraint's weight in the nonce data grammar, which was 1.3 (given in Table 6.1), we can conclude that the PCG constraint is operative in the nonce data but not in the native Finnish lexicon.

Table 6.2.1 PCG Constraint Result in Native Finnish Phonotactic Grammar

Constraint	Weight	Num. Forms Violated
*[+geminate][-word_boundary][+voiceless_stop]	0.13	113

Because the constraint received some, albeit small, weight, the result raises the question of whether stops in the position following a geminate stop behave differently than other consonants in the same position, that is, whether voiceless stops are somehow special in this position. To explore this question, additional constraints that tested variations on the PCG constraint were added to the phonotactic grammar. Two of these constraints that received weight are listed in Table 6.2.2. These constraints have the same formulation as the PCG constraint, except that instead of a voiceless stop following the geminate there is a tap and a fricative, respectively. Both of these constraints received more weight than the PCG constraint, providing additional evidence that the PCG constraint does not have a unique effect in the grammar.

Table 6.2.2 Variations on the PCG Constraint in Native Finnish Phonotactic Grammar

Constraint	Weight	Num. Forms Violated
*[+geminate,+stop][-long_vowel][+tap]	2.6	0
*[+geminate,+stop][-long_vowel][+fricative,-geminate]	0.31	70

25 We saw that the PCG constraint was operative in the maxent loanword grammar in section 5, and we confirmed that it was phonotactically real when the constraint received weight in the phonotactic grammar in section 6.1.

Four additional constraint variations on the PCG constraint were also tested; these are listed in Table 6.2.3. None of these constraints received weight in the phonotactic grammar.

Table 6.2.3 Variations on the PCG Constraint in Native Finnish Phonotactic Grammar

Constraint	Weight	Num. Forms Violated
*[+geminate,+stop][-long_vowel][+trill,-geminate]	0	58
*[+geminate,+stop][-long_vowel][+glide]	0	161
*[+geminate,+stop][-long_vowel][+liquid,-geminate]	0	326
*[+geminate,+stop][-long_vowel][+nasal,-geminate]	0	1014

As this section has shown, the PCG constraint receives a small weight in the native Finnish phonotactic grammar. The weight is significantly smaller than the weight it receives in the nonce data phonotactic grammar. Additionally, two minimally different constraints received a greater weight in the native Finnish grammar, suggesting that stops are not unique in their behavior in this position. These results provide strong evidence that the PCG constraint is not operative as a phonotactic constraint in the native Finnish lexicon.

QUANTITATIVE CONSONANT GRADATION

This section will discuss an observation made about the Quantitative Consonant Gradation constraint and will present evidence that the constraint, in its current formulation, is an overgeneralization that obscures two distinct constraints which are violated at different rates in the native Finnish lexicon. As we remember from section 1.4, Quantitative Consonant Gradation (QCG) causes lenition of a geminate voiceless stop[26] in the onset of a heavy syllable. While the standard QCG constraint makes use of syllable boundaries and weight, the software used here was limiting in this regard. The data could not be syllabified because geminates were treated as a single feature and thereby could not straddle a syllable boundary. The QCG constraint was therefore broken up into four separate constraints in the phonotactic grammar; the constraints are given in (13)-(16).

(13) *[+geminate,+stop][+syllabic][+geminate,+stop]

26 There are no voiced stops in the data here, so voice is not specified in the natural classes (see native inventory and footnotes above).

Constraint (13) disallows sequences such as /p.pap.p/ and /p.paap.p/. It received a healthy weight of 2.5 and is violated in only one form in the learning data, [to**p.puut.t**e.le.mas.sa] (constraint violation bolded). Contrasting this constraint with one that does not specify geminate stops confirms the divergent behavior of geminate stops from that of other geminates. The constraint *[+geminate][+syllabic][+geminate] receives a weight of .65, likely the result of overlap with constraint (1), and is violated in 321 forms in the data, e.g. [a.loi.tel.les.saan]. This confirms that QCG targets specifically geminate stops, not geminates in general.

(14) *[+geminate,+stop][+long_vowel]

Constraint (14) disallows sequences such as /p.paa/, and therefore partially overlaps with the sequences disallowed by constraint (13). It received a weight of .42 and is violated in 441 forms in the data, e.g. [han.ka.loit.taa].

(15) *[+geminate,+stop][+syllabic][-syllabic][-syllabic]

Constraint (15) disallows sequences such as /t.tan.s/ and /t.taan.s/. Note that because geminates are specified with a feature, constraint (15) does not overlap with constraint (13). The constraint received a weight of 1.1 and is violated in 60 forms in the data, e.g. [a.jok.kin.sa].

(16) *[+geminate,+stop][+syllabic][-syllabic][+word_boundary]

Constraint (16) disallows sequences such as /p.pas#/ and /p.paas#/. It received a weight of .97 and is violated in 81 forms in the data, e.g. [ä.lyk.kyys].

An interesting result of this unstandard treatment of QCG, i.e. breaking the constraint up into four constraints that are sensitive to separate environments, is the discovery that the constraint is not equally powerful in all its environments. In fact, it is significantly lopsided in terms of its phonotactic violations. It is almost universally obeyed in sequences of geminate stops—constraint (13)—but has frequent exceptions when syllable weight is created by a long vowel—constraint (14)—or two non-geminate consonants—constraint (15). The difference between constraint (13) and (14) is likely due to Kiparsky's observation (2003) that unstressed long vowels should be parsed as disyllabic. Because the data here are not syllabified, and because secondary stress is not surface observable, it it likely that many of these apparent violations are not actual violations. The difference in constraint violations between (13) and (15), in which weight is created by the first half of a geminate or by a non-geminate consonant, respectively, is notable.

The lopsidedness of the distribution of QCG suggests that the constraint could be broken into separate constraints, one of which would be quite powerful (constraint 13), and one which would be more commonly violated (constraints 14-16). Additional support for the hypothesis that constraint (13) should be separated from the other constraints making up QCG is that it is attested as a separate constraint in other languages. Notably, constraint (13) is the same constraint represented by the Lex Mamilla effect in Latin, which does not have general lenition in the onset of a heavy syllable. As shown in following example, Lex Mamilla results in the de-gemination of a geminate when affixed with another geminate (Ito and Mester 1998):[27]

mamma (breast) + lla (diminutive) → mamilla *mammilla

The data presented here constitute evidence that much of the power of QCG comes from the Lex Mamilla effect, while the other conditions of QCG act as much less powerful environmental constraints.

6.3 Partial Consonant Gradation as a Perceptual Effect

In the previous two sections, we explored and rejected the possibility that the PCG constraint is a phonotactic constraint of native Finnish that had previously gone unnoticed. While section 6.1 confirmed that the PCG constraint is operative in the phonotactic grammar of the nonce data, section 6.2 showed that the constraint is not operative in the phonotactic grammar of native Finnish. In the process of exploring this option, however, the native phonotactic grammar revealed an interested fact about QCG, which is that it is heavily respected in its Lex Mamilla environments, while more frequently violated in its other environments. This section will explore the second option that was

27 I was unable to find an example of lex mamilla over a long vowel; in which case this stronger formulation of lex mamilla, *[+geminate,+stop][+syllabic, -long][+geminate,+stop] is unviolated in the Finnish data.

posited at the beginning of section 6 as an account of PCG, which is that the constraint results from the perceptual process whereby Finnish speakers adapt English strings into Finnish strings. Specifically, I will argue that PCG can be explained as a perceptual effect that stems from the relative duration of voiced vs. voiceless stops.

PERCEPTUAL EFFECTS OF STOP DURATION

Lisker first observed (1957, 1969) that English voiceless stops are regularly longer in duration of closure than voiced stops. Durational differences alone cannot, however, account for the fact that [p] segments favored gemination before [b] segments as opposed to before other [p] segments. Port et al. (1982), though, demonstrates that *relative* duration of segments functions as a perceptual cue for phonological voicing in English stops. They found that absolute duration of the segments actually had very little (an increase of about 2%) explanatory power. Using these two observations, we now have a picture in which we can hypothesize that when a Finn hears the form ['a.pa.ba], the longer relative duration of the [p] segment compared to the [b] segment triggers a perceptual effect that causes the [p] more likely to be rendered as a geminate.[28] This solution of course requires that the properties of a particular segment in a word can influence the perception of those segments that preceded it. Brown et al. (2012) provide us with experimental evidence demonstrating that this is the case. That is, perceptual cues can work "backwards;" a segment perceived at time *x* can affect what a listener believes they perceived at time *x-1*. Thus, a [b] segment following a [p] segment of relatively longer duration can influence the way the speaker perceives and renders the [p] segment. In this case, the Finnish speaker renders the [p] as a geminate [pp] in order to account for the different duration lengths perceived between the two segments.

THEORIES OF LOANWORD ADAPTATION AS PERCEPTION

There are two theoretical approaches to my knowledge in which the perception explanation of PCG could be explored. The first approach posits a loanword-specific constraint along the lines of

28 Thank you to Bruce Hayes for proposing this as a possible explanation. No phonetic analysis of the specific elicitation forms was done here.

Peperkamp 2005 to account for the fact that the adaptation strategy doesn't apply as a constraint in the native Finnish grammar. Peperkamp provides examples of phonological differences between native phonology and loanword phonology, arguing that loanword adaptations do not fall out of the native phonology and therefore must be modeled in a separate system. Instead of being part of the native phonology, she argues, loanword adaptations are phonetically minimal transformations that are applied during speech perception. PCG in this case would fall under her classification of "unnecessary adaptations," namely an adaptation that doesn't exist to repair an ill-formed phonotactic structure. Instead, PCG would be accounted for under the phonetic transformations that occur from the donor string to the adapted string.

A similar yet separate approach is that of Boersma's (2009) system, in which he rejects the distinction between loanword phonology and native phonology, arguing that both are integrated into a single system. In this system, loanword adaptation is to be entirely understood in term of phonological and phonetic comprehension and production mechanisms in the speaker's native language. Specifically, the interaction of cue constraints—operating on the relation between perceptual input and phonological surface forms—and structural constraints on the comprehension side, and the interaction of faithfulness constraints and structural constraints on the production side, accounts for the apparent differences between phonology in loanwords and in native words. Unlike in Peperkamp's system, adaptations do not occur because of the difficulty speakers have in perceiving non-native sound patterns, but from a "best-guess" system, colored by native phonology, in which listeners attempt to retrieve a speaker's intended surface form. In this case, the perceived difference in relative length between voiced and voiceless stop would be retrieved by the Finnish speaker, whose phonological system is sensitive to consonant length, as cues indicating a difference in consonant length.

6.4 Additional Explanations of Partial Consonant Gradation

Section 6.3 presented evidence that PCG can be explained as a perceptual effect caused by the durational differences of voiced and voiceless stops. However, because the data used here is limited there are two other possibilities of accounting for the data that need to be mentioned.

Phonetic Effects

A possible alternative explanation of PCG is that the phenomenon is entirely explainable in terms of phonetic variation in the pronunciation of the donor nonce forms. Under this explanation, the PCG effect was created from variation in the pronunciation of particular tokens which by chance correlated with one group of forms more than the other. This possibility would be tested by phonetically analyzing the recordings of the elicitation sessions to determine whether the donor stimuli that was mapped onto geminate stops in the adaptation forms are phonetically distinct from those that did not. Specifically, to see whether the stops that geminated were uttered with greater force than the forms that were not geminated. In this were indeed the case, it would be expected that repeated sessions using the same recorded elicitations would produce a similar result with the same consultant and across sessions with different consultants. In this case, a constraint that is sensitive to the phonetic properties of the input could account for the variation seen. While this explanation is possible and therefore must be posited, the set-up of the experiment does mitigate against the possibility of accidental correlation between groups of forms. As mentioned in section 4.1, each donor form was presented to the consultant multiple times (four times for four syllable forms and three times for three syllable forms) with each time as a separate spoken elicitation. This setup means that, for example, the donor form [ˈa.pa.ˌba.ba] was adapted four times as the loan form [aap.pa.baa.ba]. The repetition of the gemination of certain forms provides a greater certainty that the adaptation strategy for this particular form is indeed real.

Random Result

The third explanation is that the gemination of certain forms was a completely random result and would not be repeated if the same elicitation recordings were used across multiple sessions with the same consultant or with sessions with different consultants. In this case, the phenomenon would not exist as a perceptual effect; it is unclear in this case what the explanation for the gemination would then be.

7. Conclusion

This project looked at loanword adaptation in Finnish, focusing on the gemination of voiceless stops. It showed that previous theories of gemination in Finnish loanword adaptation that focused on structural accounts of gemination missed perceptual effects that are influential in the adaptation process. Specifically, I presented evidence using on-line adaptations of nonce forms that gemination is influenced by the stress of the donor word as well as Finnish primary stress. Additionally, I demonstrated that gemination is influenced by a perceptual effect caused by the difference in relative length between [p] and [b] segments. I also argued that Quantitative Consonant Gradation in the native Finnish lexicon could more accurately be broken up into two separate constraints, one accounting for strong Lex Mamilla effects and one accounting for the environments of QCG that are more commonly violated.

Appendix A: Native Finnish Corpus

Below are listed the specific searches used to create a corpus of native Finnish loans. See Section 2.1, 5.2.

Corpora	Aamulehti 1995
	Aamulehti 1999
	Turun Sanomat 1999
	Turun Sanomat 1998
	Kaleva 1998-1999
	Hyvinkaan Sanomat 1997
	Keskisuomalainen 1999
Search Criteria	Nouns, singular and plural; [bf="*" pos='Noun' case='Nom']
Tokens Selected	First 2,000 hits (including duplicates)

Corpora	Aamulehti 1995
	Aamulehti 1999
	Turun Sanomat 1999
	Turun Sanomat 1998
	Kaleva 1998-1999
	Hyvinkaan Sanomat 1997
	Keskisuomalainen 1999
Search Criteria	[wf="*" pos='Verb' modality='INF']
Tokens Selected	First 2,000 frequency hits (including duplicates)

Corpora	Aamulehti 1995
	Aamulehti 1999
	Turun Sanomat 1999
	Turun Sanomat 1998
	Kaleva 1998-1999
	Hyvinkaan Sanomat 1997
	Keskisuomalainen 1999
Search Criteria	[bf="*" pos='Adjective' case='Nom']
Tokens Selected	First 2,000 frequency hits (including duplicates)

Appendix B: Nonce Wug-Tested Data

The table includes the English wug form, "input," the Finnish adapted form, "candidate," and the frequency of each given adapted form, "Num. Observed." Forms with gemination are in bold. See Section 4.2.

Input	Candidate	Num. Observed
a.ˈpa.pa.ˌpa	**apaappapaa**	1
	apaapapa	2
	apaapapaa	2
a.ˌpa.pa.ˈpa	apaapapaa	4
ˈa.pa.ˌpa.pa	**aappapaapa**	1
	aapapaapa	3
ˌa.pa.pa.ˈpa	aapapapaa	4
ˌa.pa.ˈpa.pa	**appapaapa**	1
	aapapaapa	3
a.ˈba.ba.ˌpa	abaabapaa	4
a.ˌba.ba.ˈpa	abaabapaa	4
ˈa.ba.ba.ˌpa	aababapaa	4
ˈa.ba.ˌba.pa	aababaapa	4
ˌa.ba.ba.ˈpa	aababapaa	4
ˌa.ba.ˈba.pa	ababaapa	1
	aababaapaa	1
	aababaapa	2
a.ˈba.pa.ˌba	abaapabaa	4
a.ˌba.pa.ˈba	**abaappabaa**	1
	abaapabaa	3
ˈa.ba.pa.ˌba	aabapabaa	4
ˈa.ba.ˌpa.ba	aabapaaba	4
ˌa.ba.pa.ˈba	aabapabaa	4
ˌa.ba.ˈpa.ba	aabapaaba	4
a.ˈba.pa.ˌpa	abaapapaa	4
a.ˌba.pa.ˈpa	abaapapaa	4
ˈa.ba.pa.ˌpa	aabapapaa	4
ˈa.ba.ˌpa.pa	aabapapa	1
	aabapaapa	3
ˌa.ba.pa.ˈpa	aabapapaa	4
ˌa.ba.ˈpa.pa	aabapaapa	4
a.ˈpa.ba.ˌba	apaababaa	4
a.ˌpa.ba.ˈba	**aappababaa**	1
	apaababaa	3
ˈa.pa.ba.ˌba	**aappababaa**	3
	aapababaa	1
ˈa.pa.ˌba.ba	**aappabaaba**	4
ˌa.pa.ba.ˈba	**appababaa**	2
	aappababaa	2
ˌa.pa.ˈba.ba	**appabaabaa**	1

	appabaaba	3
a.ˈpa.ba.ˌpa	apaabapaa	4
a.ˌpa.ba.ˈpa	apaabapaa	4
ˈa.pa.ba.ˌpa	**aappabapaa**	1
	aapabapaa	3
ˈa.pa.ˌba.pa	**aappabaapa**	2
	aapabaapa	2
ˌa.pa.ba.ˈpa	**aappabapaa**	2
	aapabapaa	2
ˌa.pa.ˈba.pa	**appabaapa**	1
	aappabaapa	1
	aapabaapa	2
a.ˈpa.pa.ˌba	apaapabaa	4
a.ˌpa.pa.ˈba	apaapabaa	4
ˈa.pa.pa.ˌba	aapapabaa	4
ˈa.pa.ˌpa.ba	**aappapaaba**	3
	aapapaaba	1
ˌa.pa.pa.ˈba	aapapabaa	4
ˌa.pa.ˈpa.ba	**appapaaba**	1
	aapapaaba	3
ˈa.pa.pa.ˌpa	aapapapaa	4
a.ˈba.ba	abaaba	1
ˈa.ba.ˌba	aababaa	1
ˌa.ba.ˈba	ababaa	1
a.ˈpa.pa	apaapaa	1
	apaapa	2
ˈa.pa.ˌpa	aapapaa	3
ˌa.pa.ˈpa	apapaa	1
	aapapaa	2
a.ˈpa.ba	apaaba	3
ˈa.pa.ˌba	aapabaa	3
ˌa.pa.ˈba	apabaa	1
	aapabaa	2
a.ˈba.pa	abaapa	1
	abaapaa	2
ˈa.ba.ˌpa	aabapaa	3
ˌa.ba.ˈpa	aabapaa	3

Appendix C: ChiDist Significance Tests of Constraints

The table below gives the significance tests result for each constraint that received weight in the grammar listed in Appendix B. See Section 5.3.

Constraint	**P-value**
If loan CC, donor secondary stress	0.98
If donor secondary stress, loan CC	0.98
If donor stress loan CC	0.99
If loan singleton C, donor unstressed	0.99
If donor primary stress, loan CC	0.99
*HH	0.86
If loan VV, donor stress	0.99
If loan singleton C, donor unstressed	0.99
Ident(C Length)	0.65
If loan CC, donor primary stress	0.26
If donor unstressed, loan short V	0.99
If donor unstressed, loan singleton C	0.99
If loan short V, donor unstressed	0.99
If donor primary stress, loan VV	0.69
Partial Consonant Gradation	8.80E-005
Consonant Gradation	0.0047
*Super Heavy Syllable	5.85E-020
Finnish Stress to Finnish Geminate	1.72E-012
*V#	0.0005
If loan VV, donor primary stress	3.57E-101
If loan CC, donor stress	3.63E-008
If donor stress, loan VV	6.57E-190

Appendix D: Final UCLA Phonotactic Learner Grammar and Feature Chart

The following chart gives a list of the phonotactic constraints of native Finnish as well as the weight that they were assigned by the Phonotactic Learner and the number of forms that violated each constraint in the learning data. See Section 6.2.

Constraint	**Weight**	**Num. Forms Violated**
*[+consonantal][+consonantal,-sonorant][+consonantal,-sonorant]	0	44
*[+geminate,+stop][-long_vowel][+trill,-geminate]	0	58
*[+consonantal][+consonantal][+consonantal]	0	71
*[+geminate,+stop][-long_vowel][+glide]	0	161
*[+geminate,+stop][+long_vowel][+word_boundary]	0	241
*[+geminate,+stop][-long_vowel][+liquid,-geminate]	0	326
*[+nasal,+coronal][-tap,-coronal]	0	653
*[+geminate,+stop][-long_vowel][-geminate,+nasal]	0	1014
*[+syllabic][+syllabic,+high]	0	5041
*[-nasal][+geminate]	0	6001
*[-nasal][+consonantal][+consonantal]	0	7759
*[+nasal,-labial][-fricative,+labial]	0.015	216
*[+geminate]	0.1	6136
*[+word_boundary][+glide,+labial][+high,-front]	0.117	53
*[+geminate,+stop][-long_vowel][-geminate,+stop]	0.134	113
*[-liquid][-syllabic][-syllabic]	0.256	42
*[+geminate,+stop][-long_vowel][+fricative,-geminate]	0.31	70
*[+long_vowel][+long_vowel]	0.335	0
*[+geminate,+stop][+long_vowel]	0.416	441
*[+front_harmonic][-front]	0.482	52
*[+nasal,-coronal][+coronal]	0.642	3
*[+geminate][+syllabic][+geminate]	0.653	321
*[-front][+front_harmonic]	0.685	29
*[+tap][+nasal]	0.862	0
*[+fricative,-consonantal][+word_boundary]	0.891	0
*[+nasal,-coronal][+word_boundary]	0.955	0
*[+geminate,+stop][+syllabic][-syllabic][+word_boundary]	0.965	81
*[-front][+neutral_harmonic][+neutral_harmonic][+front_harmonic]	1.056	68
*[+geminate,+stop][+syllabic][-syllabic][-syllabic]	1.096	60
*[+nasal][+liquid]	1.114	53
*[+front_harmonic][+neutral_harmonic][+neutral_harmonic][-front]	1.115	64
*[-liquid][+geminate]	1.144	135
*[-liquid,-nasal][-syllabic][+fricative,-consonantal]	1.146	0
*[+glide][+word_boundary]	1.274	0
*[-liquid,-nasal][+fricative,-consonantal][-syllabic]	1.44	0
*[+long_vowel][+geminate]	1.468	528
*[+front_harmonic][+neutral_harmonic][-front]	1.687	239
*[-syllabic][-syllabic][+geminate]	1.817	0
*[+consonantal][+sonorant,+velar]	1.823	0
*[-liquid,-nasal][-fricative,-stop][-fricative,-stop]	1.919	0
*[-front][+neutral_harmonic][+front_harmonic]	1.981	179
*[+word_boundary][+high,-front][+labial,-high,-front]	2.194	1
*[+word_boundary][+sonorant,+velar]	2.253	0
*[+word_boundary][-long_vowel][+word_boundary]	2.34	1
*[+sonorant,+velar][-syllabic]	2.351	0
*[+geminate,+stop][+syllabic][+geminate,+stop]	2.485	1
*[-syllabic][-syllabic][-syllabic]	2.627	71
*[+geminate,+stop][-long_vowel][+tap]	2.629	0
*[+front_harmonic][-front]	2.629	388
*[-high,-low,+long_vowel]	2.694	473
*[+syllabic][+syllabic][+syllabic]	2.72	125

*[+fricative,+consonantal][+tap]	2.806	1
*[+word_boundary][+high,+neutral_harmonic][+high,+neutral_harmonic]	2.864	0
*[+liquid,-trill,-geminate][+liquid,-trill,+geminate]	2.91	0
*[-front][+front_harmonic]	3.022	259
*[-fricative,+front][-fricative,+front]	3.025	0
*[+glide,+labial][+glide,+labial]	3.025	0
*[+nasal,+labial][+glide,+labial]	3.031	0
*[+fricative,+consonantal][+trill]	3.048	0
*[-sonorant,+high][+stop,+coronal]	3.05	0
*[+sonorant,+velar]	3.134	59
*[-fricative][+tap]	3.211	12
*[+stop][+nasal]	3.225	12
*[+word_boundary][-fricative,+front][+high,+neutral_harmonic]	3.411	0
*[+consonantal,-sonorant][+fricative,-consonantal]	3.515	4
*[+liquid,-trill][+trill]	3.518	0
*[-geminate,+stop,-labial][-geminate,-sonorant,+labial]	3.57	0
*[-geminate,-sonorant,+labial][-geminate,+stop,-labial]	3.575	0
*[+word_boundary][+tap]	3.589	10
*[+syllabic][+long_vowel]	3.666	86
*[+nasal,+labial][-syllabic,-labial]	3.947	3
*[+tap][-syllabic]	4.104	5
*[+stop,-coronal][+glide]	4.114	0
*[-liquid,-nasal][+geminate]	4.664	0
*[+long_vowel][+syllabic]	4.736	30
*[+tap][+word_boundary]	4.816	0
*[+word_boundary][-syllabic][-syllabic]	4.818	17
*[+glide][+consonantal]	5.274	0
*[-syllabic][+consonantal][+word_boundary]	5.42	0
*[-geminate,-coronal][+word_boundary]	5.951	0
*[+geminate][+word_boundary]	6.129	0
*[+word_boundary][+geminate]	6.163	0
*[+geminate][-syllabic]	6.205	0

The chart below is the feature chart created for the UCLA Phonotactic Learner of native Finnish. These features were used to generate the natural classes listed in the phonotactic constraints given above. Segments are listed along the left and features are listed along the top.

	velar	fricative	tap	glide	liquid	syllabic	trill	geminate	consonantal	sonorant	stop	nasal	voice	labial	round	coronal	high	low	front	back	front_harmonic	neutral	long_vowel	wordboundary
y	0	0	0	0	0	+	0	0	-	+	0	-	+	+	+	-	+	-	+	-	+	-	-	-
u	0	0	0	0	0	+	0	0	-	+	0	-	+	+	+	-	+	-	-	+	-	-	-	-
ø	0	0	0	0	0	+	0	0	-	+	0	-	+	+	+	-	-	-	+	-	+	-	-	-
o	0	0	0	0	0	+	0	0	-	+	0	-	+	+	+	-	-	-	-	+	-	-	-	-
i	0	0	0	0	0	+	0	0	-	+	0	-	+	-	-	-	+	-	+	-	-	+	-	-
e	0	0	0	0	0	+	0	0	-	+	0	-	+	-	-	-	-	-	+	-	-	+	-	-
æ	0	0	0	0	0	+	0	0	-	+	0	-	+	-	-	-	-	+	+	-	+	-	-	-
a	0	0	0	0	0	+	0	0	-	+	0	-	+	-	-	-	-	+	-	+	-	-	-	-
yy	0	0	0	0	0	+	0	0	-	+	0	-	+	+	+	-	+	-	+	-	+	-	+	-
uu	0	0	0	0	0	+	0	0	-	+	0	-	+	+	+	-	+	-	-	+	-	-	+	-
øø	0	0	0	0	0	+	0	0	-	+	0	-	+	+	+	-	-	-	+	-	+	-	+	-
oo	0	0	0	0	0	+	0	0	-	+	0	-	+	+	+	-	-	-	-	+	-	-	+	-
ii	0	0	0	0	0	+	0	0	-	+	0	-	+	-	-	-	+	-	+	-	-	+	+	-
ee	0	0	0	0	0	+	0	0	-	+	0	-	+	-	-	-	-	-	+	-	-	+	+	-
ææ	0	0	0	0	0	+	0	0	-	+	0	-	+	-	-	-	-	+	+	-	+	-	+	-
aa	0	0	0	0	0	+	0	0	-	+	0	-	+	-	-	-	-	+	-	+	-	-	+	-
p	-	-	-	-	-	-	-	-	+	-	+	-	-	+	-	-	0	0	0	0	0	0	0	-
pp	-	-	-	-	-	-	-	+	+	-	+	-	-	+	-	-	0	0	0	0	0	0	0	-
k	+	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	+	-	0	0	0	0	0	-
kk	+	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-	+	-	0	0	0	0	0	-
m	-	-	-	-	-	-	-	-	+	+	-	+	+	+	-	-	0	0	0	0	0	0	0	-
v	-	-	-	+	-	-	-	-	-	+	-	-	+	+	0	-	0	0	0	0	0	0	0	-
j	-	-	-	+	-	-	-	-	-	+	-	-	+	-	-	-	+	-	+	-	0	0	0	-
h	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	0	0	0	0	0	-
ng	+	-	-	-	-	-	-	-	+	+	-	+	+	-	-	-	+	-	0	0	0	0	0	-
t	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	+	0	0	0	0	0	0	0	-
tt	-	-	-	-	-	-	-	+	+	-	+	-	-	-	-	+	0	0	0	0	0	0	0	-
n	-	-	-	-	-	-	-	-	+	+	-	+	+	-	-	+	0	0	0	0	0	0	0	-
r	-	-	-	-	+	-	+	-	+	+	-	-	+	-	-	+	0	0	0	0	0	0	0	-
l	-	-	-	-	+	-	-	-	+	+	-	-	+	-	-	+	0	0	0	0	0	0	0	-
s	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	+	0	0	0	0	0	0	0	-
d	-	-	+	-	-	-	-	-	+	-	-	-	+	-	-	+	0	0	0	0	0	0	0	-
mm	-	-	-	-	-	-	-	+	+	+	-	+	+	+	-	-	0	0	0	0	0	0	0	-
nn	-	-	-	-	-	-	-	+	+	+	-	+	+	-	-	+	0	0	0	0	0	0	0	-
rr	-	-	-	-	+	-	+	+	+	+	-	-	+	-	-	+	0	0	0	0	0	0	0	-
ll	-	-	-	-	+	-	-	+	+	+	-	-	+	-	-	+	0	0	0	0	0	0	0	-
ss	-	+	-	-	-	-	-	+	+	-	-	-	-	-	-	+	0	0	0	0	0	0	0	-

The forms listed below are a sample of those created by the Learner using the constraint grammar provided to it. See Section 6.2.

Learner Generated Forms		
i u s a o h i r	o o j v ii k s O	p ii r A
e n A y pp y n O	m yy n O j e ss u	A A d i l i kk yy
y O m A y h A n	j AA m O s O nn O	j a n t e v j uu
p uu rr u n i h aa	u ll e A l m AA n	ii n t A kk AA p ii
k y y t y t r O	l ii h A O i rr i	r e r a r ii d uu
n o i nn o k a e	l e r uu d uu k aa	m y n i kk ee j yy
e o n i u d o s	yy n y k i mm e r	v i r e mm yy k e
l yy r y pp O v ii	O t j A s v y i	A mm e i ll A k e
i p i m e h m AA	s uu n o h j aa s	v a l r e h j a
o m oo l nn i s ii	e t s y j i e n	j e O ll yy l v ii
v i j a p uu l uu	h y m m AA r O r	v A y d AA s l e
AA m A k yy s t y	a h s u n u l a	m a ss u a t r aa
ii t e v a i n u	O O s O t yy k yy	h o r a a l ee l
p A l y s y y t	l i s yy v e v yy	v i h u a l k u
e m i l a t k aa	o s i l l uu p i	i o pp o e l ss u
s i pp i t A nn e	aa l n e e k l o	h i h j y p O t
ii p e O i p a l	e r r AA k i pp y	p ii p O O h l AA
a nn a n i k r ii	o l aa n uu t i o	i n u p e t r a
r e n O j j A r	t e m mm A m y t	AA l l O s s y e

References

Anttila, Arto.1997. Deriving Variation from Grammar. In Frans Hinskens, Roeland van Hout and Leo Wetzels (eds.), *Variation, Change and Phonological Theory,* 35-68. Amsterdam, John Benjamins.

Boersma, Paul and Silke Hamann. 2009. Loanword Adaptation as First-Language Phonological Perception. In Andrea Calabrese and W. Leo Wetzels (eds.), *Loanword Phonology*, 11-58.

Brown, Meredith, Laura C. Dilley, and Michael K. Tanenhaus. 2012. Real-time Expectations Based on Context Speech Rate can Cause Words to Appear or Disappear. In N. Myake, D. Peebles, and R. P. Cooper (Eds.) *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, 1,374-1,379. Austin, TX: Cognitive Science Society.

Campbell, Lyle. 1998. *Historical Linguistics: An Introduction*. Cambridge, MA: MIT Press.

Capodieci, Frank, Bruce Hayes, and Colin Wilson. 2008. Phonotactic Learning Program.

Goldwater, Sharon and Mark Johnson. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. In *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*.

Hanson, Kristin and Paul Kiparsky. 1996. A Parametric Theory of Poetic Meter. *Language* 72: 287-335.

Itô, Junko and Armin Mester. 1998. Markedness and Word Structure: OCP Effects in Japanese. Ms., University of California Santa Cruz.

Kang, Y. 2003. Perceptual Similarity in Loanword Adaptation: Adaptation of English Post-Vocalic Word-Final Stops to Korean. *Phonology* 20: 219: 273.

Karlsson, Fred. 1983. *Suomen kielen äänne- ja muotorakenne* [The phonological and morphological structure of Finnish]. Porvoo: WSOY.

Karlsson, Fred. 2008. *Finnish: An Essential Grammar*. New York, NY: Routledge.

Karvonen, Daniel. 1998. Finnish Loanword Phonology and the Core-Periphery Structure of the Lexicon. Handout, SWOT 4, University of Arizona, April 4.

Karvonen, Daniel. 2005. *Word Prosody in Finnish*. Dissertation, University of California, Santa Cruz.

Kiparsky, Paul. 2003. Finnish Noun Inflection. In Diane Nelson and Satu Manninen (eds.) *Generative Approaches to Finnic and Saami Linguistics*, 109-161. Stanford, CA: CSLI.

Kiparsky, Paul. 2008. Fenno-Swedish Quantity: Contrast in Stratal OT. In Bert Vaux and Andrew Nevins (eds.) *Rules, Constraints, and Phonological Phenomena*, Oxford: Oxford University Press.

Lehiste, Ilse. 1970. Suprasegmentals. Cambridge, MA. London: MIT Press.

Lisker, Leigh. 1957. Closure Duration and the Intervocalic Voiced-Voiceless Distinction in English.

Language. 33: 43-49.

Lisker, Leigh. 1969. Stop Duration and Voicing in English. Haskins Laboratory. SR 19/20: 27-35.

Peperkamp, Sharon. 2004. Lexical Exceptions in Stress Systems: Argument from Early Language Acquisition and Adult Speech Perception. *Language*, 80, 98-126.

Peperkamp, Sharon. 2005. A Psycholinguistic Theory of Loanword Adaptations. In M. Ettlinger, N. Fleischer, and M. Park-Doob (eds.), *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society*, 341-351.

Port, Robert F. and Jonathan Dalby. 1982. Consonant/Vowel Ratio as a Cue for Voicing in English. *Perception and Psychophysics*. 32(2): 141-152.

Prince, Alan and Paul Smolensky. 2002. Optimality Theory: Constraint Interaction in Generative Grammar.

Riad, Tomas. 1992. *Structures in Germanic Prosody: A Diachronic Study with Special Reference to the Nordic Languages*. Stockholm: Stockholm University.

Smolensky, Paul. 1986. Information Processing in Dynamic Systems: Foundations of Harmony Theory. In D.E. Rumelhart and J.L. McClelland (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 1:194-281. Cambridge: MIT Press.

Smolensky, Paul, Geraldine Legendre, and Yoshiro Miyata. 1992. Integrating Connectionist and Symbolic Computation for the Theory of Language. Report CU-CS-628-92, Computer Science Department at the University of Colorado at Boulder.

Suomi, Kari. 1980. *Voicing in English and Finnish Stops: A Typological Comparison with an Interlanguage Study of the Two Languages in Contact*. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 10, Department of Finnish and General Linguistics, University of Turku.

Suomi, Kari, Juhani Toivanen, and Riikka Ylitalo. 2008. *Finnish Sound Structure: Phonetics, phonology, phonotactics, and prosody*. Oulu, Finland: University of Oulu.

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf, and Patti J. Price. 1992. Segmental Durations in the Vicinity of Prosodic Phrase Boundaries. *Acoustic Society of America*. 91(3): 1707-1717.

Wilson, Colin. 2006. Learning Phonology with Substantive Bias: An Experimental and Computational Study of Velar Palatalization. *Cognitive Science*. 30(5): 945-982.