The Role of Perceptual Similarity and Gradient Phonotactic Well-Formedness in Loan Gemination Processes

by

Lilla Magyar

M.A./M.Ed., English Language and Literature, University of Pannonia (2009)
M.A/M.Ed., German Language and Literature, University of Pannonia (2009)

Submitted to the Department of Linguistics and Philosophy in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2017

Signature redacted

Author ....
Department of Linguistics and Philosophy
September 8, 2017

Certified by ..... Signature redacted ....
Adam Albright
Professor
Thesis Supervisor

Signature redacted

Accepted by.. ....
David Pesetsky
Head, Department of Linguistics and Philosophy

The Role of Perceptual Similarity and Gradient Phonotactic Well-Formedness in Loan Gemination Processes

by

Lilla Magyar

Submitted to the Department of Linguistics and Philosophy
on September 8, 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Loan gemination is a cross-linguistically widespread phenomenon: short consonants preceded by short stressed vowels in the source language are borrowed as long in loanwords. It is generally considered to be an 'unnecessary' adaptation (Peperkamp, 2005), because it does not repair any illegal sequences in the native phonotactics of the borrowing languages.

Hungarian is a particularly interesting case of a seemingly unnecessary adaptation: in the native phonology, both singletons and geminates can be found in word-final and intervocalic position (where loan gemination could potentially apply), therefore - on the face of it - there is nothing in the native phonotactics that would require gemination (Nádasdy, 1989).

In this thesis, I present a detailed case study of Hungarian loan gemination and argue that this process is heavily influenced by native phonotactics (i.e. geminate markedness which is also reflected in the distribution of geminates in the native phonology), perceptual similarity effects (faithfulness to source vowel duration), and orthography. I propose a MaxEnt model with weighted constraints which incorporates these factors and predicts the probability of productive loan gemination.

Thesis Supervisor: Adam Albright
Title: Professor

Acknowledgments

I don't think there is enough space to list all the people I am grateful to. First of all, I would like to say thank you to my committee (Adam Albright, Edward Flemming, Donca Steriade and Michael Kenstowicz) for the help, support and encouragement they gave me during this process. I am especially grateful to Adam Albright for being a great advisor and mentor during my years at MIT.

I also learned a lot from and had fun with my classmates (especially Ling-12 and especially Ben and Sudheer). I am thankful to Ling-15 for being such good students in my TA sessions. I greatly benefited from having discussions and classes with past and present faculty members including Noah Constant, Martin Hackl, Martin Walkow, Michel DeGraff, David Pesetsky, Sabine Iatridou, Irene Heim, Kai von Fintel and Roger Schwarzschild.

There are several people outside MIT who I want to say thank you to. Szilárd Szentgyörgyi was my first and most memorable linguistics instructor and mentor and this will never change - thank you for sparking my interest in linguistics.

I would never have reached this far if it hadn't been for the great Hungarian linguists at ELTE (TLP): Miklós Törkenczy, Péter Rebrus, László Kálmán and many others. Attending Ádám Nádasdy's classes was also an experience I will never forget. Not to mention the Phonology Camps.

I am also thankful to my linguist friends all over the world - there are so many that I can't even attempt to list them.

Last but not least, I'm grateful to my parents and Sudheer who have always supported me unconditionally. You are my inspiration.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Gemination in loanwords is a cross-linguistically widespread phenomenon, found in languages including Japanese (Kubozono et al., 2008; Kawahara, 2007, 2015), Hungarian (Nádasdy, 1989; Törkenczy, 1989; Kertész, 2006; Magyar, 2014, 2016), Italian (Passino, 2008; Hamann and Colombo, 2017), Finnish (Karvonen, 1998, 2005; Kiparsky, 2003; Kroll, 2014), Malayalam (Namboodiripad et al., 2014) and Telugu (Krishnamurti and Gwynn, 1985). It involves a short consonant preceded by a short (usually stressed) vowel in the source word undergoing lengthening in the loanword (e.g. English *fit* /fɪt/ → Hungarian *fitt* /fit:/). This process presents a puzzle for the following reasons:

1. The donor languages do not have geminates.
2. In the borrowing languages, both singletons and geminates are acceptable in the context where loan gemination applies (word-finally or intervocalically, following short (stressed) vowels). Furthermore, gemination does not repair illegal sequences in native phonologies, therefore it is widely regarded as an 'unnecessary adaptation' (Peperkamp, 2005).
3. Languages differ in the set of consonants they allow to geminate in loan-

> words, and it is not necessarily the set of geminates permitted in their native phonologies. There are languages which allow (almost) all consonants to be geminated in the native phonology but only a subset of them in the loan phonology (e.g. Hungarian, Finnish). Others allow a limited subset of geminates in the native phonology, and allow some other geminates in the loanword phonology which are not normally permitted in the native phonology (e.g. Japanese). Some languages do not impose much restriction on gemination in their native phonologies and allow all consonants to be geminated even in the loan phonology (e.g. Telugu, Italian).

The phenomenon in Hungarian is particularly interesting because even though both singletons and geminates are acceptable in the contexts where loan gemination applies and it has been claimed that there is nothing in the native phonotactics that would require gemination (Nádasdy, 1989), there is a clear geminate markedness hierarchy to be observed in loan gemination processes: voiceless affricates > voiceless stops > voiceless fricatives > voiced stops > nasals > liquids, voiced fricatives, and there is a high preference for geminate voiceless affricates and stops compared to other consonants. It also differs from similar processes in other languages: while in other languages the consonants which are final in source words end up being intervocalic in loanwords, in Hungarian they remain word-final, which is interesting because word-final geminates are cross-linguistically marked.

This study shows that preferences for geminates in loanwords line up with distributions of geminates in the native Hungarian phonology: although all singleton consonants have a geminate counterpart in the native phonology, different consonant classes have different distributions which line up with cross-linguistically observed geminate markedness hierarchies (Podesva, 2002; Steriade, 2004).

However, gemination hierarchies are more fine-grained in the native phonology and these facts are not sufficient to explain why there is a preference for geminate voiceless stops and affricates in loanwords. In a perception experiment, I show that one of the motivating factors for gemination in loanwords is faithfulness to source vowel duration. In English, vowels are significantly shorter before voiceless stops and affricates in closed syllables (Port, 1981; Van Santen, 1992) but this effect is much smaller in open syllables. This effect is weak and inconsistent in Hungarian (Olaszy, 2006). However, studies have shown that gemination does not have a significant effect on the preceding vowel in open syllables (Ham, 2001; Neuberger and Beke, 2017), but it significantly shortens the preceding vowel in before final consonants in monosyllables (an experiment conducted for Magyar (2014)). Therefore, gemination can be used as a strategy to shorten the preceding vowel and match source vowel duration.

Gradient patterns of loan gemination can be accounted for by the interaction of two main factors, apart from the influence of spelling: faithfulness to source vowel duration driven by perceptual similarity and gradient phonotactic well-formedness in the native phonology (most importantly, geminate markedness). To formalize this theory, I propose a Maximum Entropy model with weighted constraints, incorporating all the main factors influencing gemination in Hungarian loanwords and show that such a model is able to predict the probability of gemination in loanwords with high accuracy.

Chapter 2

Gemination in Hungarian loan adaptation

Hungarian loanwords borrowed from English, German and occasionally from French and Italian participate in loan gemination: a consonant which is pronounced as short in the borrowing language is borrowed as long. Consonants in three main contexts can undergo loan gemination: (1) word-finally in monosyllables, (2) intervocalicaly in disyllables and (3) in word-finally in polysyllables. Consonants following long vowels are not borrowed as geminates. Vowels preceding potentially geminated consonants are usually short and stressed in the source language. Source language orthography also plays an important role: if a consonant is spelled with a double letter in the source word, it will be borrowed as a phonetic geminate in the loanword. Since this phenomenon exhibits gradience and inter- and intra-speaker variation and has numerous exceptions, it can be considered more of a tendency than a rule.

In this chapter, I review some previous work, and by careful examination of loanword data and native speaker intuitions, I present the puzzles and questions raised by loan gemination and the possible hypotheses which can shed light on these issues.

2.1 Previous work

No work has presented a full analysis of gemination in English and German loanwords.[1] The most detailed presentation of loan gemination, especially in English, German and French words, is by Nádasdy (1989). He points out - by showing examples and presenting a pre-theoretical proposal - that loan gemination largely depends on position in the word, stress in the source word, and consonant class. His claim is that German words were originally borrowed from a specific variety of German, South-Eastern Urban German (henceforth SEUG), which was spoken by middle class people in Austria-Hungary and was based on the Bavarian-Austrian dialect. In this variety, consonants were pronounced with 'sharp'[2] articulation, which might have been perceived as long compared to short Hungarian consonants by Hungarian listeners. Nádasdy hypothesizes that this perceptual effect triggered gemination of consonants in loanwords borrowed from German. With this in mind, he proposes a [+German] feature for words which tend to participate in gemination and [-German] to those which do not, regardless of what the origin of a particular word is.

Törkenczy (1989), similarly to Nádasdy (1989) points out that words which are considered to be foreign have the Heavy Syllable Requirement, and generally, words with branching onsets are considered to be foreign, which might be the reason why they tend to be borrowed with final geminates.

[1]Huszthy (2002) presents an analysis of the asymmetries of gemination in Italian loanwords. The motivation and the analysis of gemination in Italian loanwords is slightly different from that of English and German loanwords, therefore it will not be explored in detail in the present work.

[2]It is unclear what exactly Nádasdy means by this term.

Kertész (2006) discusses various loanword adaptation phenomena in different languages, including gemination in loanwords borrowed into Hungarian. Building on Nádasdy (1989) and Törkenczy (1989), Kertész (2006) stressed the importance of the Heavy Syllable Requirement: that is, if a syllable in a foreign word contains a short (stressed) vowel followed by a short consonant, it has to become heavy, which is usually achieved by gemination of the final consonant. She cites two different motivations for this process to happen from personal communications: (1) English voiceless obstruents are often claimed to behave like geminates (Péter Szigetvári, p.c.), and (2) loan gemination is the result of analogical extension (László Kálmán, p.c.) of a process that used to be influenced by perceptual similarity (SEUG consonants perceived as long, claimed by Nádasdy (1989)).

2.2 Loanword corpus

There is a relatively small number of loanwords potentially undergoing gemination in loanwords, and not all of them have a consistent spelling in Hungarian. If the Hungarianized word is spelled with a double letter, it is generally the indication of the presence of a phonetic geminate. If it is spelled with a single letter (in the absence of standardized spelling), there is still a possibility of pronouncing that consonant as a geminate.

I constructed a list of 122 loanwords (which is found in the Appendix A). The data were compiled from words listed in Nádasdy (1989), online dictionary and corpora searches (Hungarian Webcorpus (Halácsy et al., 2004) and Google search. As I do not know of any specific loanword corpora, the list of words is most probably not exhaustive. It contains words with consonants which undergo or could potentially undergo gemination, because they contain consonants in one of the following positions: (1) word-finally in monosyllables, (2) intervocalically in disyllables and (3) word-finally in polysyllables. The data are labeled with the following variables: meaning, word type (monosyllable, disyllable, polysyllable), source word spelling (single letter, digraph, trigraph), consonant position (final or intervocalic), consonant class (voiceless stop / affricate / fricative, voiced stop / fricative, nasal, liquid), preceding vowel stress (stressed or unstressed), and gemination (yes, optional (preferred), optional (dispreferred), no). 'Optional' means that both geminate and singleton forms are equally acceptable - there is no specific preference for either form. 'Optional, preferred' means that both forms are acceptable, but the geminate form is more standard. 'Optional, dispreferred' means that there is a possibility of gemination, but most speakers use the singleton form (that is the standard form) and the geminate form is rarely used or even stigmatized. These labels are based on native speaker intuitions and and not corpus counts: they were included to provide information on which forms are considered to be the standard variety.

2.2.1 Word-final gemination monosyllables

Word-final position in monosyllables is the most common environment for loan gemination.

There are 29 monosyllables in the corpus which end in voiceless stops. Consonants which are spelled with a single letter can be obligatorily pronounced as a geminate (n=6), optionally geminated (gemination preferred: n=1, dispreferred: 2), optionally geminated (without specific preference for either geminate or singleton form: n=6), or optionally geminated (gemination is dispreferred: n=2). Voiceless stops spelled with a double letter are always geminated (n=6), and those spelled with a digraph can be obligatorily (n=6) or optionally (n=1) geminated.

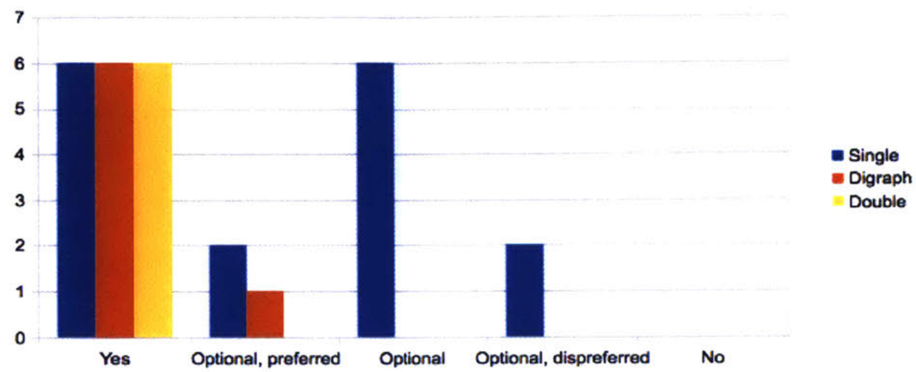

Figure 2-1: Word-final gemination in monosyllabic loanwords: voiceless stops

There are 13 words in the corpus which end in voiceless affricates. They are either spelled with a digraph (n=8) or a trigraph (n=5), and are in both cases geminated.

Figure 2-2: Word-final gemination in monosyllabic loanwords: voiceless affricates

There are 10 monosyllables in the corpus ending in voiceless fricatives. Those spelled with single letters in the source words can be optionally geminated (without specific preference: n=1 or with preference for gemination: n=1). Those with a digraph (n=3), trigraph (n=2) and a double letter (n=3) spelling are obligatorily geminated.

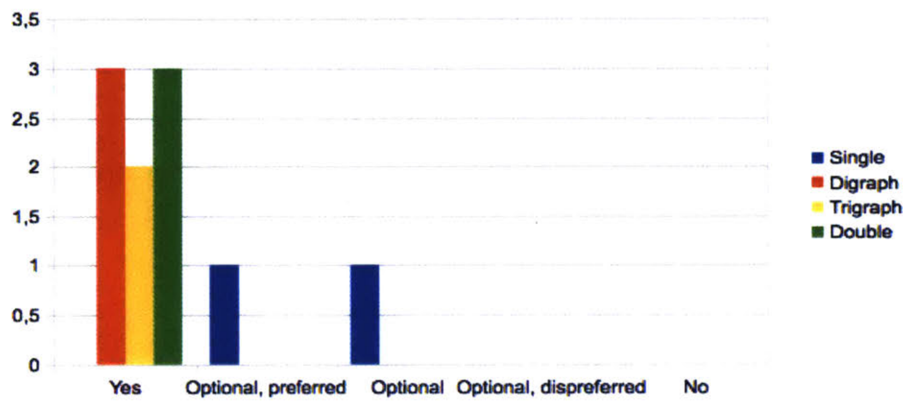

Figure 2-3: Word-final gemination in monosyllabic loanwords: voiceless fricatives

14 words in the corpus end in voiced stops and all of them are spelled with a single letter in the source word. There is an option of gemination, but it is dispreferred (n=10). There is one item in which gemination is preferred and another one in which gemination is never an option. There are no words ending in voiced stops which are obligatorily borrowed with geminates or for which there is an equal preference for the geminated and the singleton form.

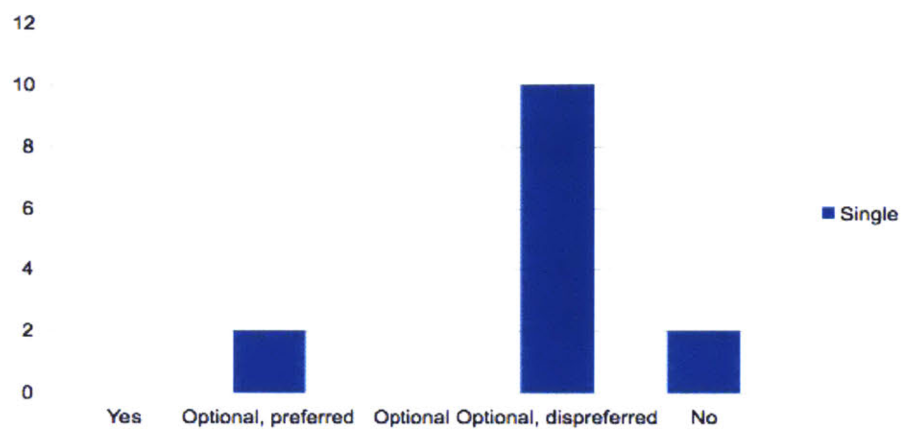

Figure 2-4: Word-final gemination in monosyllabic loanwords: voiced stops

There are only five examples of well-known loanwords ending in nasals. Those spelled with a single letter in the source word can have a geminate (n=1), an optional, preferred geminate (n=1) or a singleton (n=1) pronunciation in the loanword. Those spelled with a double letter are optionally geminated, but there is a preference for gemination (n=2).

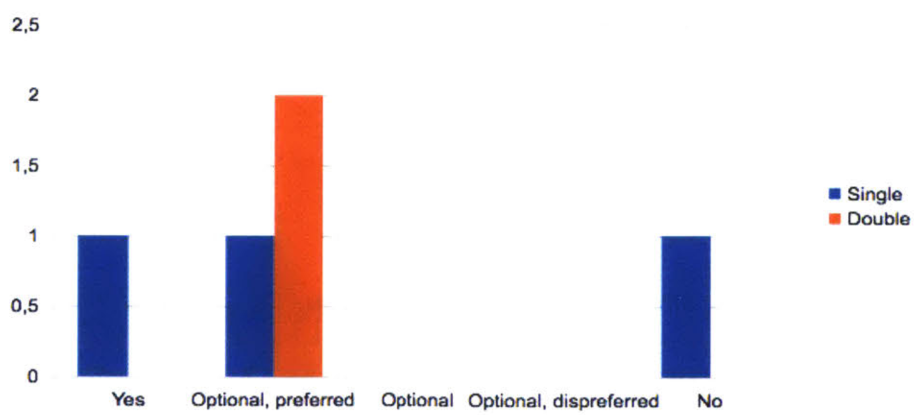

Figure 2-5: Word-final gemination in monosyllabic loanwords: nasals

All well-known monosyllabic loanwords ending in liquids are spelled with a double letter in the source word and all of them are geminated in the loanword.

Figure 2-6: Word-final gemination in monosyllabic loanwords: liquids

There is only one example of an English / German loanword ending in a voiced fricative: it is *dzsessz* 'jazz'. Its traditional pronunciation is with a long voiceless alveolar fricative instead of a voiced one, but younger speakers do pronounced it with a long (or sometimes with a shortened) voiced fricative. It is not clear whether this word was borrowed directly from English or was borrowed through German: in the latter case, the devoicing can be attributed to the German pronunciation of the final consonant. Otherwise, it is due to the markedness of voiced fricatives both in Hungarian and in other languages.

2.2.2 Intervocalic gemination in disyllables

There is a relatively small number of loanwords with potential intervocalic gemination. There is an extremely small number of intervocalic consonants spelled with a single letter in English and German, as in these languages short vowels are usually followed by double letters or digraphs. The ones with single letter spellings are borrowed with a short vowel, with the exception of *dopping* 'doping'.[3] Double letters tend to be pronounced as geminates in the Hungarianized form. Most loanwords consisting of a monosyllabic root and the suffix *-er* are pronounced long, but many of those ending in *-y* do not (as *-i* is also a diminutive suffix in Hungarian and is never preceded by geminates). In general, intervocalic consonants are only geminated when they are spelled with a double letter or consist of a monosyllabic root + a foreign suffix, usually *-er*.

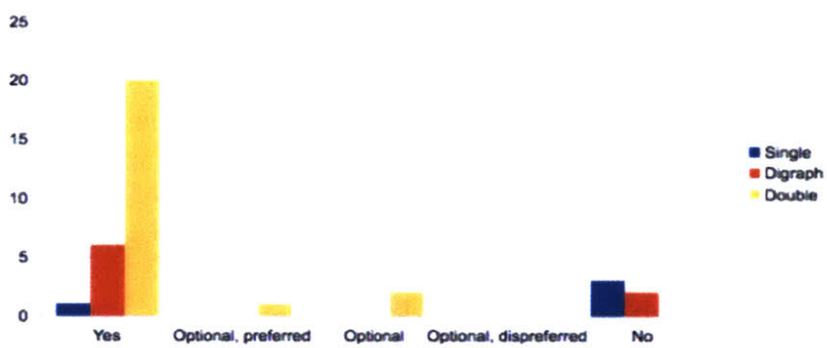

Figure 2-7: Intervocalic gemination in disyllabic loanwords

2.2.3 Word-final gemination in polysyllables

There are a few words which are not monosyllables but have final geminates. These words were mostly borrowed from French or German, and had a double letter spelling or a final stressed syllable in the source language (e.g *piruett* 'pirouette'. Exceptions are *hotel* and *panel* which were borrowed through German and they

[3]Another notable exception is *mafia*. However, it is an Italian loanword and is presumably pronounced with a long consonant to mimic post-tonic lengthening in Italian. Italian words have different gemination processes, which will not be discussed in much detail in this thesis.

have final stress in the source word - the reason for this could be the lack of gemination of liquids without orthographic reflex.

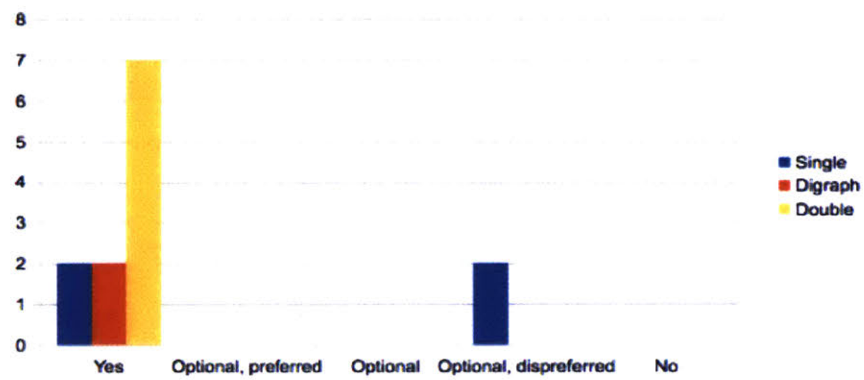

Figure 2-8: Word-final gemination in disyllabic loanwords

2.3 Productive loan gemination: native speaker intuition elicitation task

Loan gemination is most common and productive word-finally in monosyllables. Based on the examples shown above, voiceless obstruents are more likely to be geminated without the influence of orthography than voiced consonants, especially voiceless stops and affricates. There is some variation in the gemination of voiced stops and nasals without the influence of orthography, but it is strongly dispreferred for voiced stops and even more so in the case of nasals.

Due to the limited amount of loanword data existing in the Hungarian lexicon, it is hard to find out which consonants are more likely to be geminated than others. It can be the case that some consonants are often geminated in already existing, well-known loanwords, but when new words are being borrowed, people will prefer singleton forms.

Furthermore, it is hard to assess the frequency of singleton and geminate forms, as there is considerable inter- and intra-speaker variation and preferences for gemination are also gradient (e.g. labels like 'optional', 'dispreferred', etc.). Google and corpus search does not help much because for many loanwords, spelling is not standardized: many words are spelled with a single letter but are often pronounced as a geminate. Spelling mistakes / errors can sometimes be revealing, but they do not necessarily correspond to how people actually pronounce those words.

Therefore, it is important to see what native speaker's intuitions are on real or hypothetical English loanwords.

2.3.1 Participants

Five native speakers of Hungarian participated in the elicitation task, two of whom had a background in linguistics and the other three have also learned about speech sounds in school. All of them had some familiarity with English.

2.3.2 Items

Elicitation was conducted in different ways: in person or in email. Participants were provided with a list of items. Each word was listed in a new line, but the order of items was as follows: items ending in [t], [k], [p], [tʃ], [ts], [f], [ʃ], [s], [d], [g], [b], [n], [m], [l], [z]. The oder of presentation was voiceless stops, voiceless affricates, voiceless fricatives, voiced stops, nasals, liquids and voiced fricatives, but the consonant class labels were not specified. For each consonant, examples were included for the vowels [æ], [ɛ], [ɪ], and [ʌ], for single and double letter spelling and for complex onset. The exact order of presentation is given below.

- Voiceless stops:
  - *bat, bet, bit, but, batt, bett, bitt, butt, blat, blet, blit, blut*

  - *bak, bek, bik, buk, back, beck, bick, buck, black, bleck, blick, bluck*
  - *bap, bep, bip, bup, bapp, bepp, bipp, bupp, blap, blep, blip, blup*

- Voiceless affricates:
  - *batch, betch, bitch, butch, blatch, bletch, blitch, blutch*
  - *bats, bets, bits, buts, blats, blats, blits, bluts*

- Voiceless fricatives:
  - *baf, bef, bif, buf, baff, beff, biff, buff, blaf, blef, blif, bluf*
  - *bash, besh, bish, bush, blash, blesh, blish, blush*
  - *bas, bes, bis, bus, bass, bess, biss, buss, blas, bles, blis, blus*

- Voiced stops:
  - *bad, bed, bid, bud, badd, bedd, bidd, budd, blad, bled, blid, blud*
  - *bag, beg, big, bug, bagg, begg, bigg, bugg, blag, bleg, blig, blug*
  - *bab, beb, bib, bub, babb, bebb, bibb, bubb, blab, bleb, blib, blub*

- Nasals:
  - *ban, ben, bin, bun, bann, benn, binn, bunn, blan, blen, blin, blun*
  - *bam, bem, bim, bum, bamm, bemm, bimm, bumm, blann, blenn, blinn, blunn*

- Liquids:
  - *bal, bel, bil, bul, ball, bell, bill, bull*

- Voiced fricatives:
  - *baz, bez, biz, buz, bazz, bezz, bizz, buzz, blaz, blez, bliz, bluz*

– *bav, bev, biv, buv, blav, blev, bliv, bluv*

Only [l] was included among liquids, as vowels preceding [r] are never short in loanwords, therefore gemination is not triggered. Among voiced fricatives, [ʒ] was not included for the same reasons. Double letter spelling of [v] was not included, either, since it is unattested both in English and Hungarian in word-final position.

2.3.3 Methods

Participants were asked to look at the items as if those were English loanwords and decide if they preferred to borrow them with a long or a short consonant.

2.3.4 Results

Participants gave three kinds of responses: either a form with a long consonant, a short consonant or both. Vowels were short in all items. They indicated their responses regarding consonant length by writing 'long', 'short' or 'long/short'. The percentages were calculated from their answers by counting responses listing both as 0.5. Results are shown in the following table for words which did not have a double letter spelling: words with double letters were borrowed with geminates by all participants. Only one participant borrowed words with complex onsets consistently with final geminates.

	Geminate (%)	N
Voiceless affricates	83%	72
Voiceless stops	62%	120
Voiceless fricatives	37%	96
Voiced stops	20%	120
Nasals	12%	64
Liquids	0%	20
Voiced fricatives	0%	64

Table 2.1: Results for words without double letter spelling

Before fitting a mixed effects logistic regression, the variable 'consonant class' was difference coded because it is important to see whether there is a significant difference in geminate and singleton responses between different consonant classes. Consonant classes (as variables) were ordered according to the probability of getting a geminate response for a comparison between 'adjacent' consonant classes. This order was established without the items containing orthographic geminates. Double (short/long) responses were listed as separate line in the data file.

- Voiced Fricatives - Liquids
- Nasals - Voiced Fricatives
- Voiced Stops - Nasals
- Voiceless Fricatives - Voiced Stops
- Voiceless Stops - Voiceless Fricatives
- Voiceless Stops - Voiceless Affricates

The variable 'OrthGem' (orthographic geminate with a binary response: 'yes' or 'no') was treatment-coded, and 'no' was set as the baseline.

A mixed effects logistic regression was fitted in R with the following formula: glmer(Geminate ~ C.Class+OrthGem+(1|Subject), family = binomial(link = "logit"), data=loantest, glmerControl(optimizer="bobyqa", optCtrl = list(maxfun = 100000))).

There are significant main effects for both C.Class (consonant class) and OrthGem (orthographic geminate):

C.Class 6 1084.92 297.53 <2.2e-16 ***

OrthGem 1 1138.39 341.00 <2.2e-16 ***

The pairwise comparisons show significant effects between many of the 'adjacent' pairs.

	Estimate	Std. Error	z value	Pr
(Intercept)	-5.7049	27.7620	-0.205	0.83719
VoicedFricatives-Liquids	-0.5366	48.5790	-0.011	0.99119
Nasals-VoicedFricatives	17.4707	95.8128	0.182	0.85531
VoicedStops-Nasals	1.0229	0.4272	2.395	0.01664 *
VoicelessFricatives-VoicedStops	1.1585	0.2524	4.589	4.44e-06 ***
VoicelessStops-VoicelessFricatives	0.6007	0.2133	2.816	0.00486 **
VoicelessStops-VoicelessAffricates	1.9041	0.4043	4.709	2.49e-06 ***
OrthGemyes	38.4631	45.3991	0.847	0.39687

Table 2.2: Fixed effects

There is a significant difference between voiceless affricates and voiceless stops, voiceless stops and voiceless fricatives, voiceless fricatives and voiced stops, and voiced stops and nasals. There are no significant differences between the rest of the consonant classes with respect to geminate and singleton responses.

The results show that, in native speaker judgements of loanwords, voiceless affricates geminate the most, followed by voiceless stops. Voiceless stops geminate more often than voiceless fricatives. Voiceless fricatives geminate more than voiced stops, and voiced stops geminate more than nasals. There is no significant ranking order between the rest of the consonant classes: nasals, liquids and voiced fricatives.

2.4 Questions and hypotheses

Patterns of loan gemination present a puzzle for different reasons. Although in principle all singleton consonants have a geminate counterpart in the native Hungarian phonology, only a subset of these consonants can be geminated in loanwords, namely voiceless obstruents and to a very small extent voiced stops and nasals. Voiceless obstruents are cross-linguistically less marked as geminates than voiced and more sonorous consonants. This raises the question whether this reflects universal markedness hierarchies or if we look closer into the distribution of geminates and singletons in the native phonology, we will find similar patterns.

The fact that loan gemination is truly productive only word-finally in monosyllables raises further questions about gradient phonotactic well-formedness in the native phonology, as geminates and singletons are thought to be equally well-formed and well-attested in all of the positions in which consonants can potentially undergo gemination in loanwords.

Furthermore, loan adaptation processes - especially the so-called 'unnecessary' adaptations (which loan gemination has been claimed to be, cf. Peperkamp (2005)) - generally impose some sort of faithfulness requirement on the loan word to the source word form, which is often influenced by perceptual similarity. Productive loan gemination is preferred in the case of two consonant classes: voiceless stops and affricates. In English, these are the consonants which trigger shortening of the preceding vowel, and this effect is much stronger in closed syllables (Port, 1981). In Hungarian, gemination of the same consonants has been shown to have a similar effect: a significant effect of shortening before geminate voiceless stops and affricates has been shown in monosyllables in word-final position (Magyar, 2014), but there is no such effect on the preceding vowel in intervocalic position (Ham, 2001; Neuberger and Beke, 2017). This suggests that gemination could be used as a strategy to retain the shortness of the source vowel in this particular context, and could explain why gemination is more likely to affect certain consonants and certain positions in words.

Based on the facts, two hypotheses can be proposed and tested:

- Hypothesis I: Loan gemination is shaped by gradient phonotactic well-formedness and geminate markedness in the native Hungarian phonology.
- Hypothesis II: Gemination is a strategy to preserve the shortness of the source vowel.

Chapter 3

The gradient phonotactic well-formedness effect

In a pseudo loanword adaptation task described in Chapter 3, native speakers of Hungarian were asked whether they would borrow monosyllabic loanwords with final consonants as geminates or singletons. Results for geminates (excluding words with a double letter spelling) based on their responses are as follows: voiceless affricates (83%) > voiceless stops (62%) > voiceless fricatives (37%) > voiced stops (20%) > nasals (12%) > liquids, voiced fricatives (0%). There was no significant difference between nasals and liquids and between liquids and voiced fricatives.

The fact that voiceless obstruents are more likely to be geminated than voiced obstruents, and voiced obstruents are more likely to be geminated than sonorants (except for voiced fricatives, which are not geminated) suggests that geminate markedness potentially has an effect on loan gemination processes. This also raises the question whether this is the effect of universal markedness but nothing specific to the native Hungarian phonology (in which practically all consonants can be geminated), or whether these are patterns to be observed in the native phonology / lexical frequencies as well.

In this chapter, I provide an overview of gemination in the native phonology and results of a corpus study of the distribution of geminates and singletons. Then I show how a grammar with weighted constraints can account for these distributions. I also report the results of a well-formedness task on novel words and show how the grammar of the native phonology alone is not sufficient to account for native speaker judgements on novel forms.

3.1 Gemination in the native phonology

There are 25 consonants in the native phonology of Hungarian. All short consonants have a long counterpart, but not all of these consonants participate in loan gemination processes. Palatal stops and palatal nasals do not occur in English and German loanwords. Palatal glides do occur in some German names (e.g. *Bayer*), but there is an option of geminating the consonant or lengthening the previous vowel, and there is not enough data to make generalizations about the behavior of /j/ in loan adaptation processes (*contra* Nádasdy (1989)), who claims that /j/ is always geminated in loanwords). Voiced affricates very rarely occur in loanwords in contexts where they could potentially undergo gemination, but are always geminated in intervocalic and word-final contexts in the native Hungarian phonology.

Singleton consonants can occur in word-initial, intervocalic, and word-final position, can be flanked by other consonants and be part of complex onsets and codas. Geminates can be found in intervocalic and word-final position, but are degeminated when flanked by other consonants, and cannot co-occur with other consonants in complex onsets and codas (Siptár and Törkenczy, 2000). Both long and short consonants can occur in word-final and intervocalic position, following short stressed vowels (that is, in the context where loan gemination takes place).

Both short and long consonants can occur after long vowels as well, but geminates are extremely rare in this context, and even then, either degemination or shortening of the long vowel takes place.

Currently, there are no studies available which show the exact distribution of geminates in different contexts. The goal of this chapter is to provide a detailed overview of the distribution of geminates in the two main contexts where loan gemination potentially occurs (that is, word-finally (and very rarely in intervocalic position), following short stressed vowels), and to show the types and rates of gemination in different contexts, and at different stages of Hungarian (before and after English and German loanwords entered the language). The data were collected using regular expressions and python scripts from corpora and word lists (Hungarian Webcorpus (Halácsy et al., 2004) and a reverse alphabetized dictionary (Papp, 1969). The corpus study shows that some of the patterns observed in loan gemination processes had already been visible in the native phonology before English and German loanwords were borrowed into Hungarian, and that these patterns line up with cross-linguistic implicational hierarchies of gemination to a certain extent.

3.1.1 Word-final gemination in monosyllables

Word-final position, following short vowels in monosyllables is the most common context for loan gemination: if a monosyllabic word is borrowed from English which contains a singleton consonant following a short stressed vowel in the source word, there is a high chance that the consonant will be geminated in the loanword. This used to be a very productive process in the adaptation of German as well as less recent English loanwords. Therefore, it is important to see whether gemination was frequent in this context even before these loanwords were borrowed into the language.

The following figure shows the distribution of geminates for each consonant class in two data sets: native words (without German and English borrowings) and all current words (well-integrated loanwords which are known to all native speakers if Hungarian are included). Consonant classes are represented as follows: T: (geminate voiceless stops), T:S (geminate voiceless affricates), S: (geminate voiceless fricatives), D: (geminate voiced stops), N: (geminate nasals), L: (geminate liquids), and Z: (geminate voiced fricatives). The blue line represents the native data set and the orange line is the current data set with well-known loanwords.

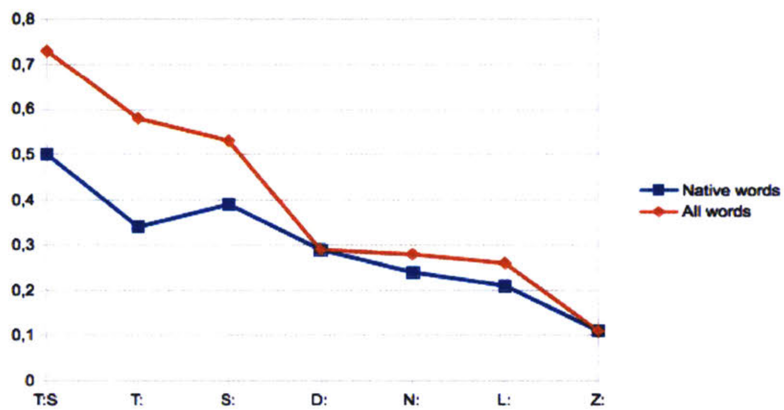

Figure 3-1: Distribution of word-final geminates after short vowels in monosyllables

What we can observe is that the rate of gemination was considerably lower in the native phonology before German and English loanwords began to enter the language, but even then, voiceless obstruents were more frequently geminated than voiced consonants, and voiced fricatives were the least likely to be geminated. This trend continues after the adaptation of German and English loanwords, but the number of geminate voiceless affricates, stops and fricatives drastically increases. There is a slight increase in the number of geminate nasals and liquids, but the geminate / singleton ratio stays the same for voiced stops and fricatives. The distribution of geminates in the earlier native phonology was as follows: voiceless affricates (50% (n=24)) > voiceless fricatives (39% (n=39)) > voiceless stops (34% (n=65)) > voiced stops (29% (n=55)) > nasals (24% (n=33)) > liquids (21% (n=57)) > voiced fricatives (11% (n=13)), while in the current native phonology, it is as follows: voiceless affricates (73%, n=45) > voiceless stops (59% (n=114) > voiceless fricatives (53% (n=77) > voiced stops (29% (n=66)) > nasals (27% (n=40)) > liquids (26% (n=62) > voiced fricatives (11% (n=13)). In both stages of the language, the order of the distributions of geminates is similar, which is indicated by a high rank order correlation coefficient (Kendall's Tau = 0.9).

Geminates after long vowels are extremely rare in this context, as illustrated by the following figure. The yellow line shows the distribution of word-final geminates for various consonant classes after long vowels in monosyllables (the distribution is the same in native words and the current lexicon), compared to the distribution of geminates after short vowels in the same context in native words (blue line) and the current lexicon (orange line). The distribution of geminates following short vowels at the end of monosyllabic words is as follows: liquids (3% (n=76)) > voiceless stops (2% (n=55)) > voiceless affricates (0% (n=12)), voiced stops (0% (n=41)), nasals (0% (n=36)), liquids (0% (n=76)), and voiced fricatives (0% (n=39)).

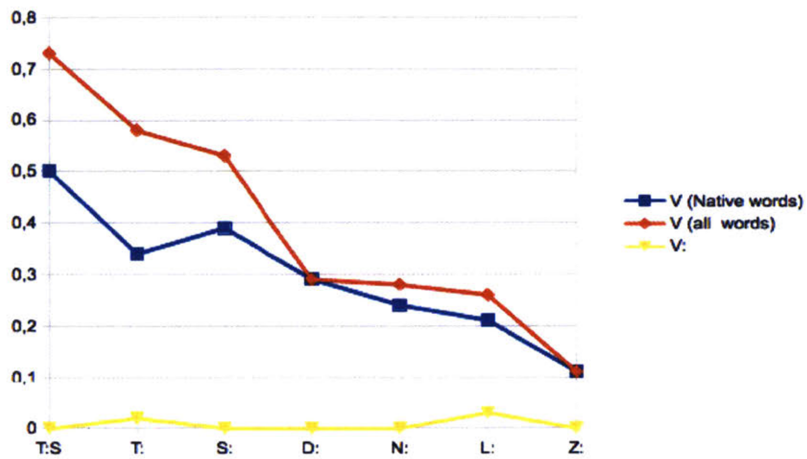

Figure 3-2: Distribution of word-final geminates after long vowels in monosyllables

3.1.2 Intervocalic gemination in disyllables

In loanwords, gemination is much less common in intervocalic position following short stressed vowels than in word-final position, but this is another context where gemination does happen occasionally, however, only when the consonant in the source word was spelled with a double, or the word consists of a monosyllabic root and a (foreign) suffix.

The following figure shows the distribution of geminates in disyllables in intervocalic position, following short stressed vowels. As previously, consonant classes are represented as follows: T: (geminate voiceless stops), T:S (geminate voiceless affricates), S: (geminate voiceless fricatives), D: (geminate voiced stops), N: (geminate nasals), L: (geminate liquids), and Z: (geminate voiced fricatives).

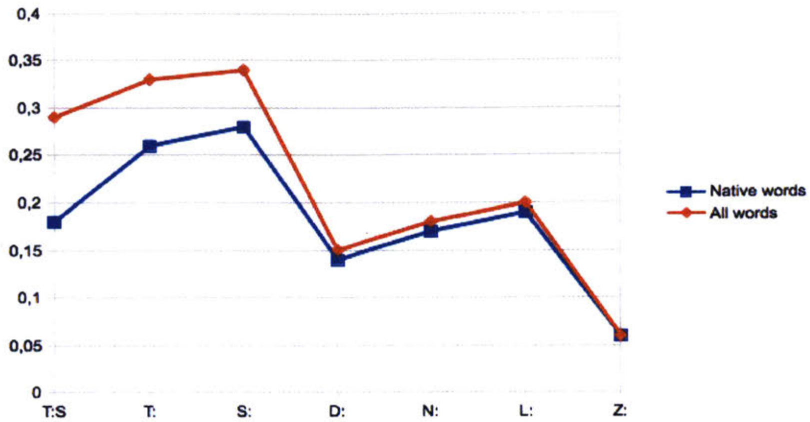

Figure 3-3: Distribution of intervocalic geminates in disyllables in the Hungarian lexicon

As is clearly seen, the distribution of geminates does not follow the consistent, hierarchical patterns that can be observed in the case of monosyllables. Even here, voiceless consonants are more frequently geminated than voiced consonants, but in general, gemination is much less frequent than in the case of monosyllables. The overall pattern does not change with the adaptation of German and English words, but there is an increase in the gemination of voiceless affricates, stops and fricatives. The order of the distribution of geminates in the native lexicon (without English and German loanwords) is as follows: voiceless fricatives (29% (n=185)) > voiceless stops (26% (n=433)) > liquids (19% (n=453)) > voiceless affricates (18% (n=134)) > voiced stops (14% (n=305)) > voiced fricatives (0% (n=172)). The order is similar in the current lexicon (including some well-integrated loanwords):

Similarly to word-final position, gemination after long vowels is rare in intervocalic contexts, as shown by the following table. As previously, gemination after long vowels both in the native and the current lexicon (yellow line) is shown in comparison to gemination after short vowels in native words (blue line) and the current lexicon including well-integrated German and English loanwords (orange line).

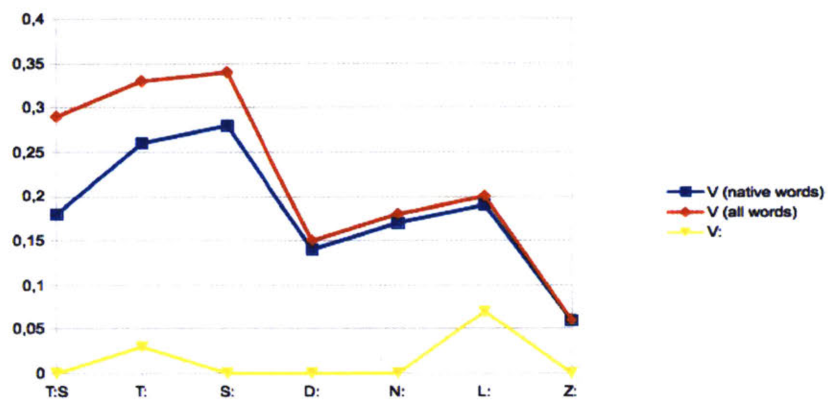

Figure 3-4: Distribution of intervocalic geminates in disyllables in the Hungarian lexicon

There is a possible explanation for the lack of clear patterns of the geminate markedness hierarchy in intervocalic position (as opposed to the clear patterns word-finally in monosyllables). The monosyllabic data contains all possible monosyllables, including morphologically complex forms which are generally suppletive or are not treated as morphologically complex by native speakers of Hungarian, as it is not easy to separate the roots from suffixes in monosyllables. At the same time, the data set containing intervocalic geminates cannot be an exhaustive list of words or forms with intervocalic geminates: most of intervocalic gemination is the result of productive morphophonological processes and it is impossible to list all forms with inflectional and derivational suffixes. Therefore, the present list is only a small subset of the possible geminates found in intervocalic position.

3.1.3 Outline of a grammar

Based on the distribution of geminates in the native phonology (before German and English loanwords were part of the vocabulary), we can sketch a grammar. Gemination is more common in monosyllables word-finally than in polysyllables intervocalically, which suggests that there should be a bimoraic word minimality requirement. This constraint, of course, is violated fairly often, as singletons are generally preferred over geminates in the native phonology.

- BIMORAICMINIMALITY: Words must have the length of at least two moras.

The following markedness constraint express the generalization that geminates tend not to occur after long vowels, and that certain consonant classes cannot occur as geminates.

- *V:CC: Geminates cannot occur after long vowels.
- *ZZ: Geminate voiced fricatives are not permitted.
- *LL: Geminate liquids are not permitted.
- *NN: Geminate nasals are not permitted.
- *DD: Geminate voiced stops are not permitted.
- *SS: Geminate voiceless fricatives are not permitted.
- *TT: Geminate voiceless stops are not permitted.
- *TTS: Geminate voiceless affricates are not permitted.

As the categorical ranking of these constraints could only predict distributions like 0-100% (with fixed rankings) and 0-50% (with partially ranked / unranked constraints), constraints have to be weighted to account for fine-grained distributions. The MaxEnt Grammar Tool (Hayes and Wilson, 2008) was used, which weights constraints based on the observed probabilities of potential surface forms and the number of violations of each constraint by each candidate. The following weights were assigned to the constraints:

Constraint	Weight
*ZZ	3.5
*V:CC	2.7
*NN	2.5
*DD	2.4
*LL	2.2
*TT	1.9
*SS, *tts	1.8
BIMORAICMINIMALITY	1.4

Table 3.1: Constraints with weights

As we can see, geminate markedness constraints and *V:CC received a greater weight than BIMORAICMINIMALITY. Since geminates are generally more marked than singletons, the bimoraic word minimality requirement is often violated by actual surface forms.

The two figures below show how this grammar fits the distribution of word-final and intervocalic geminates for each consonant class after short and long vowels. The blue line indicates the distribution of geminates based on corpus frequencies and the orange line shows the predictions of the grammar. There is a close but imperfect fit between actual distributions and gemination predicted by the grammar. The reason for this is that the distribution of geminates is inconsistent between word-final and intervocalic positions for certain consonant classes.

For example, voiceless affricates are frequently geminated in word-final position in monosyllables but not in intervocalic position, while liquids are generally not favored as geminates but do geminate more frequently after long vowels than other consonants.

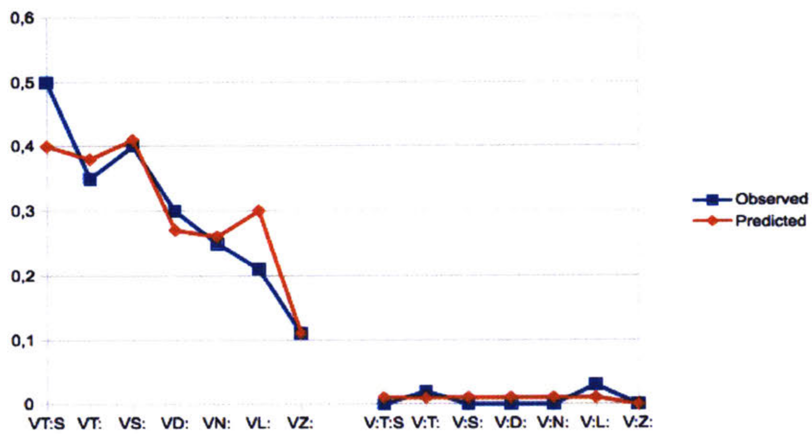

Figure 3-5: Word-final geminate frequencies observed in the corpus and predicted by the grammar

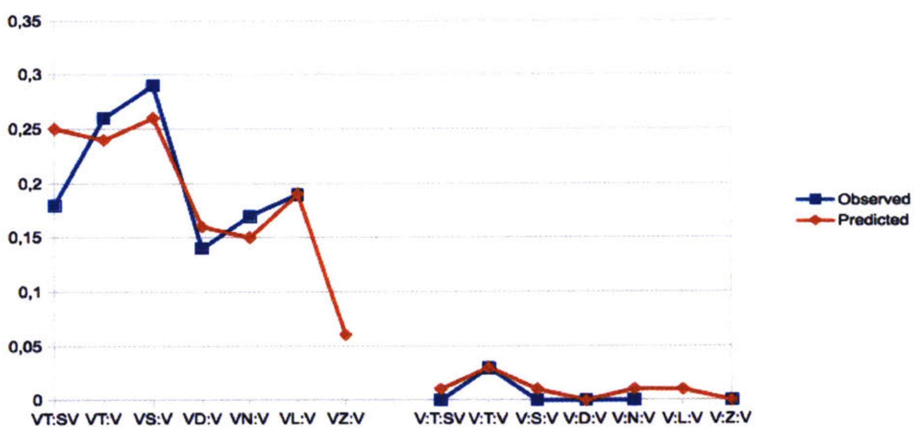

Figure 3-6: Intervocalic geminate frequencies observed in the corpus and predicted by the grammar

Predictions of the grammar were compared to the native Hungarian data set without English and German loanwords, because the grammar was constructed without other factors, for example, perceptual similarity or similarity to other forms. The other data set containing well-integrated English and German loanwords was shaped by other factors as well (including perceptual effects, faithfulness to source words and similarity to existing words) and will be the basis for well-formedness judgements on novel forms as well as new loanwords.

3.2 Well-formedness of novel words

Loan gemination is most common and productive in word-final position following short stressed vowels in monosyllables. If the consonant is spelled with a double letter or a digraph, it is most likely to be borrowed as a geminate. Certain consonants also tend to be adopted as geminates even when they are spelled with a single letter in the source word. Another well-known context for loan gemination is following short stressed vowels in intervocalic position. However, in this context, consonants only geminate when there is a double letter spelling in the source word or the loanword is complex, consisting of a monosyllabic root and a suffix.

In this section, I present the results of two nonce word well-formedness judgement tasks. One of the experiments is testing the well-formedness of monsyllables ending in short vowel + singleton / geminate sequences. The other one is testing disyllabic words containing short vowel + singleton / geminate + short vowel sequences. The results show that gemination patterns are more consistent with geminate / singleton distributions in the native phonology in the case of monosyllables than for disyllables.

3.2.1 Experiment I: Word-final gemination following short vowels in monosyllables

There is a relatively small number monosyllables (containing short vowels) in the native Hungarian lexicon, and the distribution of singletons and geminates show considerable variation across individual consonants and consonant classes. Geminate voiceless stops and affricates are almost as well-attested as their singleton forms, voiceless fricatives are less common as geminates, but still do occur more often than geminate nasals and liquids. Voiced stops are rare and voiced fricatives are unattested in word-final position in unsuffixed words. However, certain voiced stops (/b/ and /d/) and the voiced fricative [z] occur at the end of monosyllabic but non-monomorphemic words (e.g. *vedd* /vɛd:/ - imperative form of *vesz* 'take', *jobb* /job:/ - comparative form of *jó* 'good', *hozz* /hoz:/ - imperative form of 'bring', etc.).

The goal of this well-formedness judgement task testing the acceptability of nonce monosyllables (briefly described in Magyar (2014) and Magyar (2016)) was to investigate whether native speaker judgements reflect the patterns seen in the corpus of native Hungarian words and well-integrated older loans, and to provide a larger number of data (since the native Hungarian data for this context are sparse and have many accidental gaps).

The test contained 144 target items (72 word pairs, see Appendix B): combinations of short vowels and short consonants or geminates. All words ended in a short vowel + short consonant or geminate sequence. 115 native speakers of Hungarian participated in the experiment, who were recruited on Facebook and various university mailing lists. They volunteered for the experiment and were not paid. All of them were living in Hungary at the time of the experiment.

The task was administered online and participants remained anonymous. Participants were presented with a word pair and asked to decide which member of the word pair sounded more plausible as a Hungarian word or a Hungarianized loanword. All word pairs were minimal pairs, containing a monosyllable ending in a short vowel + singleton sequence and another one with the same vowel and the same consonant in a geminated form (e.g. *mok* /mok/ - *mokk* /mok:/). The order of presentation was balanced, that is, sometimes the word with the singleton came first, sometimes the one with the geminate. Although the stimuli were written, not spoken, participants were asked to treat spelling as a strict representation of pronunciation.

Results

Well-formedness judgement results for word-final gemination are shown in comparison to corpus data in Figure 3-7. The blue line indicates the distribution of geminates in the native Hungarian dataset (without English and German loanwords). The orange line represents gemination in all current words in the Hungarian language (including well-integrated, widely used loanwords). The yellow lines shows gemination in well-formedness judgements: voiceless affricates (65% (n=664) > voiceless stops (51% (n=913) > voiceless fricatives (46% (n=996)) > liquids (41% (n=664)) > nasals (29% (n=664)) > voiced stops (28% (n=996) > voiced fricatives (15% (n=996)).

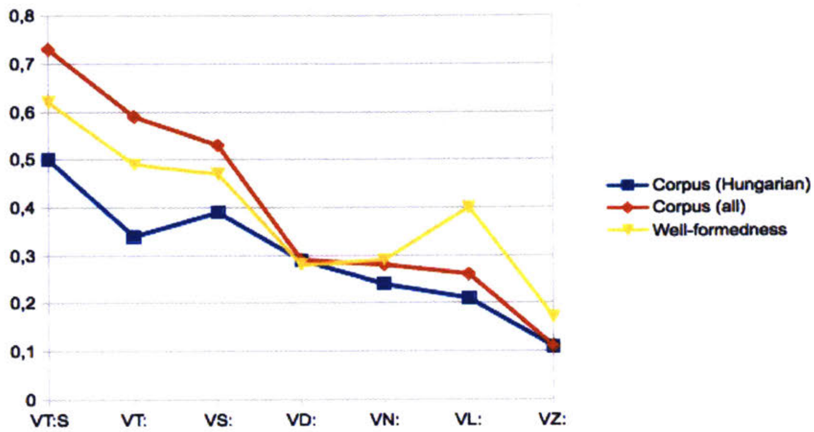

Figure 3-7: Word-final geminate frequencies

The distribution of geminates for different consonant classes shows similar tendencies in all three data sets, but the rate of gemination for voiceless obstruents is much higher in the data set including loanwords and in well-formedness judgements on novel words than in native Hungarian words. The proportion of geminates for voiceless obstruents is smaller in well-formedness judgements on novel words than in the data set containing a mix of native words and loanwords which are part of the native vocabulary. This is expected, given the hypothesis that gemination is due to perceptual effects in loanwords: such effect could not play a role in the adaptation / well-formedness judgement of novel forms, as participants were only presented with written forms corresponding to pronunciation, but they were not presented with auditory stimuli. Therefore the only factor they could have been influenced by is the distribution of geminates in the native phonology.

There is one crucial and conspicuous difference between corpus data and well-formedness judgements: geminate liquids were rated fairly highly by native speakers in well-formedness tasks even though they are fairly marked geminates cross-linguistically as well as in the native phonology.[1]

3.2.2 Experiment II: Intervocalic gemination following short vowels in disyllables

Although loan gemination is most common and productive in word-final position following short stressed vowels in monosyllabic words, there are some examples of intervocalic gemination as well. The context for this type of gemination is a short consonant spelled as a double letter preceded by a short stressed vowel or flanked by two short vowels. There are hardly any examples of consonants spelled with a single letter undergoing loan gemination in intervocalic position. However, the reason for this might be the fact that vowels preceding consonants spelled with a single letter are generally pronounced long in the donor languages which Hungarian speakers borrow words from, and loan gemination does not apply after long vowels.

Even in the presence of orthographic geminates or digraphs, not all intervocalic consonants are pronounced as geminates in these contexts. There is a large amount of degemination, especially in more recent loanwords and in the adaptation of foreign names.

In the native Hungarian lexicon, there is a higher number of disyllables (containing short vowels) than monosyllables, and the distribution of singletons and geminates is quite different in these two contexts. In the case of some consonants in word-final position, the distribution of singletons and geminates is almost equal, whereas in intervocalic position, geminates are rare across the board compared to singletons.

[1]It is unclear why geminate liquids were so popular with native speakers of Hungarian. One possibility is that there are several high frequency monosyllables ending in geminate liquids in Hungarian. However, this is only speculation and it has not been tested as a hypothesis.

The nonce disyllable well-formedness task was designed to test whether native speakers are sensitive to the patterns in the Hungarian lexicon and (dis)preferences for gemination in the adaptation of English or German disyllabic words with intervocalic consonants.

The questionnaire contained 71 disyllabic nonce word pairs (142 items, see Appendix B) with a consonant in intervocalic position: members of each word pair were minimal pairs, only differing in the length of the word-medial consonant (e.g. *biki* /biki/ and *bikki* /bik:i/). All consonants which may occur in loanwords were included among the test items. Items were constructed to look like possible English loanwords ending in *-ic* (spelled and pronounced as /ik/ in the Hungarianized version), *-y* (spelled and pronounced as [i]) and *-ish* (spelled as *-is* and pronounced as /ɪʃ/, and possible German loanwords ending in *-er*. The vowel preceding the intervocalic consonants were /ɒ, ɛ, i, u/, represented in spelling by *a*, *e*, *i* and *u*, respectively.

The task was administered online and 87 native speakers of Hungarian participated in it. They were asked to choose whether the member of each word pair containing a singleton or a geminate is more acceptable as a possible Hungarian word or a Hungarianized loanword. The presentation of items was also similar to the task on monosyllables: written stimuli were used and spelling fully represented pronunciation. The order of presentation was balanced: sometimes the form containing the geminate was the first option, sometimes the other way around.

Results

The distribution of intervocalic geminates in native speaker judgements compared to corpus data are shown in Figure 3-8. Well-formedness judgement data are represented by the yellow line: voiceless stops (39% (n=1042)) > nasals (31% (n=696)) > voiced stops (30% (n=957)) > voiceless fricatives (28% (n=1036)) >

voiceless affricates (27% (n=690)) > voiced fricatives (10% (n=1039)).

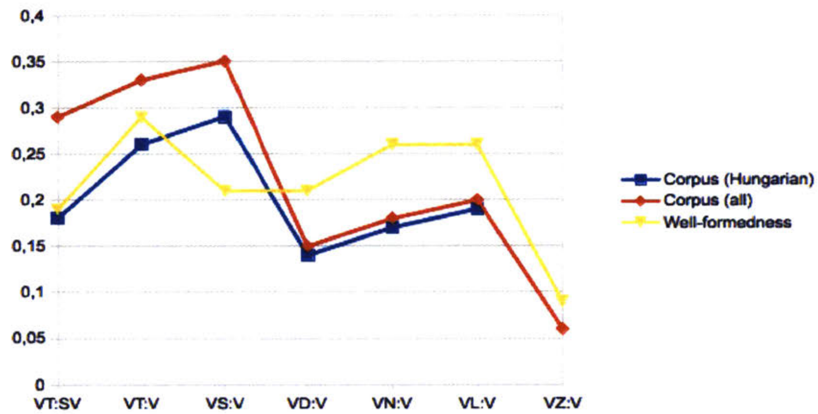

Figure 3-8: Intervocalic geminate frequencies

It is clear from the figure that gemination patterns in well-formedness judgements are not consistent with corpus data. Even in the corpus, intervocalic gemination did not line up with geminate markedness hierarchies as well as gemination in monosyllables did, but both word-finally and intervocalically, there was a clear difference between gemination of voiceless and voiced consonants: voiceless consonants were more frequently geminated in both positions. However, in native speaker judgements, there is not even a clear difference between voiceless and voiced geminates. This indicates that native speakers of Hungarian do not have much intuition for replicating the distribution of intervocalic gemination in the corpus or in loanwords. This would be hard to do because (1) intervocalic geminates are rare in the native phonology and in loanwords as well (unless the consonant is spelled with a double letter in the source word) and (2) it is hard to decide how to calculate the distribution of intervocalic gemination, as there is countless data as a result of productive morphophonological processes.

3.2.3 Testing the native grammar on novel words

The grammar outlined in this chapter is the grammar of native Hungarian words before German and English loanwords began to enter the language and only consists of markedness constraints imposing restrictions on word minimality, vowel and consonant length in VC sequences and gemination for different consonant classes.

The current Hungarian vocabulary (including well-integrated loanwords as well) is shaped by other factors as well, for example, faithfulness (influenced by perceptual similarity) to words forms in the donor language. As was shown previously, well-formedness judgements on word-final gemination are a closer match to these data.

Figure 3-9 shows the distribution of geminates found in the native phonology (without English and German loanwords) (blue line), predicted by the grammar (orange line) and in native speaker judgements on novel words (yellow line). Predictions of the model are not accurate for voiceless affricates and liquids: it underestimates the proportion of geminates for voiceless affricates and overestimates the gemination of liquids. The reason for this, as mentioned earlier, is that the use of markedness constraints cannot account for the high rate of gemination in one context and the low rate of gemination in another position for certain consonant classes compared to others (e.g. voiceless affricates are very often geminated word-finally and very rarely geminated intervocalically).[2] Interestingly, native speaker judgements also show similar patterns of liquid gemination. Apart from these deviations, well-formedness judgements turned out to be as expected: for voiceless consonants, the proportions of geminates are larger - especially in the case of

[2]Geminate markedness should not be confined to position. Moreover, gemination in word-final position in general is more marked than gemination in intervocalic position, as it is harder to perceive.

voiceless affricates and voiceless stops.

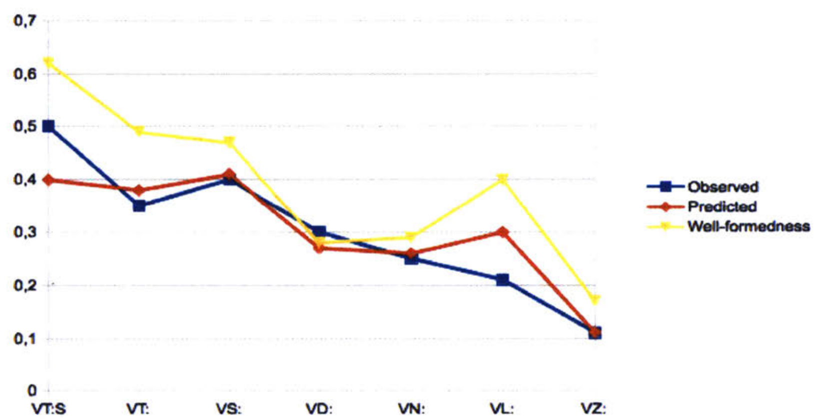

Figure 3-9: Word-final geminate frequencies observed in the corpus, grammar fit and well-formedness judgements

Figure 3-10 shows the distribution of geminates in intervocalic position in the native Hungarian lexicon (blue line), in the grammar based on the native phonology (orange line) and native speaker judgements (yellow line). Similarly to the previous comparison with the current lexicon (including well-integrated loanwords as well), native speaker judgements in intervocalic position do not line up with the native grammar, either.

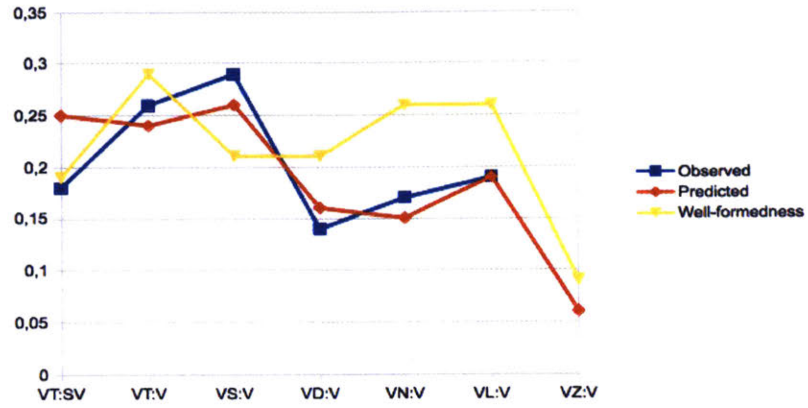

Figure 3-10: Intervocalic geminate frequencies observed in the corpus, grammar fit and well-formedness judgements

3.3 Learnability of geminate markedness and patterns of loan gemination

In the previous sections, we have seen that loan gemination is only productive in word-final position in monosyllables, following short vowels. In intervocalic position in disyllabic words, consonants are only geminated when spelled with a double letter in the source word or ending in suffixes like *-er*. We have also seen that patterns of loan gemination have roughly been modeled on distributions of geminates already present in the language. The following figure shows the distribution of geminates by consonant class in different sets of monosyllables: (1) native Hungarian words (blue line), (2) all words (including English and German loanwords) (orange line), (3) nonce words judged by native speakers of Hungarian (yellow line), and (4) loan gemination as it is happening currently (green line).

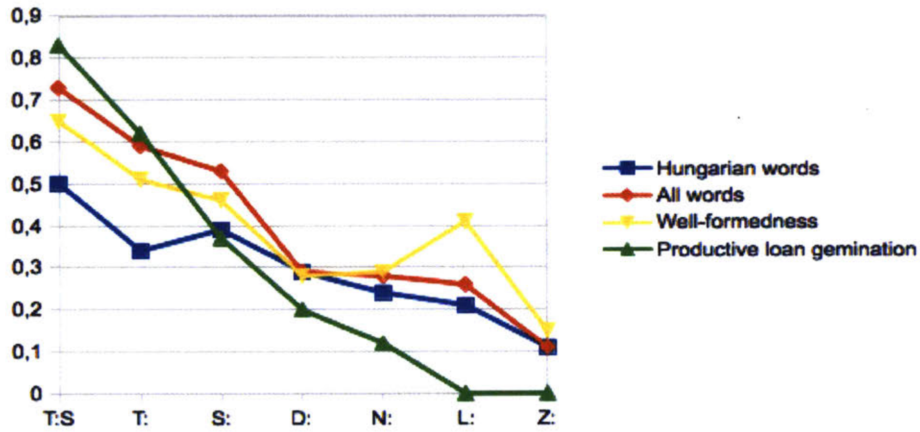

Figure 3-11: Word-final gemination by consonant class in different data sets

As shown in Figure 3-11, the distribution of most consonant classes as geminates is fairly consistent across different data sets, except for liquids, which show considerable fluctuation. In all data sets, geminate voiceless stops, affricates and fricatives are more common than geminate voiced stops and nasals, and geminate voiced fricatives are extremely rare. Geminate liquids are in the same league as voiced stops with respect to frequency in Hungarian words and Hungarian words + loanwords combined, but are much more frequent in loanword and nonce word well-formedness data. There are no English or German loanwords in the Hungarian language containing a short vowel followed by a liquid which did not contain orthographic geminates in the source words. Source words containing double consonant letters in word-final position in monosyllables are always borrowed into Hungarian with a phonetic geminate, and this is what could have inspired participants to prefer the geminated form of liquids (especially [l]) in many cases.

There is also a much more fine-grained difference between consonant classes with respect to their distribution of geminates in the Hungarian grammar. The order of frequency for geminates is as follows: voiceless affricates > voiceless stops

> voiceless fricatives > voiced stops > nasals > liquids > voiced fricatives. This is in line with some of the implicational hierarchies of universal geminate markedness reported in the literature: voiceless obstruents are more often geminated than other consonants, and the likelihood of gemination is inversely correlated with sonority. It can be accounted for by the following markedness constraints in this particular ranking order of hierarchy: *ZZ >> *LL >> *NN >> *DD >> *SS >> *TT >> *TTS.

*ZZ: Geminate voiced fricatives are prohibited.
*LL: Geminate liquids are prohibited.
*NN: Geminate nasals are prohibited.
*DD: Geminate voiced stops are prohibited.
*SS: Geminate voiceless fricatives are prohibited.
*TT: Geminate voiceless stops are prohibited.
*TTS: Geminate voiceless affricates are prohibited.

As was shown in earlier chapters, the proportions of distribution of geminates in intervocalic position in disyllabic words is similar to the word-final context in monosyllables. However, in intervocalic position, the number of geminates is quite low compared to singletons for each consonant class, whereas in word-final position, geminates of certain consonant classes are as frequent as singletons, or even more frequent. This can be explained by the principle of Bimoraic Minimality, which forces gemination in monosyllables containing short vowels:

BIMORAIC MINIMALITY: All words should consist of at least two moras.

This constraint is often violated, of course, and some of the markedness constraints against gemination must be ranked above and some below it. However, while trying to implement it as a strict ranking in categorical OT, we would run into problems. These problems can be solved using weighted constraints. That type of analysis will be presented in a later chapter.

The goal of this section is to investigate whether the patterns of gemination present in the native Hungarian phonology can be learned based on phonotactic constraints discovered by the learner and be generalized and applied to novel words, be it loanwords or nonce words. The question also arises whether such patterns emerge when the learner has access to fine grained detail or is only able to extract the patterns when attention is drawn to the relevant information. The results of two sets of simulations will be reported: (1) a model trained on native Hungarian words and tested on loanwords, and (2) a model trained on all the words present in the current Hungarian lexicon (including German and English loanwords) and tested on nonce word well-formedness judgement data.

3.3.1 Settings

The UCLA Phonotactic Learner can be run using two data sets (training and testing data), and a feature file (containing all the distinctive features of consonants and vowels. Gram size and O/E can also be adjusted. Gram size is the number of feature matrices in the constraint. The O/E (observed over expected) value is a measure of constraint effectiveness. It is the ratio of the number of times a constraint is violated in the learning data to the number of times it would be expected to be violated, based on the grammar learned so far. The more powerful a constraint is, the lower O/E value it has.

In this chapter, the results of two sets of simulations will be reported. Details of these simulations are shown in the following table.

	Training data	Testing data	Gram size	O/E
1a	native monosyll., full	loanwords	3	0.35
1b	native monosyll, VC	loanwords	2	0.35
1c	native monosyll., VC(class)	loanwords	2	-
1d	native monosyll., final C	loanwords	2	-
1e	native monosyll., final C(class)	loanwords	2	-
2a	all monosyll., full	nonce words	3	0.35
2b	all monosyll., VC	nonce words	2	0.35
2c	all monosyll., VC(class)	nonce words	2	-
2d	all monosyll., final C	nonce words	2	-
2e	all monosyll., final C(class)	nonce words	2	-

Table 3.2: Simulations

The goal of Simulations 1a-e is to investigate whether patterns of loan gemination could have been learned from the native lexicon that existed before the adaptation of German and English loanwords. In Simulation 1a, monosyllables (excluding recent loanwords) were used as training and loanwords as testing data. Since most of these words consist of three sounds, gram size was set at 3. As the learning process tends to take extremely long when gram size of 3 is used, the O/E value was set at .35, so that the learner can discover the most powerful and relevant constraints and terminate after a reasonable amount of time. In Simulation 1b, the grammar was learned on rhymes of the learning data from Simulation 1a. Simulation 1c is a further simplified version of 1b: the final consonants of rhymes are represented as consonant classes, instead of individual consonants (e.g. not [p], [t] and [k], but T (voiceless stops)). In simulations 1d and 1e, the learner was given no other information but the final consonants: individual consonants in 1d and consonant classes in 1e. When consonant classes instead of individual consonants were included in the learning data, a feature file was used which contained the features which distinguish manners but not places of articulation. Gram size was reduced to 2 when the maximum length of phoneme strings was 2. O/E was not specified in those cases when the grammar can be learned by discovering relatively few constraints and only finding the most powerful constraints would cause the generalizations to be coarse and rather inaccurate.

The goal of Simulations 2a-e is to determine whether well-formedness judgements can be based on generalizations learned from the current Hungarian lexicon (including fairly recent loanwords as well). In 2a, the training data were full forms of monosyllabic words containing a short vowel + singleton / geminate sequence, including native Hungarian words, older and fairly recent borrowings. The grammar was tested on nonce word well-formedness judgement data. 2b and 2c were trained on the rhymes of the training data from 2a: 2b has final consonants and 2c has final consonant classes. 2d and 2e were trained on the final consonants (2d) and consonant classes (2e) of 2a, without any other information given to the learner.

3.3.2 Presentation of results

As described earlier, the model assigns scores to forms based on sums of violations for constraints (see *Definition 1* in Hayes and Wilson (2008). The maxent value can be calculated based on the scores assigned by the learner as follows: P*(x)=exp(-h(x)), where x is the form and h(x) is the score assigned to each form. This was done by a python script which goes through a list of scores and converts them to maxent values using the aforementioned formula.

Maxent values can be converted to two kinds of probabilities to suit our purposes. Hayes and Wilson (2008) use global probabilities, that is, they calculate the probability of a form (e.g. *ibb* [ib:]) given all the other forms, which is calculated as follows:

$$P(ibb) = \frac{exp(-h(ibb))}{exp(-h(ibb)) + exp(-h(opp)) + exp(-h(et)) + ...}$$

$$P(ibb) + P(ib) + P(opp) + P(et) + ... = 1$$

This is useful when we are interested in the probability of forms compared to all the other forms in the language or in any set of data. However, when calculating the probability of geminates compared to singletons, we are generally comparing distributions within consonant classes or individual consonants, and not in comparison to all other forms in the language. Therefore, we have to calculate the probability of a form given another form (e.g. the probability of *ibb* [ib:] given *ib* [ib]), that is, local probability:

$$P(ibb) = \frac{exp(-h(ib))}{exp(-h(ibb)) + exp(-h(ib))}$$

$$P(ibb) + P(ib) = 1$$

Local probability provides a better comparison with the loan gemination and nonce word well-formedness data, because a direct comparison of geminate and singleton forms corresponds more closely to how people borrow loanwords and to the task participants were asked to perform in the experiment.

3.3.3 Results

All learning simulations were run three times. The results of those simulations will be reported which are the most accurate, that is, the closest match to the data they were run on.

3.3.4 Simulation 1: Learning loan gemination from the native phonology

In the first set of simulations, the learner was trained on lexical frequencies of a word list containing monosyllables ending in a short vowel + short consonant / geminate sequence. This list does not include English and German loanwords, but includes all monosyllables containing a short vowel + short / long consonant sequences which had been present in the language before English and German loanwords were borrowed. The learner was tested on older and more recent English and German loanwords. It has been claimed that German loanwords were borrowed into Hungarian as a result of a perceptual similarity effect: in the dialect these words were borrowed from, word-final consonants following short stressed vowels were actually pronounced as long, and gemination in English loanword is modeled on the adaptation of German loanwords.

Although this perceptual effect might have played an important role in the adaptation of loanwords, the process could have been affected by lexical frequencies in the native Hungarian lexicon. As described earlier, among native Hungarian monosyllables, there is a fairly large number of geminates among voiceless stops, affricates and fricatives compared to other consonant classes (see Figure 3-1): this is the same pattern which can be observed cross-linguistically and in the adaptation of loanwords.

In 1a, the probability (well-formedness) of loanwords was predicted based on a list of whole word shapes in the native Hungarian lexicon. In 1b and 1c, it was predicted based on rhymes: in 1b, training data contained vowel + consonant sequences, while in 1c, final consonants were represented as consonant classes (that is, [aT], instead of [at], [ak] and [ap]). In 1d and 1e, probabilities were predicted based on final consonants: in 1d, consonants were represented individually (e.g. [p], [t] and [k]), whereas in 1e, they were simplified to consonant classes (T as voiceless stops).

In all five simulations, results were compared to observed probabilities of a loanword having a final geminate. In 1a, probability scores (on a 0-1 scale) were assigned to each loanword. When the final consonant in a given loanword is always geminated, the score 1 was assigned to it. When the word-final consonant is always pronounced as a singleton, it was assigned a probability of 0. When there is free variation, the score of the form was 0.5. When the variation is very rare or stigmatized, the form was assigned a score to 0.1 to 0.2 (if the geminated form is much less common than the one with the singleton). Most loanwords have a 0, 0.5 or a 1. In 1b and 1c, the learner was trained on rhymes, therefore the expected probabilities were assigned to the testing data accordingly: the probabilities of existing loanwords were pooled over rhymes (e.g. loanwords ending in [it] or [it:] were assigned a score, not individual loanwords such as *fitt* or *snitt*) in 1b and also consonant classes in 1c. In 1d and 1e, the expected probability of word-final consonants (1d) and consonant classes (1e) were calculated based on all words ending in each consonant (1d) or consonant class (1e): for example, if there are a certain number of words ending in [t] (1d) (or [p], [t], [k] as T (voiceless stop) (1e)) and all of them end in a geminate, then the score of word final [t:] in loanwords is 1.

Results are shown in the following scatter plots. In 3.2, probabilities assigned to loanwords in Simulation 1a are on the x axis and the observed probabilities are on y. In 3.3, the results of Simulation 1b are plotted against the observed probabilities of loanword rhymes (vowel + geminate consonant). 3.4 shows rhymes, too, but final consonants are represented as consonant classes. In 3.5 and 3.6, predicted and observed probabilities of final geminates (as consonants (3.5) and consonant classes (3.6) are correlated using Kendall's Tau to check for concordant pairs.

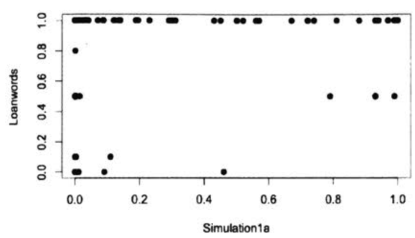

Table 3.3: 1a-Loanwords (Kendall's Tau=0.17)

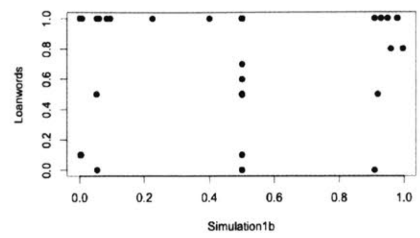

Table 3.4: 1b-Loanwords (Kendall's Tau=0.09)

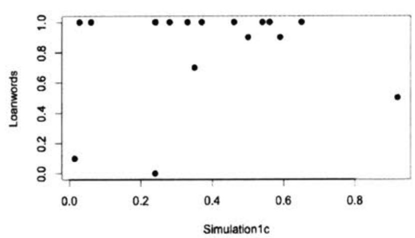

Table 3.5: 1c-Loanwords(Kendall's Tau=0.11)

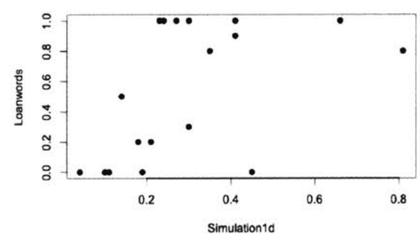

Table 3.6: 1d-Loanwords(Kendall's Tau=0.5)

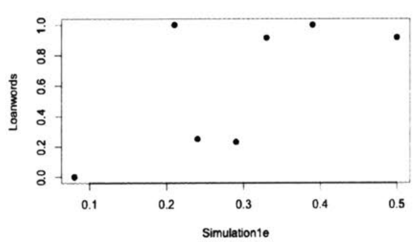

Table 3.7: 1e-Loanwords (Kendall's Tau=0.66)

What we can observe is that the less detail the learner is given, the more accurate its predictions are on loan gemination. In 1a, with a gram size of 3, several constraints are learned which pose restrictions on the whole shape of the word, which may be higher ranked than constraints regulating consonant length in word-final position. It often occurs that two forms which only differ in the initial consonant and would behave the same way as loanwords (e.g. *pakk* 'package' and *lakk* 'polish' are assigned a different score by the learner, simply because of the discovered a constraint banning the [+lateral][-cons, +low, +back][+long] sequence. These two words, however, do have the same probability of gemination in the loanword adaptation process.

Predicted probabilities assigned to rhymes in Simulation 1b are a bad match to actual loanword well-formedness, too. In this case, the learner may assign different probabilities to two rhymes containing the same geminate but a different vowel, but in reality, two loanwords having such rhymes would be equally likely to be geminated. For example, the learner assigned the probability of 0.9 to words ending in [ok:] but only 0.2 to those ending in [ɛkː], however, words ending in both rhymes has the observed probability of 1 in existing loanwords. We run into similar problems in 1c, where consonants are represented as consonant classes, because consonant classes show a fairly uniform behavior in loanwords with respect to gemination, regardless of the preceding consonant.

When the learner is only provided with information on final consonants (1d) or consonant classes (1e), the accuracy of predicting the probability of gemination in loanwords improves. Since loan gemination is fairly uniform across consonant classes, probabilities predicted by 1e are a better match than those assigned by 1d, but still are not a perfect match. The learner was able to predict the actual corpus frequencies fairly accurately, as is shown in the following table. TT stands for geminate voiceless stops, TTS for geminate voiceless affricates, SS for geminate voiceless fricatives, DD for geminate voiced stops, NN for geminate nasals, LL for geminate liquids, and ZZ for geminate voiced fricatives.

Geminates	1e	Corpus	Loanwords
TT	0.33	0.34	0.915
TTS	0.5	0.5	0.917
SS	0.39	0.39	1
DD	0.29	0.29	0.23
NN	0.24	0.24	0.25
LL	0.21	0.21	1
ZZ	0.08	0.11	0

Table 3.8: Predicted probabilities of word-final gemination in consonant classes

We can see that the probabilities predicted by Simulation 1e line up well with frequencies found in the corpus (that is, the training data). The learner discovered markedness constraints using distinctive features, which can be translated to the following types of constraints: constraints on individual consonant classes (e.g. *ZZ) or groups of consonant classes (*geminate sonorants, *short voiceless obstruents). These constraints are similar to the grammar sketched in the beginning of this chapter. But there is some mismatch between these probabilities and the loanword data. In loanwords, voiceless stops, affricates and fricatives are practically always geminated, voiced stops and nasals are sometimes geminated, and voiced fricatives are never geminated. These patterns are to be observed in the native phonology as well, but are not very conspicuous. The main difference between corpus data / learned probabilities and early loanword adaptation is in the gemination of liquids, which is due to the fact that no German and English words which contained a single *l* in spelling was borrowed into Hungarian, and orthographic geminates have always been borrowed as phonetic geminates. Therefore, we do not know if an [l] speled as a single letter would have been borrowed as a singleton or a geminate at the time when German and English loanwords started entering the language, but we know that these days it would be borrowed as a single [l].

These results indicate that the earliest process of loan gemination cannot be accounted for exclusively by native phonotactic well-formedness. Although voiceless obstruents, the most frequently geminated consonants on loanwords, are the most common geminates in the native phonology, but unlike in loanwords, they are still less frequent or at most as frequent as singletons. Therefore, other factors, such as spelling and perceptual similarity, must have played a role in the early adaptation of final consonants in loanwords.

3.3.5 Simulation 2: Learning to judge novel forms based on the current lexicon (including more recent loanwords as well)

The goal of Simulations 1a-e was to test whether the early adaptation of German and English loanwords could have been based on lexical frequencies in the native phonology. We have found that the early adaptation of loanwords was not entirely mirorring lexical frequences and other factors must have been at play, too. Loan gemination used to be extremely widespread, but today, it is becoming more and more optional and restricted to voiceless obstruents, especially voiceless affricates and stops. Simulations 2a-e were run in order to see if well-formedness judgements and current loan adaptation processes are influenced by the distribution of geminates in the contemporary lexicon of Hungarian, including native words and well-known loanwords, and whether they mirror lexical frequencies more closely than earlier processes of loanword adaptation did.

Data sets are different but the setup is very similar to Simulations 1a-e. The training data contain the list of monosyllabic words (both native vocabulary and loanwords) ending in a short vowel + short consonant / geminate sequence currently used in the Hungarian language. Testing data were nonce word well-formedness judgements and loan gemination - the way it is productive today. 1a had full words, 1b and 1c rhymes, and 1d and 1e final consonants.

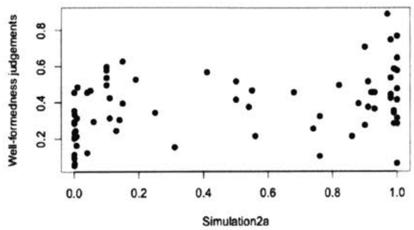

Table 3.9: 2a-Well-formedness judgements (Kendall's Tau=0.38)

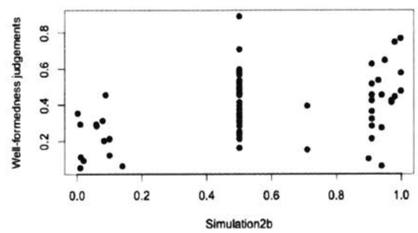

Table 3.10: 2b-Well-formedness judgements (Kendall's Tau=0.38)

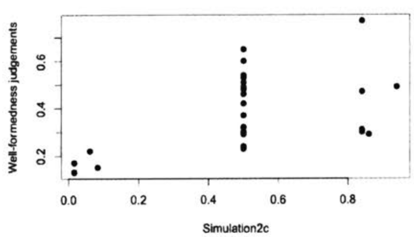

Table 3.11: 2c-Well-formedness judgements (Kendall's Tau=0.49)

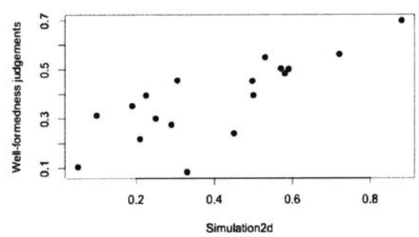

Table 3.12: 2d-Well-formedness judgements (Kendall's Tau=0.78)

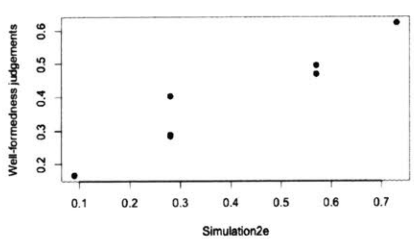

Table 3.13: 2e-Well-formedness judgements (Kendall's Tau=0.95)

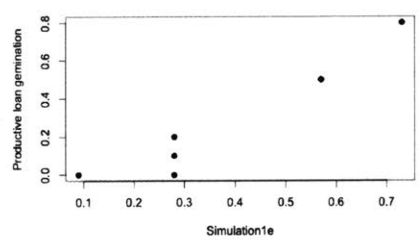

Table 3.14: 2e-Productive loan gemination (Kendall's Tau=0.95)

In all simulations, predicted and observed probabilities line up better than in

Simulations 1a-e, but even here, we can see that the more detail the learner is provided with, the less accurate its predictions are, since loan gemination is generally extended to consonant classes and is not influenced by fine-grained differences. However, in judging the well-formedness of novel forms, some of the fine grained differences do seem to matter, as is shown in 3.11, but the match between predicted and observed probabilities is perfect only when final consonant classes are considered and apart from that, the learner is not provided with any other specific information (e.g. preceding vowel).

Geminate	2e	Corpus	Well-formedness	Loan gemination
TT	0.57	0.59	0.49	0.62
TTS	0.73	0.73	0.62	0.83
SS	0.57	0.53	0.47	0.37
DD	0.28	0.29	0.28	0.2
NN	0.28	0.28	0.29	0.12
LL	0.28	0.26	0.40	0
ZZ	0.09	0.11	0.17	0

Table 3.15: Predicted probabilities of word-final gemination in consonant classes

As is shown in Table 3.15, voiceless affricates are geminated the most frequently in the current Hungarian vocabulary, which includes well-known loanwords as well. Geminate voiceless stops and fricatives are also very common. There are only small differences between the distribution of geminates in the case of voiced stops, nasals and liquids, and the learner did not make those differences. Voiced fricatives very rarely occur as geminates. The learner discovered the following constraints to account for these patterns:

Constraint	Shorthand	Definition	Weight
*[+del.rel,-long]	*TS	No short affricates	1.414
*[+voice,+long]	*DD, *NN, *LL, *ZZ	No long voiced consonants	0.966
*[+son,-lat,-long]	*N	No short nasals	0.826
*[+son,-lat,+long]	*NN	No long nasals	0.808
*[+del.rel,+long]	*TTS	No long affricates	0.682
*[-voice,-long]	*T, *TS, *S	No short voiceless consonants	0.267

Table 3.16: Predicted probabilities of word-final gemination in consonant classes

Results of Simulation 2 indicate that well-formedness judgements on novel words are influenced by gradient phonotactic well-formedness which is learnable from lexical frequencies. The same patterns are reflected in native speakers' judgements on the well-formedness of novel words as well as in contemporary loan gemination processes. However, liquids are rated much higher than expected in well-formedness judgements and it is still unclear why. Contemporary loan gemination can also be partially predicted based on already existing patterns in the language, but for the full picture, the influence of perceptual effects also has to be taken into account.

Chapter 4

The role of perceptual similarity

As described in Chapter 3, gradient phonotactic well-formedness in the native phonology and geminate markedness alone are not sufficient to account for the loan gemination data. Gradient patterns of gemination in the native phonology do shape the structure of loanwords, but they do not predict why there is a clearer preference for gemination of voiceless stops and affricates in monosyllabic loanwords than in the native phonology and well-formedness judgements. Although geminates are more common in monosyllables than intervocalically or word-finally in polysyllables even in the native phonology, it does not explain why loan gemination does not apply in intervocalic position (unless there is a double letter in the source word spelling or the word consists of a monosyllabic root + *-er*) or word-finally in polysyllables (unless the source vowel was stressed and / or the source consonant was spelled with a double letter).

Voiceless stops and affricates show similar behavior both in English and Hungarian in that they can trigger shortening of the preceding vowel, but in Hungarian, only as geminates. In English, they shorten the preceding vowel to varying degrees depending on context, as reported by Port (1981). As shown by the figure below, there is some amount of shortening for both tense and lax vowels in each context, but the effect is much stronger in the case of monosyllables in word-final position.

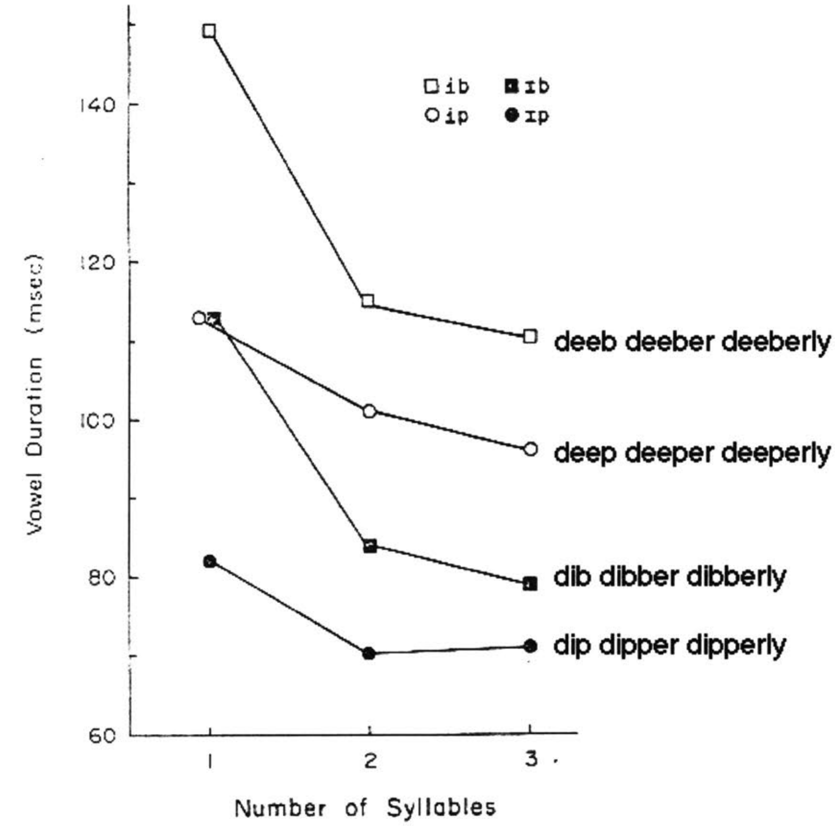

Figure 4-1: Vowel durations of lax and tense vowels before voiced and voiceless stops (Port, 1981)

However, this effect has been shown to be much weaker and very inconsistent in Hungarian (Magdics, 1966; Kovács, 2000; Olaszy, 2006). An example from (Olaszy, 2006), the vowel [ɛ] is shown in a CVC context followed by voiced vs. voiceless stops in Table 4.1. Results were similar for other voiced-voiceless pairs as well: there is only a difference of a few milliseconds between vowels before voiced and voiceless consonants, and it is not always the vowel preceding the voiceless consonant that is shorter.

[ɛ]	Mean duration (ms)
before [b]	90
before [p]	89
before [d]	92
before [t]	92
before [g]	91
before [k]	101

Table 4.1: Mean durations of [ɛ] followed by voiced and voiceless stops (based on Olaszy (2006))

Apart from the effect of voicing on the duration of the previous vowel, there have been different claims about the effect of gemination on the preceding vowel. Ham (2001) reports that Hungarian is one of those languages in which vowel quantity is not influenced by gemination, however, it was not investigated in much detail in his work. Neuberger and Beke (2017) analyzed the spontaneous speech of male native speakers of Hungarian and investigated the effect of duration voiceless stops have on the preceding vowel. They have found that gemination does not have a significant effect on the duration of the preceding vowel. Both studies investigated gemination in intervocalic position.

No studies have investigated the effect of gemination on preceding vowel duration in monosyllables, apart from the preliminary experiment with two speakers I have conducted for Magyar (2014). The results of this experiment were different from what was found about intervocalic geminates. In this context, gemination has a significant effect on the duration of the preceding vowel, as shown in the table below. The vowel [ɛ] was used and there were five repetitions by each speaker and for each consonant ([p], [t], [k], [ts], [tʃ]).

[ɛ]	Before geminate (ms)	Before singleton (ms)
Before voiceless stops	56	96
Before voiceless affricates	63	102

Table 4.2: Mean durations of [ɛ] before geminate and singleton voiceless stops and affricates

Although the difference is bigger between the length of vowels in monosyllables than in polysyllables, vowels are shorter in polysyllables, therefore we would expect gemination in polysyllabic words both word-finally and intervocalically. However, this is not the case: without the influence of orthography, consonants are not geminated in intervocalic position in loanwords and word-final consonants are only geminated in polysyllables when the preceding vowel was stressed in the source word. The explanation for this is that, first, consonant duration is easier to perceive in intervocalic position (because there are more cues) than in final position. In addition, even if English vowels are shorter in intervocalic position than Hungarian vowels, gemination could not be used as a strategy to shorten the preceding vowel, as intervocalic geminates do not have this effect. In word-final position, source word stress matters: if the preceding vowel was stressed in the source word, the final consonant is likely to be geminated, although there are not many examples for this word shape among loanwords.

The fact that voiceless stops and affricates have a shortening effect on the previous vowel in some ways in both languages (as singletons in English and as geminates in Hungarian), and that these consonants are the most likely to participate in loan gemination processes, suggests that gemination could be a strategy to shorten the preceding vowel. In this chapter, I report the result of an experiment which intends to test Hypothesis II:

- Hypothesis II: Gemination is a strategy to preserve the shortness of the source vowel.

'Shortness' is relative: an English vowel is considered to be extra short by native Hungarian listeners if the vowel in question is shorter than its substitute vowel (used in Hungarianized loanwords) and is closer in duration to the substitute vowel preceding geminate voiceless stops or affricates. To test this hypothesis, two sub-hypotheses should be evaluated:

- Hypothesis II-a: If an English vowel is shorter than a Hungarian vowel, Hungarian listeners perceive it as closer in duration to the same vowel preceding a geminate than to the one preceding a singleton.
- Hypothesis II-b: English long vowels preceding voiceless stops and affricates are closer in duration to Hungarian short vowels than long vowels in the same context.

4.1 Experiment

To test Hypothesis II-a and II-b, I conducted perception experiments. The goal of these experiments was to test (1) whether Hungarian listeners think that a Hungarian vowel preceding a geminate is a closer match to an English vowel which is shorter than the Hungarian vowel before a singleton, and (2) whether Hungarian listeners perceive long English vowels preceding voiceless stops and affricates as a closer match to short Hungarian vowels instead of long ones before the same consonants.

4.1.1 Participants

104 native speakers of Hungarian participated in three online perception experiments. Each subject participated in only one experiment. They volunteered to participate in the experiment and did not receive payment.

4.1.2 Methods

The experiment was administered online and consisted of three different parts. Each experiment contained 8 questions: 4 targets and 4 fillers. Items were presented in an ABX format: participants were asked to listen to the stimuli in a row and decide whether stimulus X sounded similar to stimulus A or B. Stimuli were auditory and their written forms were not presented to participants. Participants were asked to indicate their responses in a forced choice format ((1) A (2) B or (2) A (1) B).

4.1.3 Stimuli

The stimuli were recorded by two females: a native speaker of Hungarian and a native speaker of American English. A and B were monosyllabic words pronounced by the Hungarian speaker, and X was a word pronounced by the American English speaker. Each target word began with a /b/ and ended with a /k/, in order to exclude effects due to segmental differences. All three experiments contained two target words (and vowel pairs) twice: vowel length was manipulated in order to create a version in which the English vowel was shorter than the Hungarian vowel, and another one, in which it was of the same length as or longer than the Hungarian vowel. The Hungarian vowel in each word pair was the one that was used as a substitute in the loanword when the source word had the English vowel of the word pair. The English and Hungarian words differed only in the quality of the vowel, and the two Hungarian stimuli in the length of the consonant. Final stops were released by both speakers and their length was not manipulated. Target items are shown in the table below.

	V duration (ms)	C duration (ms)
back [bæk] (same length, natural)	93	135
bek [bɛk]	91	130
bekk [bɛkː]	64	230
back [bæk]	93	135
bek [bɛk] (shorter, manipulated)	64	130
bekk [bɛkː]	64	230
bick [bɪk] (shorter, natural)	52	106
bik [bik]	91	120
bikk [bikː]	52	178
bick [bɪk]	52	106
bik [bik] (same length, manipulated)	53	120
bikk [bikː]	52	178
bock [bɑk] (shorter, manipulated)	73	123
bak [bɒk]	102	138
bakk [bɒkː]	75	192
bock [bɑk] (same length, natural)	103	123
bak [bɒk]	102	138
bakk [bɒkː]	75	192
book [bʊk] (same length, manipulated)	53	125
buk [buk]	55	122
bukk [bukː]	52	230
book [bʊk] (shorter, natural)	51	125
buk [buk]	99	122
bukk [bukː]	52	230

	V duration (ms)	C duration (ms)
buck [bʌk] (shorter, natural)	73	122
bak [bɒk]	101	138
bakk [bɒkː]	55	192
buck [bʌk]	73	122
bak [bɒk] (shorter, manipulated)	42	138
bakk [bɒkː]	55	192
blook [blʊk] (shorter, natural)	45	146
bluk [bluk]	91	141
blukk [blukː]	47	217
blook [blʊk]	45	146
bluk [bluk] (same length, manipulated)	47	141
blukk [blukː]	42	217

Table 4.3: Target items (natural and manipulated)

Fillers were included to test Hypothesis II-b: that is, whether native speakers of Hungarian perceive English long (but phonetically shortened) vowels as closer to Hungarian short vowels. A small set of voiceless obstruents (stops and affricates) were used in the examples. The words contained [u] or [i]. The reason for choosing these particular vowels is that these are the only Hungarian vowels which are of the same quality as their English counterparts and at the same time, have a long-short distinction. The fact that short-long distinction for Hungarian high vowels is disappearing (White and Mády, 2008; Mády et al., 2008; Mády, 2010) did not influence the outcome of the experiment, as this length distinction was retained in the speech of the person who produced the Hungarian stimuli. Vowel length was not maipulated. Stimuli are shown in the table below.

	V duration (ms)	C duration (ms)
neek [nik]	75	130
nik [nik]	68	150
ník [ni:k]	149	118
fupe [fup]	74	111
fup [fup]	56	129
fúp [fu:p]	180	135
keet [kit]	74	120
kit [kit]	67	136
kít [ki:t]	155	119
zuke [zuk]	74	123
zuk [zuk]	72	128
zúk [zu:k]	115	114
beets [bits]	85	269
bic [bits]	74	242
bíc [bi:ts]	161	260

	V duration (ms)	C duration (ms)
slupe [slup]	59	158
szlup [slup]	80	136
szlúp [slu:p]	138	118
bleep [blip]	61	124
blip [blip]	73	117
blíp [bli:p]	144	125
kleets [klits]	61	279
klic [klits]	85	290
klíc [kli:ts]	134	254
klupe [klup]	55	148
klup [klup]	52	123
klúp [klu:p]	115	119
bleet [blit]	86	110
blit [blit]	68	129
blít [bli:t]	144	110
klute [klut]	53	108
klut [klut]	73	133
klút [klu:t]	116	119

4.1.4 Results: Gemination and vowel length

The statistical model is testing the following assumption: Hungarian words containing singletons and geminates lie along a perceptual dimension, and the English word is mapped noisily onto the same dimension, with normally distributed noise. Participants in the experiment respond with whichever Hungarian word the percept is closer to. A mixed effects probit model was fitted using the lme4 package (Bates et al., 2014) in R:

- glmer(Geminate ~ Dur*V+(Dur|Subject), family = binomial(link = 'probit'))

This model can predict the probability of choosing the Hungarian word with a geminate / singleton, depending on the length of the preceding vowel compared to the length of the vowel in the English word. 'Geminate' is the response variable ('yes' or 'no'). 'Dur' indicates whether the English vowel is shorter / longer than or of the same length as the Hungarian vowel. Differences between the duration of English and Hungarian vowels for each vowel pair are given in milliseconds as 'Dur' value. V is the English-Hungarian vowel pair ([ɪ]-[i], [æ]-[ɛ], [ɑ]-[ɒ], [ʌ]-[ɒ], and [ʊ]-[u]). The variable V was sum-coded in order to make the main effect of Dur the effect averaging over all vowels: individual vowels do not matter as all vowels were included with natural and manipulated vowel length.
As the formula glmer(Geminate ~ Dur*V+(Dur|Subject), family = binomial(link = 'probit')) did not run with Dur as a continuous variable, the model was run without the random slope for Dur or V, using the following formula:

- glm(Geminate ~ Dur*V +(1|Subject), family = binomial(link = 'probit'))

By plotting the results, it is easier to visualize the effects. Duration differences between English and Hungarian vowels are plotted on the x axis. 0 indicates that there was no difference between the length of certain English and Hungarian vowels. Positive values indicate that the English vowel is longer than the Hungarian one, while negative numbers mean that the English vowel is shorter. These data points were given in milliseconds. The percentages of participants choosing the geminate ('yes') or the singleton ('no') form are on the y axis.

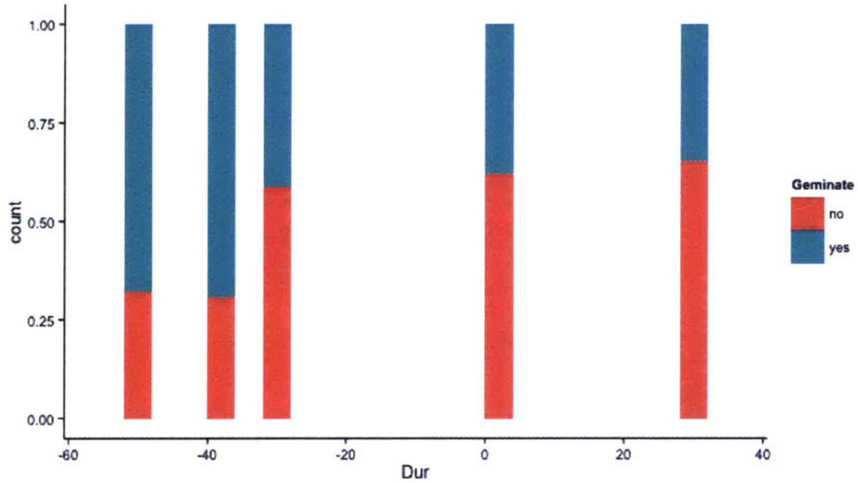

Figure 4-2: Geminate and singleton responses

It is clearly shown by the figure above that the shorter the English vowel is compared to its Hungarian substitute, the more likely native speakers of Hungarian are to respond with the geminated form (thinking that it sounds more similar to the English word). Whenever the English word is of the same length as or longer than the vowel in the Hungarian word, native speakers tend to think that the Hungarian word containing a singleton is a better match.

Fixed effects are shown in the table below. Significant effects indicate that the bigger a difference is between the English and the Hungarian vowel, the more likely listeners are to choose the geminate form as a better match. There is a significant effect on [ʊ]-[u], but as mentioned earlier, all vowels were included with natural and manipulated length so that vowels of all quality have a form in which the English vowel is shorter and another one in which the Hungarian vowel is shorter or of the same length as the English vowel.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.2946108	0.0855676	-3.443	0.000575 ***
Dur	-0.0084421	0.0034503	-2.447	0.014413 *
[I]-[i]	-0.0139499	0.1555623	-0.090	0.928546
[A]-[6]	0.0535704	0.1739002	0.308	0.758042
[U]-[u]	0.3543279	0.1726003	2.053	0.040084 *
[A]-[6]	-0.1361165	0.1946369	-0.699	0.484342
Dur:[I]-[i]	0.0054167	0.0055369	0.978	0.0327935
Dur:[A]-[6]	0.0084421	0.0079270	1.065	0.286883
Dur:[U]-[u]	-0.0024950	0.0064265	-0.388	0.697840
Dur:[A]-[6]	0.0004529	0.0088328	0.051	0.959106

Table 4.4: Fixed effects

4.1.5 Results: Vowel length before voiceless obstruents

Filler items in the experiment can be used to test Hypothesis II-b: the shortening effect of voiceless obstruents on preceding vowels is stronger in English than in Hungarian, therefore, whenever native speakers of Hungarian hear an English word containing a tense vowel followed by a voiceless obstruent, they will perceive it as more similar to a Hungarian word containing the short version of the same vowel than the one with the long vowel.

Geminate and singleton responses for the two vowels included in this part of the experiment ([i] and [u]) are plotted in the following figure. The two different vowel are on the x axis, while percentages of the geminate/ singleton responses are situated on the y axis. As is clearly seen, there is not much difference between the responses in the case of [i] and [u]. For both vowels, the majority of participants chose the Hungarian word with the short vowel as a better match for the English word with the (phonemically) long vowel.

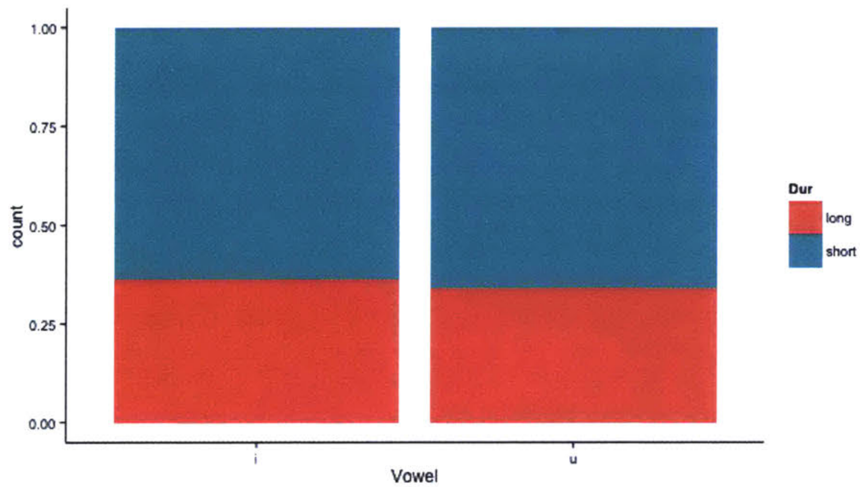

Figure 4-3: Geminate and singleton responses

It is also interesting to see the effect of the individual consonants on preceding vowels. Words in this part of the experiment ended with the following consonants: [p, t, k, ts]. Although this does not cover all the voiceless obstruent and individual differences between different places of articulation might not be relevant for loan gemination, it gives a picture of how they affect the length of the preceding vowel. Vowels are plotted on the x and percentages of 'long' and 'short' responses on the y axis. It is clearly shown by the plot that in the case of most consonants, native speakers of Hungarian were more likely to match the preceding English long vowel to the Hungarian short vowel. This effect is very strong for all consonants except for [t]. For [t], less than 50% of the participants chose the Hungarian word with the short vowel. It is not clear whether in general, there is a smaller difference between English and Hungarian vowels before [t] than before [p], [k] or [ts] or it is merely because of the items produced by only these two speakers.

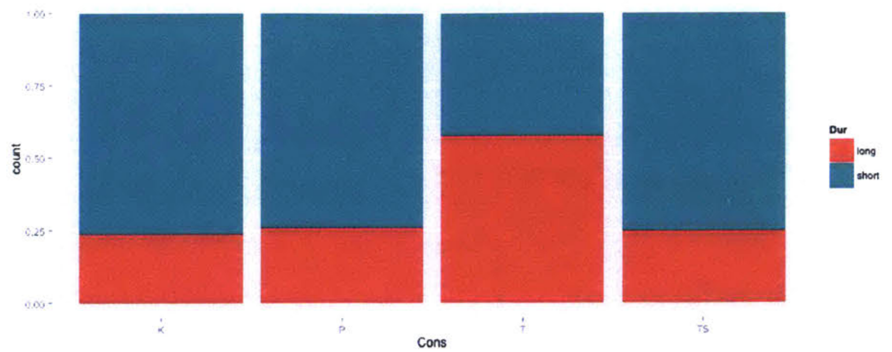

Figure 4-4: Geminate and singleton responses

4.1.6 Summary of results

The goal of the experiment was to test the two sub-hypotheses related to Hypothesis II:

- Hypothesis II-a: If an English vowel is shorter than a Hungarian vowel, Hungarian listeners perceive it as closer in duration to the same vowel preceding a geminate than to the one preceding a singleton.
- Hypothesis II-b: English long vowels preceding voiceless stops and affricates are closer in duration to Hungarian short vowels than long vowels in the same context.

Although it is clearly shown by the tables that lax English vowels preceding voiceless stops tend to be shorter than Hungarian short vowels in the same position, while English tense vowels are usually of the same length as Hungarian short vowels in this position (possibly due to pre voiceless vowel shortening), the goal of the experiment was not to prove that this is the case for only certain vowels and not for others, but to test Hypothesis II-a. It was shown that the shorter the English vowel is, the more likely Hungarian speakers are to perceive it as a closer match to a Hungarian VCC# than a Hungarian VC#. However, it depends on differences in duration between member of English-Hungarian vowel pairs: below a certain threshold, participants chose the singleton form even if the preceding vowel was shorter. Furthermore, we would not expect all participants to choose the geminated form as a better match in any case, as faithfulness to consonant length does matter, even though it is much lower ranked than faithfulness to vowel length. This is reflected in productive loan gemination processes as well: gemination is optional in many cases.

The question of vowel 'shortness' raises another issue: if actual durational differences matter, why do Hungarians borrow consonants as geminates following phonetically long vowels like [æ]? The answer is that Hungarians think in terms of categories and archetypes when they borrow loanwords. There is a perceptual basis for borrowing words containing lax vowels followed by voiceless stops and affricates: most English lax vowels are very short compared to Hungarian short vowels and are even more so when followed by these consonants. Even though [æ] is phonetically long, it is generally considered to be a short vowel and is generally matched to [ɛ] instead of the perceptually closer available match [aː]. This is the case of a source vowel being an archetype instead of an existing vowel: if loan adaptation only had a purely perceptual motivation and listeners would be exposed to the auditory form of loanwords, they would probably borrow [æ] as a long vowel and the following consonant as short.

The second part of the experiment tested and verified Hypothesis II-b: the data have shown that Hungarian long vowels are truly long before voiceless stops and affricates, while English vowels shorten considerably in this environment. Results have shown that this difference is fairly easy to perceive: most participants perceived English tense vowels preceding voiceless stops and affricates as closer to Hungarian short vowels than long vowels preceding the same consonants. This provides further evidence for the strong shortening effect of voiceless stops and affricates in English and the absence of this effect in Hungarian.

The verification of Hypotheses II-a and II-b provides evidence for the claim that gemination can be used as a strategy to preserve the shortness of the source vowel.

Chapter 5

Analysis

It has been shown that gemination in loanwords is influenced by geminate markedness and gradient phonotactic well-formedness patterns of the native Hungarian phonology.

However, faithfulness to source vowel duration increases the probability of gemination for voiceless stops and affricates in word-final position. Intervocalic consonants tend to not geminate in loanwords (unless there is a double letter spelling in the source word or the word consists of a monosyllabic root and a suffix (typically *-er*, which can be explained in two ways: (1) vowels are not significantly shortened by the following intervocalic voiceless stop or affricate in English and intervocalic Hungarian geminates do not shorten the preceding vowel, either (therefore, gemination cannot be used as a strategy to shorten the previous vowel even if the English vowel is shorter than the Hungarian vowel in this context), and (2) faithfulness to consonant duration is more important in intervocalic position, as the variety of cues is richer in this context.

Apart from gradient phonotactic well-formedness and perceptual similarity, the orthography of the donor language is also a crucial factor: since a double letter spelling represents a phonetic geminate in Hungarian (apart from (optional) degemination in certain contexts (Siptár and Törkenczy, 2000) which do not participate in loan gemination), double letter spellings of consonants in the donor language also tend to be borrowed as long consonants.

In this chapter, I first develop a grammar which is based on markedness constraints operating in the current native phonology which well-formedness judgements are based on, and show that this grammar alone is not sufficient to account for productive loan gemination patterns. Then I propose an analysis which is based on well-formedness in the native phonology, augmented with faithfulness constraints used in source-loan mappings, such as faithfulness to both phonetic and orthographic forms in the donor language.

5.1 Native grammar

Contemporary productive loan gemination is heavily influenced by native word structures, some of which had already been present before earlier German and English loans entered the language, but became more conspicuous after the adaptation of these loanwords. Although many of these loanwords were borrowed with geminates for orthographic reasons, they made already existing patterns stronger. The basis for current-day loan gemination is the contemporary Hungarian lexicon containing native words and well-integrated loanwords which are part of everyday language use. The current native phonology can be analyzed using the same constraints that apply for an earlier state of the language, but the proportions of geminates for certain consonant classes have increased.

In the native phonology, long vowels followed by geminates are statistically rare. This tendency is also observed in loan gemination processes: words with long vowels are never borrowed with geminates.

- *V:CC: Long vowels followed by long consonants are marked.

The proportion of geminates compared to singletons after short vowels in monosyllables is larger than in other contexts. This could be the case because there is a smaller number of monosyllables with short vowels than polysyllables. It is also reflected in loanword adaptation: the most common context for loan gemination is word final position in monosyllables, following short vowels.
This is enforced by a word minimality constraint. In Hungarian, there is a word minimality constraint (Siptár and Törkenczy, 2000) which does not allow words consisting of only one vowel or only a consonant and a short vowel. There are a few counterexamples to this generalization, for example, *fa* 'tree', *te* 'you' or *ó* 'oh'. Gemination patterns, however, suggest that there is another, albeit lower ranked minimality constraint which requires the word to be at least two moras. BIMORAIC MINIMALITY will ensure that word-final geminates are allowed in monosyllables in many cases, even though they are cross-linguistically marked.

- BIMORAIC MINIMALITY: The minimal length of a word should be at least two moras.

The extent of gemination also depends on consonant class. Some consonant classes are more likely to be geminated than others. In the native phonology, there is a hierarchy of gemination: voiceless affricates > voiceless stops > voiceless fricatives > voiced stops > nasals > liquids > voiced fricatives. A similar hierarchy can be observed cross-linguistically (even in languages which allow all consonants to be geminated in the native phonology) and in loan gemination processes. This pattern is not very consistent in intervocalic position, where all types of geminates are rare and loan gemination does not take place in this context, unless the consonant was spelled as a double letter in a source word or the word consists of a root + a foreign suffix well-known to Hungarian speakers. This can be captured by using geminate markedness constraints for each consonant class.

- *ZZ: Geminate voiced fricatives are marked.
- *LL: Geminate voiced liquids are marked.
- *NN: Geminate nasals are marked.
- *DD: Geminate voiced stops are marked.
- *SS: Geminate voiceless fricatives are marked.
- *TT: Geminate voiceless stops are marked.
- *TTS: Geminate voiceless affricates are marked.

5.1.1 The model

The categorical ranking of markedness constraints operating in the native phonology would not be able to capture the proportions of the distributions of singletons and geminates. Therefore I am proposing a MaxEnt model using weighted constraints, implemented in the MaxEnt Grammar Tool (Hayes and Wilson, 2008). This is a model which weights constraints based on probabilities of candidates and the sums of violations for each constraint by each candidate. Tableaux with candidates and constraint violations are listed in Appendix C.

Training data

As loan gemination does not depend on the quality of the preceding vowel and the place of articulation of consonants, I use schematic representations of word shapes with only the relevant information for training: vowel length, consonant position, consonant length, consonant manner (class), word type. The following word shapes were used as training data:

- Monosyllables:

- VC (ending in a short vowel + singleton: VT (voiceless stop), VTS (voiceless affricate), VS (voiceless fricative), VD (voiced stop), VN (nasal), VL (liquid), VZ (voiced fricative)
- VCC (ending in a short vowel + geminate: VTT (voiceless stop), VTTS (voiceless affricate), VSS (voiceless fricative), VDD (voiced stop), VNN (nasal), VLL (liquid), VZZ (voiced fricative)
- V:C (ending in a long vowel and a singleton: V:T (voiceless stop), V:TS (voiceless affricate), V:S (voiceless fricative), V:D (voiced stop), V:N (nasal), V:L (liquid), V:Z (voiced fricative))
- V:CC (ending in a long vowel and a geminate: V:TT (voiceless stop), V:TTS (voiceless affricate), V:SS (voiceless fricative), V:DD (voiced stop), V:NN (nasal), V:LL (liquid), V:ZZ (voiced fricative))

- Polysyllables:
  - VCV (intervocalic singleton, preceded by short vowel: VTV (voiceless stop), VTSV (voiceless affricate), VSV (voiceless fricative), VDV (voiced stop), VNV (nasal), VLV (liquid), VZV (voiced fricative))
  - VCCV (intervocalic geminate, preceded by short vowel: VTTV (voiceless stop), VTTSV (voiceless affricate), VSSV (voiceless fricative), VDDV (voiced stop), VNNV (nasal), VLLV (liquid), VZZV (voiced fricative)))
  - V:CV (intervocalic singleton, preceded by long vowel: V:TV (voiceless stop), V:TSV (voiceless affricate), V:SV (voiceless fricative), V:DV (voiced stop), V:NV (nasal), V:LV (liquid), V:ZV (voiced fricative))
  - V:CCV (intervocalic geminate, preceded by long vowel: V:TTV (voiceless stop), V:TTSV (voiceless affricate), V:SSV (voiceless fricative),

V:DDV (voiced stop), V:NNV (nasal), V:LLV (liquid), V:ZZV (voiced fricative))

- .VC# (word-final singleton, preceded by short vowel: .VT# (voiceless stop), .VTS# (voiceless affricate), .VS# (voiceless fricative), .VD# (voiced stop), .VN# (nasal), .VL# (liquid), .VZ# (voiced fricative))

- .VCC# (word-final geminate, preceded by short vowel: .VTT# (voiceless stop), .VTTS# (voiceless affricate), .VSS# (voiceless fricative), .VDD# (voiced stop), .VNN# (nasal), .VLL# (liquid), .VZZ# (voiced fricative))

- .V:C# (word-final singleton, preceded by a long vowel: .V:T# (voiceless stop), .V:TS# (voiceless affricate), .V:S# (voiceless fricative), .V:D# (voiced stop), .V:N# (nasal), .V:L# (liquid), .V:Z# (voiced fricative))

- .V:CC# (word-final geminate, preceded by a long vowel: .V:TT# (voiceless stop), .V:TTS# (voiceless affricate), .V:SS# (voiceless fricative), .V:DD# (voiced stop), .V:NN# (nasal), .V:LL# (liquid), .V:ZZ# (voiced fricative))

Results

Observed probabilities were assigned to each form based on their distribution in the native lexicon. The weight of each constraint was calculated based on these probabilities and the number of violations each candidate incurred for each constraint:

Constraint	Weight
*ZZ	3.64
*V:CC	3.33
*DD	2.51
*NN	2.48
*LL	2.18
BIMORAIC MINIMALITY	1.61
*SS	1.53
*TT	1.34
*TTS	1.1

Table 5.1: Native grammar: constraints and weights

The constraint penalizing geminate voiced fricatives received the highest weight, as such geminates occur very rarely in the native Hungarian phonology as well as cross-linguistically. Constraints penalizing long vowel + geminate sequences and other markedness constraints against geminate voiced stops, liquids and nasals were also assigned higher weight than BIMORAIC MINIMALITY, which is not surprising: geminates hardly ever occur following long vowels, and voiceless consonants are rarely geminated. The three consonant classes which occur the most frequently as geminates, were assigned lower weights than BIMORAIC MINIMALITY.

To measure the amount of concordant pairs between observed (based on corpus frequencies) and predicted (assigned by the grammar) probabilities, Kendall's tau was used. As shown by the following scatter plot, there is a fairly good but not perfect match between the observed and predicted probabilities.

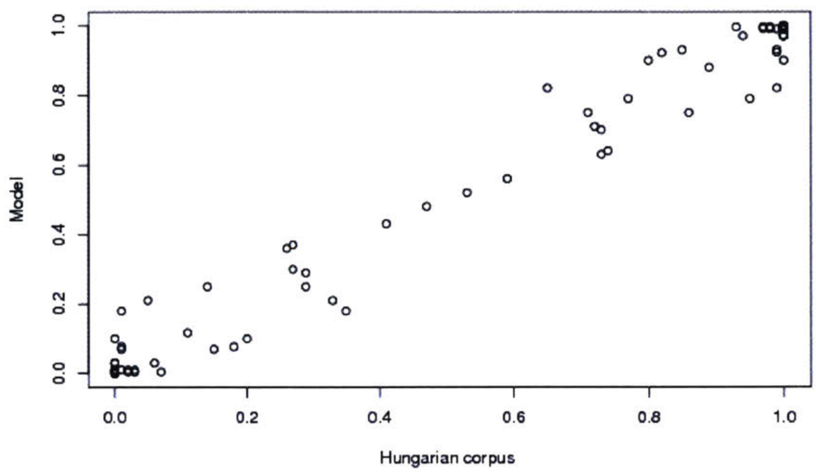

Figure 5-1: Distribution of geminates in the Hungarian phonology and probabilities assigned by the model (Kendall's Tau=0.8)

Figures 5-2, 5-3 and 5-4 are showing the distribution of word-final geminates in monosyllables (Figure 5-2), intervocalic geminates in disyllables (Figure 5-3), and word-final geminates in polysyllables (Figure 5-4).

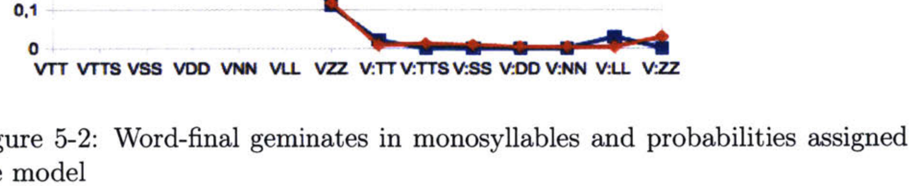

Figure 5-2: Word-final geminates in monosyllables and probabilities assigned by the model

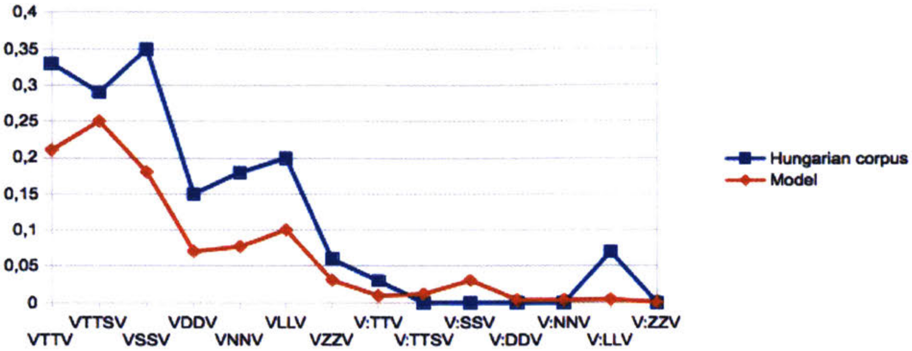

Figure 5-3: Intervocalic geminates in disyllables and probabilities assigned by the model

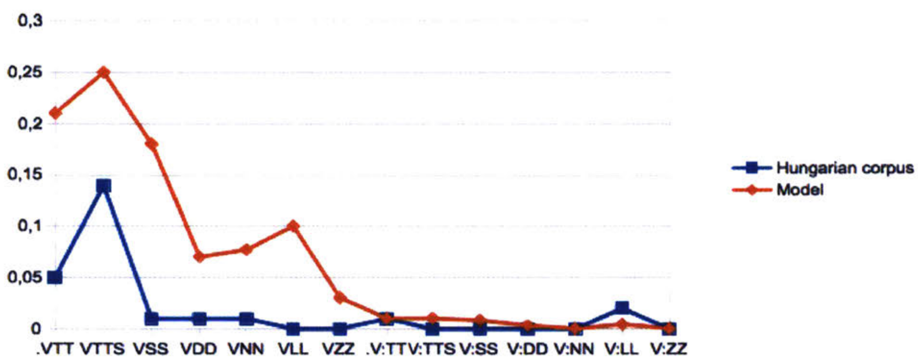

Figure 5-4: Word-final geminates in polysyllables and probabilities assigned by the model

The fit for monosyllables seems better because of the range of distributions: the range is smaller in the case of intervocalic geminates and word-final geminates in polysyllables, therefore differences between corpus frequencies and the model seem bigger. The model generally assigned lower probabilities to geminates in intervocalic position because in these cases BIMORAIC MINIMALITY does not ensure high rates of gemination. The reason why the fit is not perfect for liquids in either position is that its distribution across different positions in the corpus is not uniform:

it has a high number of geminate forms in intervocalic position after long vowels but does not geminate often in word-final position compared to other consonants.

Fit with well-formedness judgement data

The nonce word well-formedness data described in Chapter 3 can be based on the constraints learned from the current Hungarian lexicon, which contains several well-integrated loanwords along with native Hungarian words. Therefore, it is important to see how close a fit the model is to probabilities established based on well-formedness judgement data.

The following plot shows the distribution of word-final and intervocalic geminates following short vowels both based on well-formedness data and predicted by the model.

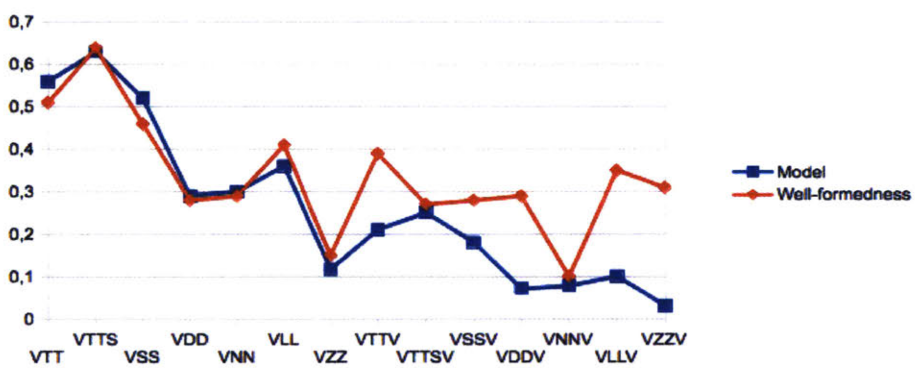

Figure 5-5: Fit with well-formedness judgement data

The fit with monosyllabic geminates is nearly perfect, but predictions of the model are fairly inconsistent with well-formedness judgements on consonants in intervocalic position (Kendall's Tau=0.5). The fact that gemination is rare in intervocalic position and that geminate markedness hierarchies are not very clear-cut in this context is reflected in well-formedness judgements.

5.2 Loanword grammar

As we have seen, it is possible to approximately predict gemination from the lexicon with weighted markedness constraints. The well-formedness judgement data also show that it is easier to learn word-final gemination in monosyllables than intervocalic gemination or word-final gemination in polysyllables.

However, we run into problems when trying to predict productive loan gemination only based on markedness constraints operating in the native phonology. The grammar is not a good fit with productive loan gemination for various reasons. For example, it cannot account for the lack of gemination in intervocalic position (when there is no orthographic influence and the word is not a monosyllabic root + suffix), it cannot predict when there is final gemination in polysyllables or when there is not, and it cannot explain why there is an even higher amount of gemination of voiceless stops and affricates. The following figure shows gemination without the influence of orthography compared to the native grammar.

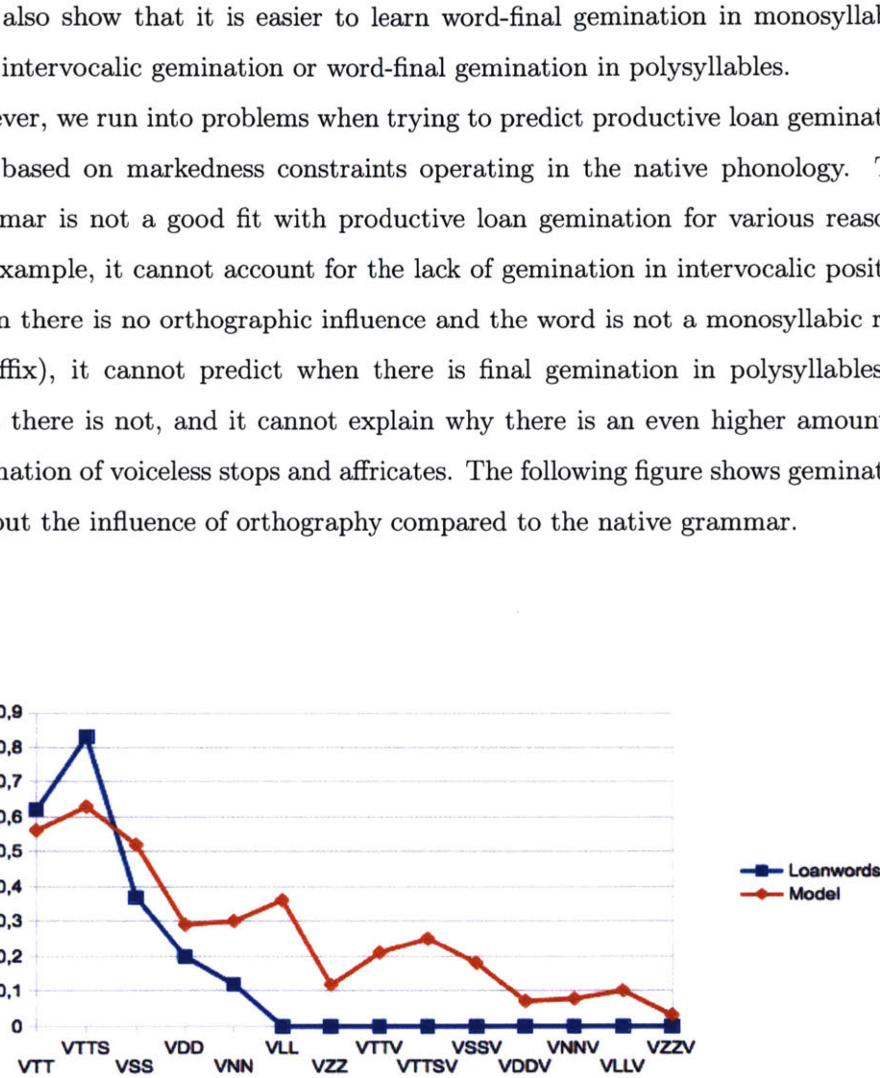

Figure 5-6: Productive loan gemination and predicted probabilities (without the influence of orthography)

5.2.1 Summary of loanword data

As described in previous sections, loan gemination occurs in the following contexts. The probability of gemination was calculated from native speaker judgements on pseudo loanwords for monosyllabic loanwords. (The task and the data are described in Chapter 2.)

1. Word-finally in monosyllables, after short vowels:

    - Voiceless stops: e.g. *fitt* 'fit' (0.62)
    - Voiceless affricates: e.g. *meccs* 'match' (0.83)
    - Voiceless fricatives: e.g. *GIF* % [gif:] (0.37)
    - Voiced stops: e.g. *blog* % [blog:] (0.2)
    - Nasals: e.g. *dzsem / jam* 'jam' % [d3Em:] (0.12)
    - Liquids: there are not many examples without the orthographic factors, only names like *Hal* <diminutive of *Harry*> (0)
    - Voiced fricatives: there are not many examples, e.g. *jazz*, *fizz*, *buzz* etc. (there is no gemination without the influence of spelling) (0)
    - All consonants are pronounced as long when there is a double spelling in the source word. The only exception is voiceless fricatives: their gemination is optional even in the case of double spelling in the source word.

2. Intervocalically in disyllables:

    - It is hard to tell whether it is a context for productive gemination without the influence of orthography, as intervocalic consonants which are preceded by short vowels are usually spelled with a double letter or a digraph in English. There is no evidence that digraphs trigger gemination: Hungarian has digraphs which are pronounced as singletons. The only context where gemination productively applies in this context is words which consist of a monosyllabic word + a foreign suffix widely known to native speakers of Hungarian (e.g. *rocker* pronounced as ɾokːɛɾ).

3. Word-finally in polysyllables:

   - There are hardly any examples without orthographic influence, and all of the examples contain voiceless stops and liquids. When the final letter is spelled with a double letter in the source word, it is geminated in the loanwords (e.g. *toalett* 'toilette' (from French)).

5.2.2 The model

The goal of this model is to account for the influence of the three main factors on loan gemination processes: gradient phonotactic well-formedness in the native Hungarian grammar, faithfulness to vowel duration (driven by perceptual similarity), and orthography. The new model is trained using constraints with fixed weights (from the previously described grammar trained on the Hungarian corpus) and new constraints which play an important role in loan adaptation processes.

Constraints

Loan gemination is not only influenced by markedness constraints from the native phonology, it is also affected by faithfulness to source vowel duration: loan gemination applies most heavily to consonants which trigger shortening of the previous vowel (that is, voiceless stops and affricates). If gemination is a strategy to shorten the preceding vowel, we would expect gemination to apply in word-final position in polysyllables more often than in monosyllables, since there is no evidence of vowel shortening in polysyllables in Hungarian, unlike in English, which means that the difference between vowel durations in this context would be even bigger in the case of word-final unstressed vowels. These examples are very rare in current loanwords (most of the examples are older and from French), but native speakers' intuition is that if the vowel was stressed in the source word or the consonant was spelled with a double letter, the consonant will be geminated in the loanword. However, if the vowel was unstressed and the consonant was spelled with a single letter in the source word, gemination does not take place. There are two constraints enforcing faithfulness to source vowel duration:

- IDENTVDUR(STR): Vowel duration in the loanword must be identical to vowel duration of the stressed vowel in the loanword.
- IDENTVDUR(UNSTR): Vowel duration in the loanword must be identical to vowel duration of unstressed vowels in the source word.

IDENTVDUR(STR) is ranked higher than IDENTVDUR(UNSTR). These constraints are evaluated using three levels of schematic vowel categories: long V, short V and extra short V. These categories are debatable and there are more fine-grained categories of vowel length (for example, English [ɪ] is inherently shorter than Hungarian [i]). Such fine grained distinctions would be used if there was a vowel quality effect involved in loan gemination (e.g. gemination of consonants following source vowel [æ] would be less preferred than geminates following source vowel [i]), but there is no evidence for such an effect based on the intuition elicitation.

Apart from faithfulness to vowel duration, faithfulness to consonant length is also an important factor contributing to dispreference for gemination in intervocalic position. As more cues are available to consonant duration in intervocalic position, faithfulness to intervocalic consonant length is a higher ranked constraint than in word-final position.

- IDENTVCV(L): Intervocalic short consonants in the source word must correspond to short consonant in the loanword.
- IDENTVC#(L): Word-final short consonants in the source word must correspond to short consonant in the loanword.

Apart from faithfulness to source vowel and consonant duration, faithfulness to source vowel orthography also plays a significant role in borrowing consonants as geminates. Whenever a consonant is spelled with a double letter in the source word, it tends to be borrowed as a geminate. This is enforced by the following constraint:

- MAXORTHGEM: A double consonant spelling in the source word must be represented as a double consonant in the loanword.

Training data

Just like in the case of Hungarian words, templatic forms were tested which are not specified for vowel quality and place of articulation for the consonant, as those factors do not appear to play a role in loan gemination processes. Underlying forms were specified as orthography of the source word (in < > brackets) and the phonetic form of the source word.

- Monosyllables:
  - *FITT-type words*, Input: <c> VC (ending in a short vowel or extra short vowel (depending on the following consonant) + singleton (spelled with a single letter): VT (voiceless stop), VTS (voiceless affricate), VS (voiceless fricative), VD (voiced stop), VN (nasal), VL (liquid), VZ (voiced fricative))

        * Candidate a: VC (with a short vowel and singleton)
        * Candidate b: VCC (with an extra short vowel and a geminate)

    - *NETT-type words*, Input: <cc> VC (ending in a short vowel or extra short vowel (depending on the following consonant) + singleton (spelled with a double letter): VT (voiceless stop), VTS (voiceless affricate), VS (voiceless fricative), VD (voiced stop), VN (nasal), VL (liquid), VZ (voiced fricative))
        * Candidate a: VC (with a short vowel and a singleton)
        * Candidate b: VCC (with an extra short vowel and a geminate)

- Polysyllables:
    - *HOKI-type words*, Input: <c> VCV (intervocalic singleton (spelled with a single letter), preceded by short vowel: VTV (voiceless stop), VTSV (voiceless affricate), VSV (voiceless fricative), VDV (voiced stop), VNV (nasal), VLV (liquid), VZV (voiced fricative))
        * Candidate a: VCV (with a short vowel and a singleton)
        * Candidate b: VCCV (with an extra short vowel and a geminate)
    - *LOBBI-type words*, Input: <cc> VCV (intervocalic singleton (spelled with a double letter), preceded by short vowel: VTV (voiceless stop), VTSV (voiceless affricate), VSV (voiceless fricative), VDV (voiced stop), VNV (nasal), VLV (liquid), VZV (voiced fricative))
        * Candidate a: VCV (with a short vowel and a singleton)
        * Candidate b: VCCV (with an extra short vowel and a geminate)
    - *BAROKK-type words*, Input: <c> .'VC# (word-final singleton (spelled with a single letter), preceded by short stressed vowel: .VT# (voice-

less stop), .VTS# (voiceless affricate), .VS# (voiceless fricative), .VD# (voiced stop), .VN# (nasal), .VL# (liquid), .VZ# (voiced fricative))

    * Candidate a: .VC# (with a short vowel and a singleton)
    * Candidate b: .VCC# (with an extra short vowel and a geminate)

- *TOALETT-type words*, Input: <cc> .'VC# (word-final singleton (spelled with a double letter), preceded by short stressed vowel: .VT# (voiceless stop), .VTS# (voiceless affricate), .VS# (voiceless fricative), .VD# (voiced stop), .VN# (nasal), .VL# (liquid), .VZ# (voiced fricative))
    * Candidate a: .VC# (with a short vowel and a singleton)
    * Candidate b: .VCC# (with an extra short vowel and a geminate)
- *PROFIT-type words*, Input:.VC# (word-final geminate, preceded by short unstressed vowel: .VTT# (voiceless stop), .VTTS# (voiceless affricate), .VSS# (voiceless fricative), .VDD# (voiced stop), .VNN# (nasal), .VLL# (liquid), .VZZ# (voiced fricative))
- Candidate a: .VC# (with a short vowel and a singleton)
- Candidate b: .VC# (with an extra short vowel and a geminate)

The training data did not contain long vowels, as words containing long vowels are never borrowed with a geminate. However, the adaptation of long vowels raises the question why phonemically long and phonetically short English vowels (due to pre-voiceless vowel shortening) are borrowed as long, if faithfulness to source vowel duration is a highly ranked constraint. The answer is that faithfulness to spelling also influences the adaptation of vowels: long monophthongs are generally spelled with a double vowel letter in English, which is automatically borrowed as a long vowel into Hungarian.

5.2.3 Results

Weights were retained in the case of constraints which were part of the native grammar: *ZZ, *V:CC, *DD, *NN, *LL, BIMORAIC MINIMALITY, *SS, *TT, and *TTS. New weights were assigned to constraints which play a role in loanword adaptation processes: IDENTVDUR(STR), IDENTVDUR(UNSTR), IDENTVDUR(STR), IDENTVCV(L), IDENTVC#(L), MAXORTHGEM. Constraints with weights are shown in the following table:

Constraint	Weight
MAXORTHGEM	19.9
IDENTVCV(L)	7.64
*ZZ	3.64
*V:CC	3.33
*DD	2.51
*NN	2.48
IDENTVDUR(STR)	2.36
*LL	2.18
BIMORAIC MINIMALITY	1.61
*SS	1.53
*TT	1.34
IDENTVC#(L)	1.25
*TTS	1.1
IDENTVDUR(UNSTR)	0

Table 5.2: Native grammar: constraints and weights

Since orthographic geminates are borrowed as phonetic geminates, MAX-ORTHGEM received the highest weight. IDENTVCV(L) was assigned the second highest weight: consonants in intervocalic position tend to not geminate, as cues to consonant duration are richer in intervocalic position, therefore faithfulness to source consonant duration in this position is much more important than in word-final position, where there is a higher chance of confusability (which is why IDENTVC#(L) was assigned a much lower weight). IDENTVDUR(STR) was assigned a weight, while IDENTVDUR(UNSTR): this ensures that even if there is a large vowel length difference between English and Hungarian vowels in final syllables of polysyllabic words (as there is no evidence for polysyllabic shortening in Hungarian (White and Mády, 2008)), final consonants do not geminate unless they were preceded by a stressed vowel in the source word.

In general, the model is a good fit with the loanword data, as shown by the following scatterplot.

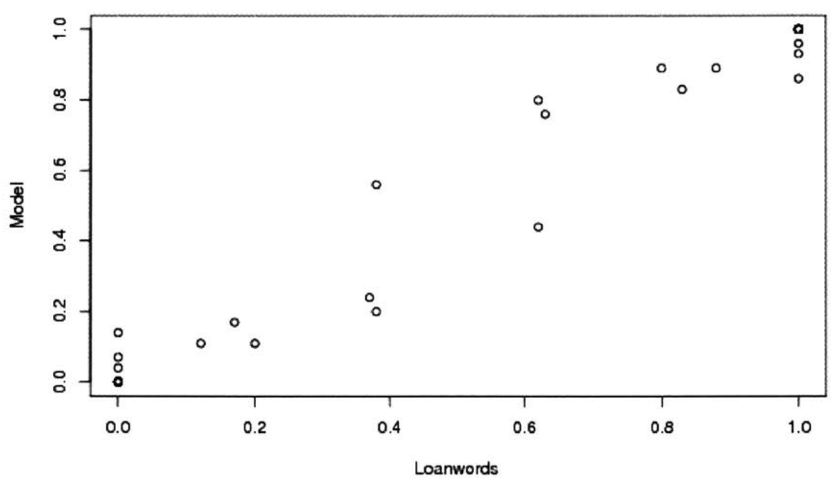

Figure 5-7: Productive loan gemination and probabilities assigned by the model (Kendall's Tau=0.93)

Kendall's Tau is 0.93, which means that there is a high amount of concordant pairs between productive loan gemination (calculated based on native speaker choices in the pseudo loanword adaptation task described in Chapter 2) and the probabilities assigned by the model.

Final gemination in monosyllabic loanwords and its fit with probabilities assigned by the model is shown in Figure 5-8. Results for words without a double letter spelling are shown on the left and results for words with a double letter spelling are shown on the right.

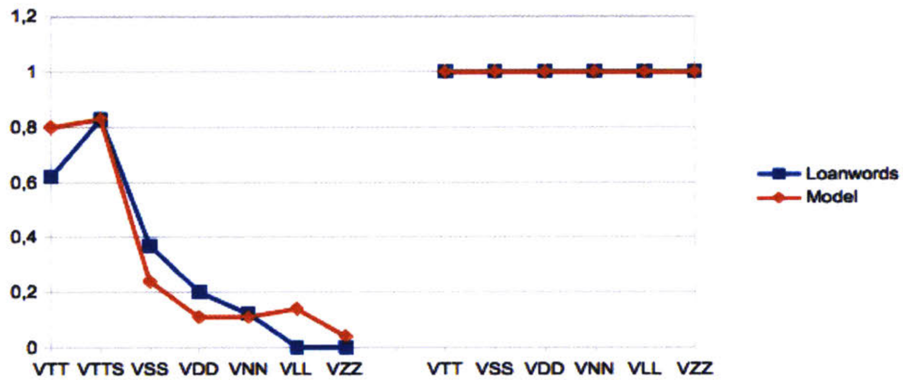

Figure 5-8: Final gemination in monosyllabic loanwords and probabilities assigned by the model

The model is a perfect fit for gemination in words with double letter spelling, due to the high weight of MAXORTHGEM. The fit with words which are not spelled with a double letter in the source word is less perfect: hierarchies of gemination are predicted well, but the model does not make a big difference between rates of gemination for voiceless stops and fricatives, even though voiceless affricates are more frequently geminated both in loanwords and in the native phonology than voiceless stops. IDENTVDUR(STR) would enforce a similar rate of gemination for both of these consonant classes, but geminate markedness constraints learned based on the native phonology would differentiate between them. However, markedness constraints like TTS and TT should received a higher weight to create a bigger difference between the rates of gemination for voiceless affricates and voiceless stops.

The fit is perfect in the case of intervocalic geminates in disyllabic words, as shown in Figure 5-9.

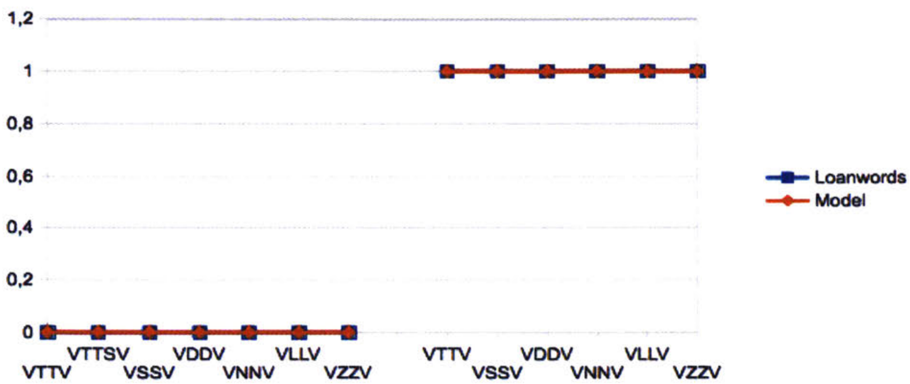

Figure 5-9: Intervocalic gemination in disyllabic loanwords and probabilities assigned by the model

Intervocalic consonants are not geminated when they were not spelled with a double letter in the source word (on the left side of the figure) and are geminated when there was a double letter in the spelling (on the right side of the figure). This is due to the high weight of MAXORTHGEM and IDENTVCV(L).

The fit is nearly perfect in the case of word-final geminates in polysyllables. On the left side of Figure 5-10, word-final consonants (without a double letter spelling) preceded by unstressed and stressed vowels are shown. On the right side of the same figure, the same type of consonants are shown which were spelled with a double letter in the source word.

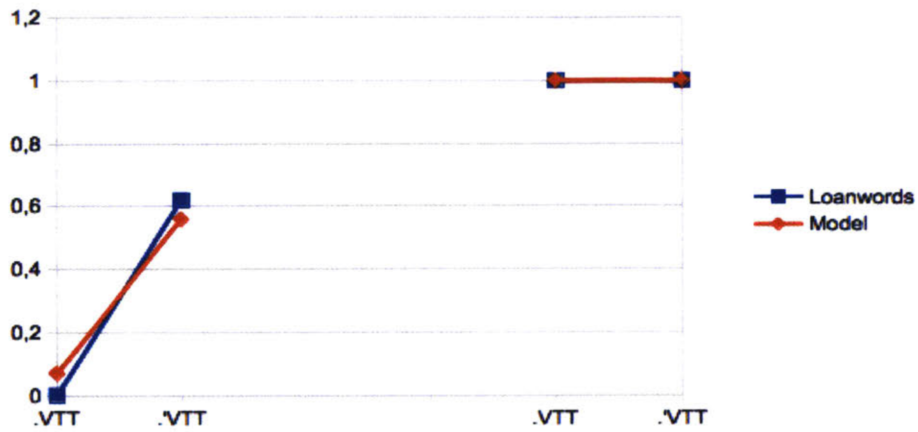

Figure 5-10: Final gemination in monosyllabic loanwords and probabilities assigned by the model

As is shown in the figure, source word stress and spelling are important factors: consonants preceded by stressed vowels are frequently geminated, those spelled with a double letter are always geminated, while those not spelled with a double letter or preceded by an unstressed vowel are not geminated.

In general, this grammar can predict the preferences for geminates and singletons. The reason why the grammar does not mirror native speaker intuitions completely is that it has the weights for markedness constraints from the native phonology, trained both on monosyllables and polysyllables. The very distinct markedness hierarchy of gemination we see in monosyllables is slightly different a polysyllables. What both hierarchies have in common is that voiceless consonants are more likely to be geminated than voiced ones.

Native speaker intuitions on loanwords are based on word shapes and well-formedness in the native lexicon instead of (cross-linguistically attested) markedness per se. Gemination hierarchies in monosyllabic loanwords are based on the well-formedness of native words (and older loanwords) rather than following the same universal markedness hierarchy for all word shapes. To get a perfect match between the observed and predicted probabilities, monosyllables should be trained on monosyllables and polysyllables on polysyllables. However, language specific geminate well-formedness and universal markedness are not mutually exclusive and often correspond very closely to each other, as we have seen in the corpus studies on other languages and Hungarian as well.

Chapter 6

Conclusions

In this thesis, I presented a detailed case study of gemination in loanwords, a cross-linguistically widespread phenomenon. I showed that loan gemination in Hungarian is motivated by two main forces: gradient phonotactic well-formedness (geminate markedness) and perceptual similarity (faithfulness to source vowel duration). It is heavily influenced by native phonotactics, but this alone is not sufficient to account for the full range of gemination patterns in loanwords.

Faithfulness to vowel duration is a possible explanation for why voiceless affricates and stops are geminated the most in loanwords. While in English voiceless stops and affricates shorten the previous vowel in closed syllables, the geminated versions of these consonants have a similar effect in Hungarian. Therefore, gemination can be used as a strategy to preserve the shortness of the source (English) vowel compared to the duration of the Hungarian vowel which is used as a substitute in loanword adaptation processes.

Faithfulness to intervocalic consonant duration is also an important factor. There is a richer variety of cues to vowel duration in intervocalic position, which means that it is easier for the listener to differentiate between long and short consonants, whereas there is a greater degree of confusability in word-final position. This can serve as an explanation to why gemination in loanwords is common in word-final and extremely rare in intervocalic position.

Apart from gradient phonotactic well-formedness and faithfulness to vowel and consonant duration, orthography also plays a major role in loan gemination processes: if a consonant is spelled with a double letter in the source word, it will most probably be borrowed as a phonetic geminate.

Appendix A

Chapter 2

A.1 Loanword corpus

This is a (non-exhaustive) list of loanwords which are used in Hungarian. Each word has the following information: origin. meaning. type. source word spelling. consonant position. consonant class and the possibility of gemination. 'Optional' means that both singleton and geminate forms are equally acceptable. there is no preference for one or the other. 'Optional, preferred' means that the geminated form is more standard or widely used. 'Optional. dispreferred' means that although some speakers pronounce a word with a geminate. this form is not considered to be standard or widely used.

Abbreviations:

C Class: Consonant Class
C Pos: Consonant Position
Disyll: disyllabic
E: English

F: French

G: German

Gr: Greek

Monosyll: monosyllabic

SW: source word

VC#: consonant in word-final. preceded by a vowel

VCV: consonant in intervocalic position

The list was extracted from the Hungarian Webcorpus (Halácsy et al. (2014)) and supplemented with items from Nádasdy. The items that were brought to my attention as instances of loan gemination by Nádasdy (1989) are indicated below.

Word	Origin	Meaning	Type	SW spelling	C Pos	C Class	Gemination
tipp	E	idea	monosyll	single	VC#	T	yes
chip	E	chip	monosyll	single	VC#	T	optional preferred
klip	E	videoclip	monosyll	single	VC#	T	yes
sokk (Nádasdy)	F	shock	monosyll	digraph	VC#	T	yes
rock	E	rock music	monosyll	digraph	VC#	T	optional preferred
vicc	G	joke	monosyll	digraph	VC#	TS	yes
giccs (Nádasdy)	G	kitsch	monosyll	trigraph	VC#	TS	yes
sitt	G	debris	monosyll	double	VC#	T	yes
blokk	G	block	monosyll	digraph	VC#	T	yes
szett	E	outfit	monosyll	single	VC#	T	yes
sikk	G	chic	monosyll	single	VC#	T	yes
hecc	G	joke	monosyll	digraph	VC#	TS	yes
puccs	G	coup	monosyll	trigraph	VC#	TS	yes
plusz	L	plus	monosyll	single	VC#	S	optional preferred
procc (Nádasdy)	G	upstart	monosyll	digraph	VC#	TS	yes
plüss	G	plush	monosyll	digraph	VC#	S	yes
snassz	G	average	monosyll	double	VC#	S	yes
szmog	E	smog	monosyll	single	VC#	D	optional dispreferred
drog	E	drug	monosyll	single	VC#	D	no
blog	E	blog	monosyll	single	VC#	D	optional dispreferred
sznob	E	snob	monosyll	single	VC#	D	optional dispreferred
klub	G	club	monosyll	single	VC#	D	optional preferred
dog	E	dog	monosyll	single	VC#	D	optional dispreferred
bob	E	bob	monosyll	single	VC#	D	optional dispreferred

Word	Origin	Meaning	Type	SW spelling	C Pos	C class	Gemination
meccs	E	match	monosyll	digraph	VC#	TS	yes
taccs (Nádasdy)	E	touch	monosyll	digraph	VC#	TS	yes
treff	G	clubs	monosyll	digraph	VC#	S	yes
dzsessz (Nádasdy)	E	jazz	monosyll	double	VC#	S	yes
jazz	E	jazz	monosyll	double	VC#	Z	optional
gin	E	gin	monosyll	single	VC#	N	no
dzsinn	G	genie	monosyll	single	VC#	N	yes
dzsem (Nádasdy)	E	jam	monosyll	single	VC#	N	optional dispreferred
stramm	G	strong	monosyll	double	VC#	N	optional preferred
gramm	Gr	gram	monosyll	single	VC#	N	optional preferred
slam (Nádasdy)	E	slam	monosyll	single	VC#	N	optional preferred
brit	E	Brit	monosyll	single	VC#	T	optional
shop	E	shop	monosyll	single	VC#	T	optional
kit (Nádasdy)	E	kit	monosyll	single	VC#	T	optional
top	E	top	monosyll	single	VC#	T	optional
stop	E	stop	monosyll	single	VC#	T	optional
rap	E	rap	monosyll	single	VC#	T	optional
hall	G	hall	monosyll	double	VC#	L	yes
kuss (Nádasdy)	G (from F)	shut up	monosyll	digraph	VC#	S	yes
brill	G	diamond	monosyll	double	VC#	L	yes
tüll	G	tulle	monosyll	double	VC#	L	yes
bit	E	bit	monosyll	single	VC#	T	optional
hit	E	hit song	monosyll	single	VC#	T	no
back	E	back	monosyll	digraph	VC#	T	yes
pop	E	pop music	monosyll	single	VC#	T	optional preferred
nett	G	neat	monosyll	double	VC#	T	yes
chat	E	chat	monosyll	single	VC#	T	optional dispreferred
flott	G	fast and easy	monosyll	double	VC#	T	yes
fitt	E	fit	monosyll	single	VC#	T	yes
friss	G	fresh	monosyll	trigraph	VC#	S	yes
gif	E	gif	monosyll	single	VC#	Z	optional
dokk	E	dock	monosyll	digraph	VC#	T	yes
snitt	G	cut	monosyll	double	VC#	T	yes
slepp	G	entourage	monosyll	double	VC#	T	yes
fess	G	handsome	monosyll	trigraph	VC#	S	yes
klassz	G	cool. good	monosyll	double	VC#	S	yes
blikk	G	wink	monosyll	digraph	VC#	T	yes
necc (Nádasdy)	G	fishnet stockings	monosyll	trigraph	VC#	TS	yes

Word	Origin	Meaning	Type	SW spelling	C Pos	C class	Gemination
nipp	G	nip	monosyll	single	VC#	T	yes
hepp	G	hobby horse	monosyll	double	VC#	T	yes
mop	E	mop	monosyll	single	VC#	T	optional
skicc	G	sketch	monosyll	digraph	VC#	TS	yes
szkeccs	E	sketch	monosyll	trigraph	VC#	TS	yes
scotch	E	Scotch whiskey	monosyll	trigraph	VC#	TS	yes
pucc	G	posh stuff	monosyll	digraph	VC#	TS	yes
csekk	F	check	monosyll	digraph	VC#	T	yes
spicc	G	tipsiness	monosyll	digraph	VC#	TS	yes
poll	E	poll	monosyll	double	VC#	L	yes
null	L	null	monosyll	double	VC#	L	yes
szvetter	E	sweater	disyll	single	VCV	T	yes
Betty	E	name	disyll	double	VCV	T	yes
petting	E	petting	disyll	double	VCV	T	yes
hippi	E	hippie	disyll	double	VCV	voicekess stop	optinal preferred
stopper	G	stopwatch	disyll	double	VCV	T	yes
floppy	E	floppy disc	disyll	double	VCV	T	optional
shopping	E	shopping	disyll	double	VCV	T	yes
szetter	E	setter	disyll	double	VCV	T	yes
mutter	G	mother	disyll	double	VCV	T	optional
fater	G	father	disyll	single	VCV	T	no
vekker	G	alarm clock	disyll	digraph	VCV	T	yes
rocker	E	rocker	disyll	digraph	VCV	T	yes
cekker	G	shopping bag	disyll	digraph	VCV	T	yes
prakker (Nádasdy)	G	carpet beater	disyll	double	VCV	T	yes
ziccer	G	good catch	disyll	digraph	VCV	TS	yes
koffer	G	baggage	disyll	double	VCV	S	yes
desszert	G (from F)	dessert	disyll	double	VCV	S	yes
smasszer	G	prison guard	disyll	double	VCV	S	yes
esszé	E (from F)	essay	disyll	double	VCV	S	yes
hobbi	E	hobby	disyll	double	VCV	D	yes
rabbi	Hebrew	rabbi	disyll	double	VCV	D	yes
Abba (Nádasdy)	name of a pop	group	disyll	double	VCV	D	yes
jiddis	G	Yiddish	disyll	double	VCV	D	yes
shimmy (Nádasdy)	E	shimmy	disyll	double	VCV	N	yes
dollár	E	dollar	disyll	double	VCV	L	yes
roller	E	roller	disyll	double	VCV	L	yes
kollázs (Nádasdy)	F	collage	disyll	double	VCV	L	yes
passzázs (Nádasdy)	F	passage	disyll	double	VCV	S	yes
szuper	L	super	disyll	double	VCV	T	no
liter	G	liter	disyll	single	VCV	T	no
Snickers	E	chocolate bar	disyll	digraph	VCV	T	no
lobbi	E	lobby	disyll	double	VCV	D	yes
hoki	E	hockey	disyll	digraph	VCV	T	no
hacker	E	hacker	disyll	digraph	VCV	T	yes
chopper	E	chopper	disyll	double	VCV	T	yes
galopp (Nádasdy)	G	gallop	disyll	double	VC#	T	yes
szonett (Nádasdy)	G (from F)	sonnet	disyll	double	VC#	T	yes
krikett	E	cricket	disyll	single	VC#	T	yes
balett	G	ballet	disyll	double	VC#	T	yes
toalett	F	toilet	disyll	double	VC#	T	yes

Word	Origin	Meaning	Type	SW spelling	C Pos	C class	Gemination
kvintett	Italian	quintet	disyll	single	VC#	T	yes
kokott (Nádasdy)	F	cocotte	disyll	double	VC#	T	yes
barakk	G (from F)	barrack	disyll	digraph	VC#	T	yes
barokk (Nádasdy)	F	baroque	disyll	digraph	VC#	T	yes
modell (Nádasdy)	G	model	disyll	double	VC#	L	yes
kartell (Nádasdy)	G	kartel	disyll	double	VC#	L	yes
hotel	G	hotel	disyll	single	VC#	L	no
panel	G	panel	disyll	single	VC#	L	no

A.2 Loanword adaptation task: responses

Subject	Word	C Class	OrthGem	Geminate
Subject1	batch	TS	no	yes
Subject2	batch	TS	no	yes
Subject3	batch	TS	no	yes
Subject3	batch	TS	no	no
Subject4	batch	TS	no	yes
Subject5	bacth	TS	no	yes
Subject1	bitch	TS	no	yes
Subject2	bitch	TS	no	yes
Subject3	bitch	TS	no	yes
Subject3	bitch	TS	no	no
Subject4	bitch	TS	no	yes
Subject5	bitch	TS	no	yes
Subject1	betch	TS	no	yes
Subject2	betch	TS	no	yes
Subject3	betch	TS	no	yes
Subject3	betch	TS	no	no
Subject4	betch	TS	no	yes
Subject5	betch	TS	no	yes
Subject1	butch	TS	no	yes
Subject2	butch	TS	no	yes
Subject3	butch	TS	no	yes
Subject3	butch	TS	no	no
Subject4	butch	TS	no	yes

Subject	Word	C Class	OrthGem	Geminate
Subject5	butch	TS	no	yes
Subject1	blatch	TS	no	yes
Subject2	blatch	TS	no	yes
Subject3	blatch	TS	no	yes
Subject4	blatch	TS	no	yes
Subject5	blatch	TS	no	yes
Subject1	blitch	TS	no	yes
Subject2	blitch	TS	no	yes
Subject3	blitch	TS	no	yes
Subject4	blitch	TS	no	yes
Subject5	blitch	TS	no	yes
Subject1	bletch	TS	no	yes
Subject2	bletch	TS	no	yes
Subject3	bletch	TS	no	yes
Subject4	bletch	TS	no	yes
Subject5	bletch	TS	no	yes
Subject1	blutch	TS	no	yes
Subject2	blutch	TS	no	yes
Subject3	blutch	TS	no	yes
Subject4	blutch	TS	no	yes
Subject5	blutch	TS	no	yes
Subject1	bats	TS	no	yes

Subject	Word	C Class	OrthGem	Geminate
Subject2	bats	TS	no	yes
Subject3	bats	TS	no	yes
Subject3	bats	TS	no	no
Subject4	bats	TS	no	yes
Subject5	bats	TS	no	yes
Subject1	bets	TS	no	yes
Subject2	bets	TS	no	yes
Subject3	bets	TS	no	yes
Subject3	bets	TS	no	no
Subject4	bets	TS	no	yes
Subject5	bets	TS	no	yes
Subject1	bits	TS	no	yes
Subject2	bits	TS	no	yes
Subject3	bits	TS	no	yes
Subject3	bits	TS	no	no
Subject4	bits	TS	no	yes
Subject5	bits	TS	no	yes
Subject1	buts	TS	no	yes
Subject2	buts	TS	no	yes
Subject3	buts	TS	no	yes
Subject3	buts	TS	no	no
Subject4	buts	TS	no	yes

Subject	Word	C Class	OrthGem	Geminate
Subject5	buts	TS	no	yes
Subject1	blats	TS	no	yes
Subject2	blats	TS	no	yes
Subject3	blats	TS	no	yes
Subject4	blats	TS	no	yes
Subject5	blats	TS	no	yes
Subject1	blets	TS	no	yes
Subject2	blets	TS	no	yes
Subject3	blets	TS	no	yes
Subject4	blets	TS	no	yes
Subject5	blets	TS	no	yes
Subject1	blits	TS	no	yes
Subject2	blits	TS	no	yes
Subject3	blits	TS	no	yes
Subject4	blits	TS	no	yes
Subject5	blits	TS	no	yes
Subject1	bluts	TS	no	yes
Subject2	bluts	TS	no	yes
Subject3	bluts	TS	no	yes
Subject4	bluts	TS	no	yes
Subject5	bluts	TS	no	yes

Subject	Word	C Class	OrthGem	Geminate
Subject1	bat	T	no	yes
Subject2	bat	T	no	yes
Subject2	bat	T	no	no
Subject3	bat	T	no	no
Subject4	bat	T	no	yes
Subject4	bat	T	no	no
Subject5	bat	T	no	yes
Subject5	bat	T	no	no
Subject1	bet	T	no	yes
Subject2	bet	T	no	no
Subject2	bet	T	no	yes
Subject3	bet	T	no	no
Subject4	bet	T	no	yes
Subject4	bet	T	no	no
Subject5	bet	T	no	yes
Subject5	bet	T	no	no
Subject1	bit	T	no	no
Subject2	bit	T	no	no
Subject2	bit	T	no	yes
Subject3	bit	T	no	no
Subject4	bit	T	no	yes
Subject4	bit	T	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject5	bit	T	no	yes
Subject5	bit	T	no	no
Subject1	but	T	no	no
Subject2	but	T	no	no
Subject2	but	T	no	no
Subject3	but	T	no	yes
Subject4	but	T	no	yes
Subject4	but	T	no	no
Subject5	but	T	no	yes
Subject5	but	T	no	no
Subject1	blat	T	no	yes
Subject2	blat	T	no	yes
Subject2	blat	T	no	no
Subject3	blat	T	no	yes
Subject4	blat	T	no	yes
Subject4	blat	T	no	no
Subject5	blat	T	no	yes
Subject5	blat	T	no	no
Subject1	blet	T	no	yes
Subject2	blet	T	no	yes
Subject2	blet	T	no	no
Subject3	blet	T	no	yes

Subject	Word	C Class	OrthGem	Geminate
Subject4	blet	T	no	yes
Subject4	blet	T	no	no
Subject5	blet	T	no	yes
Subject5	blet	T	no	no
Subject1	blit	T	no	yes
Subject2	blit	T	no	yes
Subject2	blit	T	no	no
Subject3	blit	T	no	yes
Subject4	blit	T	no	yes
Subject4	blit	T	no	no
Subject5	blit	T	no	yes
Subject5	blit	T	no	no
Subject1	blut	T	no	yes
Subject2	blut	T	no	yes
Subject2	blut	T	no	no
Subject3	blut	T	no	yes
Subject4	blut	T	no	yes
Subject4	blut	T	no	no
Subject5	blut	T	no	yes
Subject5	blut	T	no	no
Subject1	bap	T	no	yes

Subject	Word	C Class	OrthGem	Geminate
Subject2	bap	T	no	yes
Subject2	bap	T	no	no
Subject3	bap	T	no	no
Subject4	bap	T	no	yes
Subject4	bap	T	no	no
Subject5	bap	T	no	yes
Subject5	bap	T	no	no
Subject1	bep	T	no	yes
Subject2	bep	T	no	yes
Subject2	bep	T	no	no
Subject3	bep	T	no	no
Subject4	bep	T	no	yes
Subject4	bep	T	no	no
Subject5	bep	T	no	yes
Subject5	bep	T	no	no
Subject1	bip	T	no	yes
Subject2	bip	T	no	yes
Subject2	bip	T	no	no
Subject3	bip	T	no	no
Subject4	bip	T	no	yes
Subject4	bip	T	no	no
Subject5	bip	T	no	yes

Subject	Word	C Class	OrthGem	Geminate
Subject5	bip	T	no	no
Subject1	bup	T	no	yes
Subject2	bup	T	no	yes
Subject2	bup	T	no	no
Subject3	bup	T	no	no
Subject4	bup	T	no	yes
Subject4	bup	T	no	no
Subject5	bup	T	no	yes
Subject5	bup	T	no	no
Subject1	blap	T	no	yes
Subject2	blap	T	no	yes
Subject2	blap	T	no	no
Subject3	blap	T	no	yes
Subject4	blap	T	no	yes
Subject4	blap	T	no	no
Subject5	blap	T	no	yes
Subject5	blap	T	no	no
Subject1	blep	T	no	yes
Subject2	blep	T	no	yes
Subject2	blep	T	no	no
Subject3	blep	T	no	yes
Subject4	blep	T	no	yes

Subject	Word	C Class	OrthGem	Geminate
Subject4	blep	T	no	no
Subject5	blep	T	no	yes
Subject5	blep	T	no	no
Subject1	blip	T	no	yes
Subject2	blip	T	no	yes
Subject2	blip	T	no	no
Subject3	blip	T	no	yes
Subject4	blip	T	no	yes
Subject4	blip	T	no	no
Subject5	blip	T	no	yes
Subject5	blip	T	no	no
Subject1	blup	T	no	yes
Subject2	blup	T	no	yes
Subject2	blup	T	no	no
Subject3	blup	T	no	yes
Subject4	blup	T	no	yes
Subject4	blup	T	no	no
Subject5	blup	T	no	yes
Subject5	blup	T	no	no
Subject1	back	T	no	yes
Subject2	back	T	no	yes

Subject	Word	C Class	OrthGem	Geminate
Subject2	back	T	no	no
Subject3	back	T	no	yes
Subject4	back	T	no	yes
Subject4	back	T	no	no
Subject5	back	T	no	yes
Subject5	back	T	no	no
Subject1	beck	T	no	yes
Subject2	beck	T	no	yes
Subject2	beck	T	no	no
Subject3	beck	T	no	yes
Subject4	beck	T	no	no
Subject4	beck	T	no	yes
Subject5	beck	T	no	yes
Subject5	beck	T	no	no
Subject1	bick	T	no	yes
Subject2	bick	T	no	yes
Subject2	bick	T	no	no
Subject3	bick	T	no	yes
Subject4	bick	T	no	yes
Subject4	bick	T	no	no
Subject5	bick	T	no	yes
Subject5	bick	T	no	no
Subject1	buck	T	no	yes
Subject2	buck	T	no	yes
Subject2	buck	T	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject3	buck	T	no	yes
Subject4	buck	T	no	yes
Subject4	buck	T	no	no
Subject5	buck	T	no	yes
Subject5	buck	T	no	no
Subject1	black	T	no	yes
Subject2	black	T	no	yes
Subject2	black	T	no	no
Subject3	black	T	no	yes
Subject4	black	T	no	yes
Subject4	black	T	no	no
Subject5	black	T	no	yes
Subject5	black	T	no	no
Subject1	bleck	T	no	yes
Subject2	bleck	T	no	yes
Subject2	bleck	T	no	no
Subject3	bleck	T	no	yes
Subject4	bleck	T	no	yes
Subject4	bleck	T	no	no
Subject5	bleck	T	no	yes
Subject5	bleck	T	no	no
Subject1	blick	T	no	yes

Subject	Word	C Class	OrthGem	Geminate
Subject2	blick	T	no	yes
Subject2	blick	T	no	no
Subject3	blick	T	no	yes
Subject4	blick	T	no	yes
Subject4	blick	T	no	no
Subject5	blick	T	no	yes
Subject5	blick	T	no	no
Subject1	bluck	T	no	yes
Subject2	bluck	T	no	yes
Subject2	bluck	T	no	no
Subject3	bluck	T	no	yes
Subject4	bluck	T	no	yes
Subject4	bluck	T	no	no
Subject5	bluck	T	no	yes
Subject5	bluck	T	no	no
Subject1	batt	T	yes	yes
Subject2	batt	T	yes	yes
Subject3	batt	T	yes	yes
Subject4	batt	T	yes	yes
Subject5	batt	T	yes	yes
Subject1	bett	T	yes	yes
Subject2	bett	T	yes	yes
Subject3	bett	T	yes	yes
Subject4	bett	T	yes	yes

Subject	Word	C Class	OrthGem	Geminate
Subject5	bett	T	yes	yes
Subject1	bitt	T	yes	yes
Subject2	bitt	T	yes	yes
Subject3	bitt	T	yes	yes
Subject4	bitt	T	yes	yes
Subject5	bitt	T	yes	yes
Subject1	butt	T	yes	yes
Subject2	butt	T	yes	yes
Subject3	butt	T	yes	yes
Subject4	butt	T	yes	yes
Subject5	butt	T	yes	yes
Subject1	bapp	T	yes	yes
Subject2	bapp	T	yes	yes
Subject3	bapp	T	yes	yes
Subject4	bapp	T	yes	yes
Subject5	bapp	T	yes	yes
Subject1	bepp	T	yes	yes
Subject2	bepp	T	yes	yes
Subject3	bepp	T	yes	yes
Subject4	bepp	T	yes	yes
Subject5	bepp	T	yes	yes
Subject1	bipp	T	yes	yes
Subject2	bipp	T	yes	yes

Subject	Word	C Class	OrthGem	Geminate
Subject5	bett	T	yes	yes
Subject3	bipp	T	yes	yes
Subject4	bipp	T	yes	yes
Subject5	bipp	T	yes	yes
Subject1	bupp	T	yes	yes
Subject2	bupp	T	yes	yes
Subject3	bupp	T	yes	yes
Subject4	bupp	T	yes	yes
Subject5	bupp	T	yes	yes
Subject1	baf	S	no	yes
Subject1	baf	S	no	no
Subject2	baf	S	no	yes
Subject2	baf	S	no	no
Subject3	baf	S	no	no
Subject4	baf	S	no	no
Subject5	baf	S	no	yes
Subject5	baf	S	no	no
Subject1	bef	S	no	yes
Subject1	bef	S	no	no
Subject2	bef	S	no	yes
Subject2	bef	S	no	no
Subject3	bef	S	no	no
Subject4	bef	S	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject5	bef	S	no	yes
Subject5	bef	S	no	no
Subject1	bif	S	no	yes
Subject1	bif	S	no	no
Subject2	bif	S	no	yes
Subject2	bif	S	no	no
Subject3	bif	S	no	no
Subject4	bif	S	no	no
Subject5	bif	S	no	yes
Subject5	bif	S	no	no
Subject1	buf	S	no	yes
Subject1	buf	S	no	no
Subject2	buf	S	no	yes
Subject2	buf	S	no	no
Subject3	buf	S	no	no
Subject4	buf	S	no	no
Subject5	buf	S	no	yes
Subject5	buf	S	no	no
Subject1	blaf	S	no	yes
Subject1	blaf	S	no	no
Subject2	blaf	S	no	yes
Subject2	blaf	S	no	no
Subject3	blaf	S	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject4	blaf	S	no	no
Subject5	blaf	S	no	yes
Subject5	blaf	S	no	no
Subject1	blef	S	no	yes
Subject1	blef	S	no	no
Subject2	blef	S	no	yes
Subject2	blef	S	no	no
Subject3	blef	S	no	no
Subject4	blef	S	no	no
Subject5	blef	S	no	yes
Subject5	blef	S	no	no
Subject1	blif	S	no	yes
Subject1	blif	S	no	no
Subject2	blif	S	no	yes
Subject2	blif	S	no	no
Subject3	blif	S	no	yes
Subject4	blif	S	no	no
Subject5	blif	S	no	yes
Subject5	blif	S	no	no
Subject1	bluf	S	no	yes
Subject1	bluf	S	no	no
Subject2	bluf	S	no	yes
Subject2	bluf	S	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject3	bluf	S	no	no
Subject4	bluf	S	no	no
Subject5	bluf	S	no	yes
Subject5	bluf	S	no	no
Subject1	bas	S	no	yes
Subject1	bas	S	no	no
Subject2	bas	S	no	yes
Subject2	bas	S	no	no
Subject3	bas	S	no	no
Subject4	bas	S	no	no
Subject5	bas	S	no	yes
Subject5	bas	S	no	no
Subject1	bes	S	no	yes
Subject1	bes	S	no	no
Subject2	bes	S	no	yes
Subject2	bes	S	no	no
Subject3	bes	S	no	no
Subject4	bes	S	no	no
Subject5	bes	S	no	yes
Subject5	bes	S	no	no
Subject1	bis	S	no	yes
Subject1	bis	S	no	no
Subject2	bis	S	no	yes

Subject	Word	C Class	OrthGem	Geminate
Subject2	bis	S	no	no
Subject3	bis	S	no	no
Subject4	bis	S	no	no
Subject5	bis	S	no	yes
Subject5	bis	S	no	no
Subject1	bus	S	no	yes
Subject1	bus	S	no	no
Subject2	bus	S	no	yes
Subject2	bus	S	no	no
Subject3	bus	S	no	no
Subject4	bus	S	no	no
Subject5	bus	S	no	yes
Subject5	bus	S	no	no
Subject1	blas	S	no	yes
Subject1	blas	S	no	no
Subject2	blas	S	no	yes
Subject2	blas	S	no	no
Subject3	blas	S	no	yes
Subject4	blas	S	no	no
Subject5	blas	S	no	yes
Subject5	blas	S	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject1	bles	S	no	yes
Subject1	bles	S	no	no
Subject2	bles	S	no	yes
Subject2	bles	S	no	no
Subject3	bles	S	no	yes
Subject4	bles	S	no	no
Subject5	bles	S	no	yes
Subject5	bles	S	no	no
Subject1	blis	S	no	yes
Subject1	blis	S	no	no
Subject2	blis	S	no	yes
Subject2	blis	S	no	no
Subject3	blis	S	no	yes
Subject4	blis	S	no	no
Subject5	blis	S	no	yes
Subject5	blis	S	no	no
Subject1	blus	S	no	yes
Subject1	blus	S	no	no
Subject2	blus	S	no	yes
Subject2	blus	S	no	no
Subject3	blus	S	no	yes
Subject4	blus	S	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject5	blus	S	no	yes
Subject5	blus	S	no	no
Subject1	bash	S	no	yes
Subject1	bash	S	no	no
Subject2	bash	S	no	yes
Subject2	bash	S	no	no
Subject3	bash	S	no	yes
Subject4	bash	S	no	yes
Subject4	bash	S	no	no
Subject5	bash	S	no	yes
Subject5	bash	S	no	no
Subject1	besh	S	no	yes
Subject1	besh	S	no	no
Subject2	besh	S	no	yes
Subject2	besh	S	no	no
Subject3	besh	S	no	yes
Subject4	besh	S	no	yes
Subject4	besh	S	no	no
Subject5	besh	S	no	yes
Subject5	besh	S	no	no
Subject1	bish	S	no	yes
Subject1	bish	S	no	no
Subject2	bish	S	no	yes

Subject	Word	C Class	OrthGem	Geminate
Subject2	bish	S	no	no
Subject3	bish	S	no	yes
Subject4	bish	S	no	yes
Subject4	bish	S	no	no
Subject5	bish	S	no	yes
Subject5	bish	S	no	no
Subject1	bush	S	no	yes
Subject1	bush	S	no	no
Subject2	bush	S	no	yes
Subject2	bush	S	no	no
Subject3	bush	S	no	yes
Subject4	bush	S	no	yes
Subject4	bush	S	no	no
Subject5	bush	S	no	yes
Subject5	bush	S	no	no
Subject1	blash	S	no	yes
Subject1	blash	S	no	no
Subject2	blash	S	no	yes
Subject2	blash	S	no	no
Subject3	blash	S	no	yes
Subject4	blash	S	no	yes
Subject4	blash	S	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject5	blash	S	no	yes
Subject5	blash	S	no	no
Subject1	blesh	S	no	yes
Subject1	blesh	S	no	no
Subject2	blesh	S	no	yes
Subject2	blesh	S	no	no
Subject3	blesh	S	no	yes
Subject4	blesh	S	no	yes
Subject4	blesh	S	no	no
Subject5	blesh	S	no	yes
Subject5	blesh	S	no	no
Subject1	blish	S	no	yes
Subject1	blish	S	no	no
Subject2	blish	S	no	yes
Subject2	blish	S	no	no
Subject3	blish	S	no	yes
Subject4	blish	S	no	yes
Subject4	blish	S	no	no
Subject5	blish	S	no	yes
Subject5	blish	S	no	no
Subject1	blush	S	no	yes
Subject1	blush	S	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject2	blush	S	no	yes
Subject2	blush	S	no	no
Subject3	blush	S	no	yes
Subject4	blush	S	no	yes
Subject4	blush	S	no	no
Subject5	blush	S	no	yes
Subject5	blush	S	no	no
Subject1	bass	S	yes	yes
Subject2	bass	S	yes	yes
Subject3	bass	S	yes	yes
Subject4	bass	S	yes	yes
Subject5	bass	S	yes	yes
Subject1	bess	S	yes	yes
Subject2	bess	S	yes	yes
Subject3	bess	S	yes	yes
Subject4	bess	S	yes	yes
Subject5	bess	S	yes	yes
Subject1	biss	S	yes	yes
Subject2	biss	S	yes	yes
Subject3	biss	S	yes	yes
Subject4	biss	S	yes	yes

Subject	Word	C Class	OrthGem	Geminate
Subject5	biss	S	yes	yes
Subject1	buss	S	yes	yes
Subject2	buss	S	yes	yes
Subject3	buss	S	yes	yes
Subject4	buss	S	yes	yes
Subject5	buss	S	yes	yes
Subject1	bad	D	no	yes
Subject1	bad	D	no	no
Subject2	bad	D	no	no
Subject3	bad	D	no	no
Subject4	bad	D	no	no
Subject5	bad	D	no	no
Subject1	bed	D	no	yes
Subject1	bed	D	no	no
Subject2	bed	D	no	no
Subject3	bed	D	no	no
Subject4	bed	D	no	no
Subject5	bed	D	no	no
Subject1	bid	D	no	yes
Subject1	bid	D	no	no
Subject2	bid	D	no	no
Subject3	bid	D	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject4	bid	D	no	no
Subject5	bid	D	no	no
Subject1	bud	D	no	yes
Subject1	bud	D	no	no
Subject2	bud	D	no	no
Subject3	bud	D	no	no
Subject4	bud	D	no	no
Subject5	bud	D	no	no
Subject1	blad	D	no	yes
Subject1	blad	D	no	no
Subject2	blad	D	no	no
Subject3	blad	D	no	yes
Subject4	blad	D	no	no
Subject5	blad	D	no	no
Subject1	bled	D	no	yes
Subject1	bled	D	no	no
Subject2	bled	D	no	no
Subject3	bled	D	no	yes
Subject4	bled	D	no	no
Subject5	bled	D	no	no
Subject1	blid	D	no	yes
Subject1	blid	D	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject2	blid	D	no	no
Subject3	blid	D	no	yes
Subject4	blid	D	no	no
Subject5	blid	D	no	no
Subject1	blud	D	no	yes
Subject1	blud	D	no	no
Subject2	blud	D	no	no
Subject3	blud	D	no	yes
Subject4	blud	D	no	no
Subject5	blud	D	no	no
Subject1	bab	D	no	yes
Subject1	bab	D	no	no
Subject2	bab	D	no	no
Subject3	bab	D	no	no
Subject4	bab	D	no	no
Subject5	bab	D	no	no
Subject1	beb	D	no	yes
Subject1	beb	D	no	no
Subject2	beb	D	no	no
Subject3	beb	D	no	no
Subject4	beb	D	no	no
Subject5	beb	D	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject1	bib	D	no	yes
Subject1	bib	D	no	no
Subject2	bib	D	no	no
Subject3	bib	D	no	no
Subject4	bib	D	no	no
Subject5	bib	D	no	no
Subject1	bub	D	no	yes
Subject1	bub	D	no	no
Subject2	bub	D	no	no
Subject3	bub	D	no	no
Subject4	bub	D	no	no
Subject5	bub	D	no	no
Subject1	blab	D	no	yes
Subject1	blab	D	no	no
Subject2	blab	D	no	no
Subject3	blab	D	no	yes
Subject4	blab	D	no	no
Subject5	blab	D	no	no
Subject1	bleb	D	no	yes
Subject1	bleb	D	no	no
Subject2	bleb	D	no	no
Subject3	bleb	D	no	yes

Subject	Word	C Class	OrthGem	Geminate
Subject4	bleb	D	no	no
Subject5	bleb	D	no	no
Subject1	blib	D	no	yes
Subject1	blib	D	no	no
Subject2	blib	D	no	no
Subject3	blib	D	no	yes
Subject4	blib	D	no	no
Subject5	blib	D	no	no
Subject1	blib	D	no	yes
Subject1	blub	D	no	no
Subject2	blub	D	no	no
Subject3	blub	D	no	yes
Subject4	blub	D	no	no
Subject5	blub	D	no	no
Subject1	bag	D	no	yes
Subject1	bag	D	no	no
Subject2	bag	D	no	no
Subject3	bag	D	no	no
Subject4	bag	D	no	no
Subject5	bag	D	no	no
Subject1	beg	D	no	yes
Subject1	beg	D	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject2	beg	D	no	no
Subject3	beg	D	no	no
Subject4	beg	D	no	no
Subject5	beg	D	no	no
Subject1	big	D	no	yes
Subject1	big	D	no	no
Subject2	big	D	no	no
Subject3	big	D	no	no
Subject4	big	D	no	no
Subject5	big	D	no	no
Subject1	bug	D	no	yes
Subject1	bug	D	no	no
Subject2	bug	D	no	no
Subject3	bug	D	no	no
Subject4	bug	D	no	no
Subject5	bug	D	no	no
Subject1	blag	D	no	yes
Subject1	blag	D	no	no
Subject2	blag	D	no	no
Subject3	blag	D	no	yes
Subject4	blag	D	no	no
Subject5	blag	D	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject1	bleg	D	no	yes
Subject1	bleg	D	no	no
Subject2	bleg	D	no	no
Subject3	bleg	D	no	yes
Subject4	bleg	D	no	no
Subject5	bleg	D	no	no
Subject1	blig	D	no	yes
Subject1	blig	D	no	no
Subject2	blig	D	no	no
Subject3	blig	D	no	yes
Subject4	blig	D	no	no
Subject5	blig	D	no	no
Subject1	blug	D	no	yes
Subject1	blab	D	no	no
Subject2	blug	D	no	no
Subject3	blug	D	no	yes
Subject4	blug	D	no	no
Subject5	blug	D	no	no
Subject1	ban	N	no	no
Subject2	ban	N	no	no
Subject3	ban	N	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject4	ban	N	no	no
Subject5	ban	N	no	no
Subject1	ben	N	no	no
Subject2	ben	N	no	no
Subject3	ben	N	no	no
Subject4	ben	N	no	no
Subject5	ben	N	no	no
Subject1	bin	N	no	no
Subject2	bin	N	no	no
Subject3	bin	N	no	no
Subject4	bin	N	no	no
Subject5	bin	N	no	no
Subject1	bun	N	no	no
Subject2	bun	N	no	no
Subject3	bun	N	no	no
Subject4	bun	N	no	no
Subject5	bun	N	no	no
Subject1	blan	N	no	no
Subject2	blan	N	no	no
Subject3	blan	N	no	yes
Subject4	blan	N	no	no
Subject5	blan	N	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject1	blen	N	no	no
Subject2	blen	N	no	no
Subject3	blen	N	no	yes
Subject4	blen	N	no	no
Subject5	blen	N	no	no
Subject1	blin	N	no	no
Subject2	blin	N	no	no
Subject3	blin	N	no	yes
Subject4	blin	N	no	no
Subject5	blin	N	no	no
Subject1	blun	N	no	no
Subject2	blun	N	no	no
Subject3	blun	N	no	yes
Subject4	blun	N	no	no
Subject5	blun	N	no	no
Subject1	bam	N	no	no
Subject2	bam	N	no	no
Subject3	bam	N	no	no
Subject4	bam	N	no	no
Subject5	bam	N	no	no
Subject1	bem	N	no	no
Subject2	bem	N	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject3	bem	N	no	no
Subject4	bem	N	no	no
Subject5	bem	N	no	no
Subject1	bim	N	no	no
Subject2	bim	N	no	no
Subject3	bim	N	no	no
Subject4	bim	N	no	no
Subject5	bim	N	no	no
Subject1	bum	N	no	no
Subject2	bum	N	no	no
Subject3	bum	N	no	no
Subject4	bum	N	no	no
Subject5	bum	N	no	no
Subject1	blam	N	no	no
Subject2	blam	N	no	no
Subject3	blam	N	no	yes
Subject4	blam	N	no	no
Subject5	blam	N	no	no
Subject1	blem	N	no	no
Subject2	blem	N	no	no
Subject3	blem	N	no	yes
Subject4	blem	N	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject5	blem	N	no	no
Subject1	blim	N	no	no
Subject2	blim	N	no	no
Subject3	blim	N	no	yes
Subject4	blim	N	no	no
Subject5	blim	N	no	no
Subject1	blum	N	no	no
Subject2	blum	N	no	no
Subject3	blum	N	no	yes
Subject4	blum	N	no	no
Subject5	blum	N	no	no
Subject1	bann	N	yes	yes
Subject2	bann	N	yes	yes
Subject3	bann	N	yes	yes
Subject4	bann	N	yes	yes
Subject5	bann	N	yes	yes
Subject1	benn	N	yes	yes
Subject2	benn	N	yes	yes
Subject3	benn	N	yes	yes
Subject4	benn	N	yes	yes
Subject5	benn	N	yes	yes

Subject	Word	C Class	OrthGem	Geminate
Subject1	binn	N	yes	yes
Subject2	binn	N	yes	yes
Subject3	binn	N	yes	yes
Subject4	binn	N	yes	yes
Subject5	binn	N	yes	yes
Subject1	bunn	N	yes	yes
Subject2	bunn	N	yes	yes
Subject3	bunn	N	yes	yes
Subject4	bunn	N	yes	yes
Subject5	bunn	N	yes	yes
Subject1	bamm	N	yes	yes
Subject2	bamm	N	yes	yes
Subject3	bamm	N	yes	yes
Subject4	bamm	N	yes	yes
Subject5	bamm	N	yes	yes
Subject1	bemm	N	yes	yes
Subject2	bemm	N	yes	yes
Subject3	bemm	N	yes	yes
Subject4	bemm	N	yes	yes
Subject5	bemm	N	yes	yes
Subject1	bimm	N	yes	yes
Subject2	bimm	N	yes	yes

Subject	Word	C Class	OrthGem	Geminate
Subject3	bimm	N	yes	yes
Subject4	bimm	N	yes	yes
Subject5	bimm	N	yes	yes
Subject1	bumm	N	yes	yes
Subject2	bumm	N	yes	yes
Subject3	bumm	N	yes	yes
Subject4	bumm	N	yes	yes
Subject5	bumm	N	yes	yes
Subject1	bal	L	no	no
Subject2	bal	L	no	no
Subject3	bal	L	no	no
Subject4	bal	L	no	no
Subject5	bal	L	no	no
Subject1	bel	L	no	no
Subject2	bel	L	no	no
Subject3	bel	L	no	no
Subject4	bel	L	no	no
Subject5	bel	L	no	no
Subject1	bil	L	no	no
Subject2	bil	L	no	no
Subject3	bil	L	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject4	bil	L	no	no
Subject5	bil	L	no	no
Subject1	bul	L	no	no
Subject2	bul	L	no	no
Subject3	bul	L	no	no
Subject4	bul	L	no	no
Subject5	bul	L	no	no
Subject1	ball	L	yes	yes
Subject2	ball	L	yes	yes
Subject3	ball	L	yes	yes
Subject4	ball	L	yes	yes
Subject5	ball	L	yes	yes
Subject1	bell	L	yes	yes
Subject2	bell	L	yes	yes
Subject3	bell	L	yes	yes
Subject4	bell	L	yes	yes
Subject5	bell	L	yes	yes
Subject1	bill	L	yes	yes
Subject2	bill	L	yes	yes
Subject3	bill	L	yes	yes
Subject4	bill	L	yes	yes

Subject	Word	C Class	OrthGem	Geminate
Subject5	bill	L	yes	yes
Subject1	bull	L	yes	yes
Subject2	bull	L	yes	yes
Subject3	bull	L	yes	yes
Subject4	bull	L	yes	yes
Subject5	bull	L	yes	yes
Subject1	bazz	Z	yes	yes
Subject2	bazz	Z	yes	yes
Subject3	bazz	Z	yes	yes
Subject4	bazz	Z	yes	yes
Subject5	bazz	Z	yes	yes
Subject1	bezz	Z	yes	yes
Subject2	bezz	Z	yes	yes
Subject3	bezz	Z	yes	yes
Subject4	bezz	Z	yes	yes
Subject5	bezz	Z	yes	yes
Subject1	bizz	Z	yes	yes
Subject2	bizz	Z	yes	yes
Subject3	bizz	Z	yes	yes

Subject	Word	C Class	OrthGem	Geminate
Subject4	bizz	Z	yes	yes
Subject5	bizz	Z	yes	yes
Subject1	buzz	Z	yes	yes
Subject2	buzz	Z	yes	yes
Subject3	buzz	Z	yes	yes
Subject4	buzz	Z	yes	yes
Subject5	buzz	Z	yes	yes
Subject1	baz	Z	no	no
Subject2	baz	Z	no	no
Subject3	baz	Z	no	no
Subject4	baz	Z	no	no
Subject5	baz	Z	no	no
Subject1	bez	Z	no	no
Subject2	bez	Z	no	no
Subject3	bez	Z	no	no
Subject4	bez	Z	no	no
Subject5	bez	Z	no	no
Subject1	biz	Z	no	no
Subject2	biz	Z	no	no
Subject3	biz	Z	no	no
Subject4	biz	Z	no	no
Subject5	biz	Z	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject1	buz	Z	no	no
Subject2	buz	Z	no	no
Subject3	buz	Z	no	no
Subject4	buz	Z	no	no
Subject5	buz	Z	no	no
Subject1	blaz	Z	no	no
Subject2	blaz	Z	no	no
Subject3	blaz	Z	no	no
Subject4	blaz	Z	no	no
Subject5	blaz	Z	no	no
Subject1	blez	Z	no	no
Subject2	blez	Z	no	no
Subject3	blez	Z	no	no
Subject4	blez	Z	no	no
Subject5	blez	Z	no	no
Subject1	bliz	Z	no	no
Subject2	bliz	Z	no	no
Subject3	bliz	Z	no	no
Subject4	bliz	Z	no	no
Subject5	bliz	Z	no	no
Subject1	bluz	Z	no	no
Subject2	bluz	Z	no	no
Subject3	bluz	Z	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject4	bluz	Z	no	no
Subject5	bluz	Z	no	no
Subject1	bav	Z	no	no
Subject2	bav	Z	no	no
Subject3	bav	Z	no	no
Subject4	bav	Z	no	no
Subject5	bav	Z	no	no
Subject1	bev	Z	no	no
Subject2	bev	Z	no	no
Subject3	bev	Z	no	no
Subject4	bev	Z	no	no
Subject5	bev	Z	no	no
Subject1	biv	Z	no	no
Subject2	biv	Z	no	no
Subject3	biv	Z	no	no
Subject4	biv	Z	no	no
Subject5	biv	Z	no	no
Subject1	buv	Z	no	no
Subject2	buv	Z	no	no
Subject3	buv	Z	no	no
Subject4	buv	Z	no	no

Subject	Word	C Class	OrthGem	Geminate
Subject5	buv	Z	no	no
Subject1	blav	Z	no	no
Subject2	blav	Z	no	no
Subject3	blav	Z	no	no
Subject4	blav	Z	no	no
Subject5	blav	Z	no	no
Subject1	blev	Z	no	no
Subject2	blev	Z	no	no
Subject3	blev	Z	no	no
Subject4	blev	Z	no	no
Subject5	blev	Z	no	no
Subject1	bliv	Z	no	no
Subject2	bliv	Z	no	no
Subject3	bliv	Z	no	no
Subject4	bliv	Z	no	no
Subject5	bliv	Z	no	no
Subject1	bluv	Z	no	no
Subject2	bluv	Z	no	no
Subject3	bluv	Z	no	no
Subject4	bluv	Z	no	no
Subject5	bluv	Z	no	no

Appendix B

Chapter 3

B.1 Distribution of singletons and geminates in the corpus

B.1.1 Native Hungarian words

Word-final consonants in monosyllables following short vowels

	Singleton	Geminate	Total	Geminate(%)
[p]	10	7	17	41.18%
[t]	15	8	23	34.78%
[k]	18	7	25	28.00%
Voiceless stops	43	22	65	33.85%
[ts]	2	9	11	81.81%
[tʃ]	10	3	13	23.08%
Voiceless affricates	12	12	24	50.00%

	Singleton	Geminate	Total	Geminate(%)
[f]	3	5	8	62.50%
[ʃ]	14	10	24	41.67%
[s]	19	8	27	29.63%
Voiceless fricatives	36	23	59	38.98%
[b]	13	3	16	18.17%
[d]	11	9	20	45.00%
[g]	15	4	19	21.05%
Voiced stops	39	16	55	29.09%
[m]	11	2	13	15.38%
[n]	14	6	20	30.00%
Nasals	25	8	33	24.24%
[l]	24	7	31	22.58%
[r]	21	5	26	19.23%
Liquids	45	12	57	21.05%
[v]	2	0	2	0.00%
[z]	8	1	9	11.11%
[ʒ]	2	0	2	0.00%
Voiced fricatives	12	1	13	11.11%

Word-final consonants in monosyllables following long vowels

	Singleton	Geminate	Total	Geminate
[p]	16	1	17	5.88%
[t]	25	0	25	0.00%
[k]	13	0	13	0.00%
Voiceless stops	54	1	55	1.82%
[ts]	7	0	7	0.00%
[tʃ]	5	0	5	0.00%
Voiceless affricates	12	0	12	0.00%
[f]	3	0	3	0.00%
[ʃ]	17	0	17	0.00%
[s]	13	0	13	0.00%
Voiceless fricatives	33	0	33	0.00%
[b]	4	0	4	0.00%
[d]	12	0	12	0.00%
[g]	25	0	25	0.00%
Voiced stops	41	0	41	0.00%
[m]	15	0	15	0.00%
[n]	21	0	21	0.00%
Nasals	36	0	36	0.00%
[l]	25	2	27	7.41%
[r]	49	0	49	0.00%
Liquids	74	2	76	2.63%
[v]	11	0	11	0.00%
[z]	26	0	26	0.00%
[ʒ]	2	0	2	0.00%
Voiced fricatives	39	0	39	0.00%

Intervocalic consonants in disyllables following short vowels

	Singleton	Geminate	Total	Geminate(%)
[p]	87	37	124	29.84%
[t]	107	26	133	19.55%
[k]	128	48	176	27.28%
Voiceless stops	322	111	433	25.64%
[ts]	60	12	72	16.67%
[tʃ]	50	12	62	19.35%
Voiceless affricates	110	24	134	17.91%
[f]	31	19	50	38.00%
[ʃ]	45	5	50	10.00%
[s]	56	29	85	34.12%
Voiceless fricatives	132	53	185	28.65%
[b]	99	26	125	20.80%
[d]	61	3	64	4.69%
[g]	102	14	116	12.07%
Voiced stops	262	43	305	14.10%
[m]	79	11	90	12.22%
[n]	43	14	57	24.56%
Nasals	122	25	147	17.01%
[l]	151	62	213	29.11%
[r]	215	25	240	10.42%
Liquids	366	87	453	19.21%
[v]	90	0	90	0.00%
[z]	67	0	67	0.00%
[ʒ]	15	0	15	0.00%
Voiced fricatives	172	0	172	0.00%

Intervocalic consonants in disyllables following long vowels

	Singleton	Geminate	Total	Geminate
[p]	53	1	54	1.85%
[t]	96	4	100	4.00%
[k]	62	2	64	3.13%
Voiceless stops	211	7	218	3.21%
[ts]	22	0	22	0.00%
[tʃ]	22	0	22	0.00%
Voiceless affricates	44	0	44	0.00%
[f]	31	0	31	0.00%
[ʃ]	67	0	67	0.00%
[s]	63	0	63	0.00%
Voiceless fricatives	161	0	161	0.00%
[b]	49	0	49	0.00%
[d]	66	0	66	0.00%
[g]	87	0	87	0.00%
Voiced stops	202	0	202	0.00%
[m]	64	0	64	0.00%
[n]	53	0	53	0.00%
Nasals	117	0	117	0.00%
[l]	101	21	122	17.21%
[r]	168	0	168	0.00%
Liquids	269	21	290	7.24%
[v]	87	0	87	0.00%
[z]	97	0	97	0.00%
[ʒ]	14	0	14	0.00%
Voiced fricatives	198	0	198	0.00%

B.1.2 All Hungarian words

Word-final consonants in monosyllables following short vowels

	Singleton	Geminate	Total	Geminate(%)
[p]	13	18	31	58.06%
[t]	16	21	37	56.76%
[k]	18	28	46	60.86%
Voiceless stops	47	67	114	58.77%
[ts]	3	23	26	88.46%
[tʃ]	9	10	19	52.63%
Voiceless affricates	12	33	45	73.33%
[f]	3	8	11	72.73%
[ʃ]	14	14	28	50.00%
[s]	19	19	38	50.00%
Voiceless fricatives	36	41	77	53.25%
[b]	17	5	22	22.73%
[d]	11	9	20	45.00%
[g]	19	5	24	20.83%
Voiced stops	47	19	66	28.79%
[m]	12	4	16	25.00%
[n]	17	7	24	29.17%
Nasals	29	11	40	27.50%
[l]	25	11	36	30.56%
[r]	21	5	26	19.23%
Liquids	46	16	62	25.80%
[v]	2	0	2	0.00%
[z]	8	1	9	11.11%
[ʒ]	2	0	2	0.00%
Voiced fricatives	12	1	13	11.11%

Word-final consonants in monosyllables following long vowels

	Singleton	Geminate	Total	Geminate
[p]	16	1	17	5.88%
[t]	25	0	25	0.00%
[k]	13	0	13	0.00%
Voiceless stops	54	1	55	1.82%
[ts]	7	0	7	0.00%
[tʃ]	5	0	5	0.00%
Voiceless affricates	12	0	12	0.00%
[f]	3	0	3	0.00%
[ʃ]	17	0	17	0.00%
[s]	13	0	13	0.00%
Voiceless fricatives	33	0	33	0.00%
[b]	4	0	4	0.00%
[d]	12	0	12	0.00%
[g]	25	0	25	0.00%
Voiced stops	41	0	41	0.00%
[m]	15	0	15	0.00%
[n]	21	0	21	0.00%
Nasals	36	0	36	0.00%
[l]	25	2	27	7.41%
[r]	49	0	49	0.00%
Liquids	74	2	76	2.63%
[v]	11	0	11	0.00%
[z]	26	0	26	0.00%
[ʒ]	2	0	2	0.00%
Voiced fricatives	39	0	39	0.00%

Intervocalic consonants in disyllables following short vowels

	Singleton	Geminate	Total	Geminate(%)
[p]	96	54	150	36.00%
[t]	122	37	159	23.27%
[k]	120	79	199	36.69%
Voiceless stops	338	170	508	33.46%
[ts]	65	31	96	32.29%
[tS]	50	16	66	36.73%
Voiceless affricates	115	47	162	29.01%
[f]	37	25	62	40.32%
[ʃ]	42	5	47	10.64%
[s]	57	42	99	42.42%
Voiceless fricatives	136	72	208	34.62%
[b]	102	29	131	22.13%
[d]	72	5	77	6.49%
[g]	110	17	127	13.39%
Voiced stops	284	51	335	15.22%
[m]	97	16	113	14.16%
[n]	56	17	73	23.29%
Nasals	153	33	186	17.74%
[l]	174	73	247	9.55%
[r]	231	26	257	10.12%
Liquids	405	99	504	19.64%
[v]	100	0	100	0.00%
[z]	72	11	83	13.25%
[ʒ]	15	0	15	0.00%
Voiced fricatives	187	11	198	5.56%

Intervocalic consonants in disyllables following long vowels

	Singleton	Geminate	Total	Geminate
[p]	53	1	54	1.85%
[t]	96	4	100	4.00%
[k]	62	2	64	3.13%
Voiceless stops	211	7	218	3.21%
[ts]	22	0	22	0.00%
[tʃ]	22	0	22	0.00%
Voiceless affricates	44	0	44	0.00%
[f]	31	0	31	0.00%
[ʃ]	67	0	67	0.00%
[s]	63	0	63	0.00%
Voiceless fricatives	161	0	161	0.00%
[b]	49	0	49	0.00%
[d]	66	0	66	0.00%
[g]	87	0	87	0.00%
Voiced stops	202	0	202	0.00%
[m]	64	0	64	0.00%
[n]	53	0	53	0.00%
Nasals	117	0	117	0.00%
[l]	101	21	122	17.21%
[r]	168	0	168	0.00%
Liquids	269	21	290	7.24%
[v]	87	0	87	0.00%
[z]	97	0	97	0.00%
[ʒ]	14	0	14	0.00%
Voiced fricatives	198	0	198	0.00%

Word-final consonants polysyllables following short vowels

	Singleton	Geminate	Total	Geminate
[p]	571	39	610	6.39%
[t]	8286	1023	9309	10.99%
[k]	12120	80	12200	0.66%
Voiceless stops	20977	1142	22119	5.16%
[ts]	89	14	103	13.59%
[tʃ]	75	13	88	14.77%
Voiceless affricates	164	27	191	14.13%
[f]	15	8	23	34.78%
[ʃ]	4757	10	4767	0.21%
[s]	440	36	476	7.56%
Voiceless fricatives	5212	54	5266	1.03%
[b]	201	1	202	0.50%
[d]	1498	3	1501	0.20%
[g]	3348	17	3365	0.51%
Voiced stops	5047	21	5068	0.42%
[m]	1405	9	1414	0.64%
[n]	3111	17	3128	0.54%
Nasals	4516	26	4542	4.76%
[l]	9529	50	9579	0.52%
[r]	2395	5	2400	0.21%
Liquids	11924	55	11979	0.46%
[v]	59	0	59	0.00%
[z]	3183	0	3183	0.00%
[ʒ]	6	0	6	0.00%
Voiced fricatives	3248	0	3248	0.00%

B.2 Nonce word well-formedness experiments

B.2.1 Monosyllables

zat(t)

det(t)

szit(t)

szut(t)

dap(p)

fep(p)

mip(p)

kup(p)

gak(k)

lek(k)

nik(k)

cuk(k)

lac(c)s

tec(c)s

tic(c)s

luc(c)s

nac(c)

fec(c)

tic(c)

zuc(c)

taf(f)

kef(f)

csif(f)

buf(f)

gas(s)

pes(s)

fis(s)

vus(s)

csas(s)z

kes(s)z

nis(s)z

dus(s)z

nab(b)

keb(b)

tib(b)

fub(b)

gad(d)

ned(d)

tid(d)

csud(d)

kag(g)

deg(g)

lig(g)

szug(g)

lam(m)

kem(m)

nim(m)

csum(m)

pan(n)

sen(n)

din(n)

szun(n)

kal(l)

pel(l)

zil(l)

tul(l)

dar(r)

ter(r)

kir(r)

vur(r)

kav(v)

mev(v)

csiv(v)

duv(v)

kaz(z)

dez(z)

giz(z)

csuz(z)

naz(z)s

lez(z)s

miz(z)s

nuz(z)s

B.2.2 Disyllables

dak(k)ik

bik(k)i

mek(k)er

luk(k)is

fap(p)i

mep(p)ik

szip(p)er

nup(p)is

szat(t)er

set(t)ik

bit(t)i

dut(t)is

fuc(c)is

tac(c)i

vec(c)er

bic(c)ik

dac(c)ser

szuc(c)sik

lec(c)si

nic(c)si

tif(f)i

vaf(f)ik

lef(f)is

nuf(f)er

vas(s)i

res(s)is

lis(s)er

gus(s)ik

tas(s)zer

csis(s)zi

fes(s)zis

lus(s)zik

peb(b)ik

tab(b)er

szib(b)i

med(d)is

nad(d)i

pid(d)ik

lud(d)er

mag(g)er

teg(g)is

fig(g)i

vug(g)ik

fam(m)i

lem(m)ik

szim(m)is

csum(m)er

len(n)is

gun(n)i

bin(n)ik

csan(n)er

tul(l)er

zel(l)is

jil(l)ik

kal(l)i

ner(r)is

kir(r)er

lur(r)ik

csar(r)i

tev(v)er

piv(v)is

pav(v)i

nuv(v)ik

guz(z)is

csaz(z)i

szez(z)ik

kiz(z)er

mez(z)sik

duz(z)si

taz(z)sis

fiz(z)ser

Appendix C

Chapter 5

C.1 Tableaux: Native grammar

Table C.1

Input: VT	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VT	0.41	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
VTT	0.59	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0

Table C.2

Input: V:T	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:T	0.98	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:TT	0.02	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0

Table C.3

Input: VTS	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VTS	0.27	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
VTTS	0.73	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.4

Input: V:TS	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:TS	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:TTS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0

Table C.5

Input: VS	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VS	0.47	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
VSS	0.53	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0

Table C.6

Input: V:S	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:S	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:SS	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0

Table C.7

Input: VD	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VD	0.72	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
VDD	0.29	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0

Table C.8

Input: V:D	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:D	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:DD	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0

Table C.9

Input: VN	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VN	0.73	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
VNN	0.27	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0

Table C.10

Input: V:N	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:N	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:NN	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0

Table C.11

Input: VL	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VL	0.74	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
VLL	0.26	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table C.12

Input: V:L	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:L	0.97	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:LL	0.03	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Table C.13

Input: VZ	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VZ	0.89	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
VZZ	0.11	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table C.14

Input: V:Z	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:Z	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:ZZ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Table C.15

Input: VTV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VTV	0.77	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VTTV	0.33	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0

Table C.16

Input: V:TV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:TV	0.97	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:TTV	0.03	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0

Table C.17

Input: VTSV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VTSV	0.71	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VTTSV	0.29	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.18

Input: V:TSV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:TSV	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:TTSV	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0

Table C.19

Input: VSV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VSV	0.65	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VSSV	0.35	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0

Table C.20

Input: V:SV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:SV	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:SSV	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Table C.21

Input: VDV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VDV	0.85	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VDDV	0.15	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0

Table C.22

Input: V:DV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:DV	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:DDV	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0

Table C.23

Input: VNV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VNV	0.82	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VNNV	0.18	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0

Table C.24

Input: V:NV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:NV	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:NNV	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0

Table C.25

Input: VLV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VLV	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VLLV	0.2	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table C.26

Input: V:LV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:LV	0.93	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:LLV	0.07	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Table C.27

Input: VZV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VZV	0.94	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VZZV	0.06	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table C.28

Input: V:ZV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:ZV	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:ZZV	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Table C.29

Input: .VT	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
.VT	0.95	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
.VTT	0.05	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0

Table C.30

Input: .V:T	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
.V:T	0.99	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
.V:TT	0.01	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0

Table C.31

Input: .VTS	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VTS	0.86	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VTTS	0.14	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.32

Input: .V:TS	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:TS	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:TTS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0

Table C.33

Input: .VS	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VS	0.99	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VSS	0.01	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0

Table C.34

Input: .V:S	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:S	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:SS	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0

Table C.35

Input: .VD	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VD	0.99	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VDD	0.01	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0

Table C.36

Input: .V:D	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:D	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:DD	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0

Table C.37

Input: .VN	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VN	0.99	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VNN	0.01	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0

Table C.38

Input: .V:N	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:N	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:NN	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0

Table C.39

Input: .VL	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VL	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VLL	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table C.40

Input: .V:L	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:L	0.98	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:LL	0.02	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Table C.41

Input: .VZ	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
VZ	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VZZ	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table C.42

Input: .V:Z	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	*V:CC
V:Z	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V:ZZ	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

C.2 Loanword grammar

Table C.43

Input: <t >VxT	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VT	0.38	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0
VxTT	0.62	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.44

Input: <tt >VT	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
VxTT	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.45

Input: <ts >VxTS	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VTS	0.17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0
VxTTS	0.83	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.46

Input: <s >VS	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VS	0.63	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
VSS	0.37	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.47

Input: <ss >VS	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0
VSS	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.48

Input: <d >VD	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VD	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
VDD	0.2	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.49

Input: <dd >VD	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0
VDD	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.50

Input: <n >VN	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VN	0.88	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
VNN	0.12	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.51

Input: <nn >VN	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0
VNN	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.52

Input: <l >VL	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VL	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
VLL	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.53

Input: <ll >VL	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VL	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0
VLL	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.54

Input: <z >VZ	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VZ	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
VZZ	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.55

Input: <zz >VZ	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VZ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0
VZZ	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.56

Input: <t >VTV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VTV	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VTTV	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Table C.57

Input: <tt >VTV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VTV	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
VTTV	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Table C.58

Input: <ts >VTSV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VTSV	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VTTSV	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Table C.59

Input: <s >VSV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VSV	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VSSV	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Table C.60

Input: <ss >VSV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VSV	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
VSSV	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Table C.61

Input: <d >VDV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VDV	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VDDV	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Table C.62

Input: <dd >VDV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VDV	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
VDDV	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Table C.63

Input: <n >VNV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VNV	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VNNV	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Table C.64

Input: <nn >VNV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VNV	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
VNNV	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Table C.65

Input: <l >VLV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VLV	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VLLV	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Table C.66

Input: <ll >VLLV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VLV	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
VLLV	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Table C.67

Input: <z >VZV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VZV	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VZZV	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Table C.68

Input: <zz >VZV	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
VZV	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
VZZV	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Table C.69

Input: <t >.VxT	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
.VT	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
.VTT	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.70

Input: <tt >.VxT	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
.VT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0
.VTT	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.71

Input: <t >.'VxT	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
.VT	0.38	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
.'VTT	0.62	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Table C.72

Input: <tt >.'VxT	Pr	*ZZ	*LL	*NN	*DD	*SS	*TT	*TTS	Minimality	MaxOrthGem	IdentVDur(str)	IdentVDur(unstr)	IdentC(l)VC	IdentC(l)VCV	*V:CC
.VT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0
.'VTT	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Bibliography

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*, 2014.

Péter Halácsy, András Kornai, Nemeth Laszlo, Rung Andras, István Szakadát, and Tron Viktor. Creating open language resources for Hungarians. *Proceedings of the 4th International Conference on Language Resources and evaluation (LREC)*, 2004.

William H. Ham. *Phonetic and Phonological Aspects of Geminate Timing*. Outstanding dissertations in linguistics. Psychology Press, 2001. ISBN 9780415937603.

Silke Hamann and Ilaria E Colombo. A formal account of the interaction of orthography and perception. *Natural Language & Linguistic Theory*, pages 1–32, 2017.

Bruce Hayes and Colin Wilson. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic inquiry*, 39(3):379–440, 2008.

Bálint Huszthy. Capuccinóba completát? A mássalhangzóhossz magyar aszimmetriái idegen szavakban. *In Studia Varia*, pages 125–139, 2002.

Daniel Howard Karvonen. Finnish loanword phonology and the core-periphery structure of the lexicon. *handout of talk at SWOT 4, University of Arizona*, 1998.

Daniel Howard Karvonen. Word prosody in finnish. *Ph.D dissertation*, 2005.

Shigeto Kawahara. Sonorancy and geminacy. *University of Massachusetts Occasional Papers in Linguistics 32: Papers in Optimality III*, 2007.

Shigeto Kawahara. The phonetics of obstruent geminates, sokuon. *The Mouton Handbook of Japanese Language and Linguistics*, 2015.

Zsuzsa Kertész. Approaches to the phonological analysis of loanword adaptation. *The Even Yearbook*, 7:1–15, 2006.

Paul Kiparsky. *Finnish noun inflection.* CSLI, Stanford, CA, 2003.

Magdolna Kovács. Beszédhangok kontextusfüggő időviszonyai. In Mária Gósy, editor, *Beszédkutatás 2000, MTA Nyelvtudományi Intézet*, 2000.

Bhadriraju Krishnamurti and John Peter Lucius Gwynn. A Grammar of Modern Telugu. *Oxford University Press*, 1985.

Margaret Ilona Kroll. Vowel and Consonant Lengthening in Finnish Loanword Adaptation. *Master's thesis*, 2014.

Haruo Kubozono, Junko Ito, and Armin Mester. Consonant gemination in Japanese loanword phonology: A phonological account. *Proceedings of the 18th International Congress of Linguists*, pages 28–47, 2008.

Katalin Mády. Shortening of long high vowels in Hungarian: a perceptual loss? *Proceedings of Sociophonetics at the crossroads of speech variation, processing and communication, Pisa*, 2010.

Katalin Mády, Lasse Bombien, and Uwe D Reichel. Is Hungarian losing the vowel quantity distinction? *Proceedings of 8th International Semi,nar on Speech Production, Strasbourg*, 2008.

Klára Magdics. A magyar beszédhangok időtartama [duration of hungarian speech sounds]. *Nyelvtudományi Közlemények*, 68:125–139, 1966.

Lilla Magyar. Gemination in Hungarian loanword adaptation. *MIT Linguistics Generals paper*, 2014.

Lilla Magyar. Are Universal Markedness Hierarchies Learnable from the Lexicon? The Case of Gemination in Hungarian. In Kyeong-Min Kim et al., editor, *Proceedings of the 33rd West Coast Conference on Formal Linguistics*, pages 287–298. Cascadilla Proceedings Project, Somerville, MA, 2016.

Ádám Nádasdy. Consonant length in recent borrowings into hungarian. *Acta Linguistica Hungarica*, 39:195–213, 1989.

Savithry Namboodiripad, Marc Garellek, and Eric Bakovic. Moraic Geminates in Malayalam: Evidence from minimal word effects and loanword adaptation. *Talk at the Old World Conference in Phonology, Barcelona, Spain*, 2014.

Tilda Neuberger and András Beke. Challenges in the analysis and processing of spontaneous speech. In *Talk presented at CAPSS 2017, May 14-17, Budapest, Hungary*, 2017.

Gábor Olaszy. Hangidőtartamok es időszerkezeti elemek a magyar beszédben. In *Nyelvtudományi Értekezések 155, Akadémiai Kiadó, Budapest*, 2006.

Ferenc Papp. Reverse-Alphabetized Dictionary of the Hungarian Language. *Akadémiai Kiadó, Budapest*, 1969.

Diana Passino. *Aspects of consonantal lengthening in Italian.* Unipress, Padova, 2008.

Sharon Peperkamp. A psycholinguistic theory of loanword adaptations. In N. Fleischer & M. Park-Doob M. Ettlinger, editor, *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society*, volume 30, pages 341–352. Berkeley Linguistics Society, Berkeley, CA, 2005.

Robert J. Podesva. Segmental constraints on geminates and their implications for typology. *Talk given at LSA Annual Meeting*, 2002.

Robert F Port. Linguistic timing factors in combination. *The Journal of the Acoustical Society of America*, 69(1):262–274, 1981.

Péter Siptár and Miklós Törkenczy. *The phonology of Hungarian.* Phonology of the World's Languages S., 2000.

Donca Steriade. Sources of markedness and why they matter. *GLOW, Markedness Workshop*, 2004.

Miklós Törkenczy. Does the onset branch in Hungarian? *Acta Linguistica Hungarica*, 39:273–292, 1989.

Jan PH Van Santen. Contextual effects on vowel duration. *Speech communication*, 11(6):513–546, 1992.

Laurence White and Katalin Mády. The long and the short and the final: Phonological vowel length and prosodic timing in Hungarian. In *Proceedings of 4th Speech Prosody Conference, Campinas, Brasil*, pages 363–366, 2008.