Exceptionality in vowel harmony

by

Daniel Szeredi

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics

New York University

September, 2016

Gillian Gallagher

Dedication

A családnak

Dla rodziny

Acknowledgements

Well here we go — it seems that I have indeed finished this dissertation! This could not have happened without the help of so many people that surrounded me through these six years, and definitely not without the people who supported me through the preceding 27 years. However uncomfortable it might be for most of them, I will give acknowledgements where acknowledgements are due!

First and foremost, I am very thankful for the help and constant guidance of my advisor, Gillian Gallagher. It is very clear that without our weekly meetings and Gillian's advice I would not be here wrapping up my dissertation, not even close. Trying to imitate just a little bit how Gillian works does magic to one's productivity, but reaching her level seems impossible. She reads and gives detailed and insightful comments roughly on a tempo of 1 dissertation/2 hours, and is willing to spend literally half of January reading and fixing one's rushed grant proposal day after day after day, as history showed. Stay awesome, Gillian!

I am also thankful for the rest of my committee: their advice from the beginning right up to the defense was very valuable. Thanks to Lisa Davidson for helping me become a more rigorous phonetician (remaining errors are my own), and teaching me what to think about AX experiments. Thanks to Maria Gouskova for helping me become a more rigorous phonologist, and for the amazingly detailed comments on my drafts (sorry about all, the commas). Thanks to Susannah Levy for being on my committee and giving me such great advice — why haven't I asked you much more? And of course thanks to Frans Adriaans for all those great discussions after which I always wanted to sit down and implement something cool right away. Also a lot of thanks to John Alderete, Michael Becker, Andries Coetzee, Paul DeLacy, Edward Flemming, Florian Jaeger, Lauren Hall-Lew, John Kingston, Ruben van de Vijver, Rachel Walker, Alan Yu and Kristine Yu for great comments and suggestions about the research in this dissertation.

I was very fortunate to be able to be advised by Alec Marantz for my second Qualifying Paper; learning from Alec is an amazing experience, for further elaboration, see Kastner (2016). Stephanie Harves was a great mentor to me when dealing with diverse professional matters and when dealing with the personal side of these professional matters. Thanks to Chris Barker for the witty discussions about semantics, continuation monads and the department. I was really lucky to have had the opportunity to teach with Lucas Champollion, Brian McElree, Liina Pylkkänen and Anna Szabolcsi (köszönöm a húsvéti sonkát és még oly sok mindent!). It was amazing to learn about and work on sociolinguistic questions with Gregory Guy and John Singler. Thank you to Teresa Leung and Aura Holguin for helping me whenever I had any kind of issue and ran down in panic to the 2nd floor!

I am very lucky to have worked with Tara McAllister-Byun through most of my graduate student carreer. Tara is the best boss one can wish for, not only when one is overdrafted (thanks again!). My consultantship with Tara resulted not only in academic publications in my name, and in a confidence boost at every successful release/bug fix/acceptance, but also in a non-academic skillset I hope to be able to use for quite some time. Thank you also to Druanne Dillon for all the administrative help with this job (it was really really a lot)!

I would not have been even able to get anywhere near a linguistics PhD without the support of so many linguists in Hungary. The guidance and friendship of Péter Rebrus, Miklós Törkenczy, Péter Siptár, Zsuzsanna Bárkányi, Zoltán Kiss was what really prepared me for much of grad school, and, oh well, life in general. Thanks as well to Zoltán Bánréti, Huba Bartos, Kinga Gárdai, Beáta Gyuris, László Kálmán, Ferenc Kiefer, Katalin É. Kiss, András Komlósi, Ádám Nádasdy and Oravecz Csaba for all those years of highest quality education I got back home! And a huge thank you to Katalin Mády for letting me use the lab in Budapest!

At NYU I had the chance to work and spend my time with some of the smartest and, come on, the best people I've known. I have had amazing officemates: thank you so much Katie Wallace for all the fun in 511, in 613 and around the metropolitan area in general! Thanks Dunja Veselinović for (I would write putting up with me but that sounds negative and see I'm learning a little) setting the tone for the office, with syntax, phonology, Crispy Bacon and WordHero of course. Thanks to Adina Williams for all those hours of combined work-discussion sessions basically every day the last two years, and to Tim Mathes for our venting sessions back in the day! And of course cheers to my cohort: Sonia Kasyanenko, Jeremy Kuhn (Hungarian *does* exist, my friend!), Nate LaFave, Tal Linzen and Allison Shapp for the years of friendship.

I'm thankful for all the people I could talk to about basically anything I wanted to and to all friends I have made here at NYU. A big köszi to Ildi Szabó for all the proofreading and soup, and to Sean Martin for the amazing advice, discussions and help that he provided for mysteriously no money charged. Thanks to Suzy Ahn and Sang-Im Lee Nuna for the lovely discussions, the Korean food and the great friendships we have and to Nicole 'Trivia Master' Holliday for being there for both the ups and downs of these years. Cheers and hugs to all of my friends, centered around but not limited to Peculier/W3C/Thursday 6pm/Tuesday 8pm crew! Hugs to Isaac Bleaman, Teon Brooks, Dylan Bumford, Dan Duncan, Masha Esipova, Paloma Jeretić, Itamar Kastner (the A/C saved my life, thanks!), Maria Kouneli, Becky Laturnus, Nick MacDonald, Yining Nie, Yohei Oseki, Natalie Povilonis de Vilchez, Mary Robinson, Philip Shushurin, Sheng-Fu Wang, James Whang, Masha Westerlund, Linmin Zhang, and last, but honestly not least by far, Vera Zu!

My upperclasswomen and -men were so nice to me for years, they really helped me in so many ways navigate my way through the maze we call grad school. Thanks to Carina Bauman, Vincent Chanethom, Stacy Dickerman, Amanda Dye, Danny Erker, Tricia Irwing (oh those meetings!), Tim Leffel, Kim Leiken (Go Browns!), Inna Livitz (my open house host!), Txuss Martin (my neighbor!), Tuuli Morrill, Neil Myler (our MorphBeer performances were always inspiring!), Luiza Newlin-Łukowicz, Emily Nguyen, Kevin Roon, Salvador Mascarenhas (the great Etymology class we taught!), Cara Shousterman and Jim Wood! I got to know quite a few aspiring linguists who were or are now outside of NYU as well, and the discussions and support I received from them was an important part of how I got here: thanks Andrei Antonenko, Ruth Brillman, Dustin Chacón, Eleanor Chodroff, Megan Czvekus, Allyson Ettinger, Steven Foley, Néha Hirve, Luca Iacoponi, Greg Johnson, Christen Madsen, Claire Moore-Cantwell, Jason Ostrove, Deniz Özyıldız, Chris Parles, Rob Pasternak, Amanda Rysling, Poppy Slocum, Neda Todorović (hát hát), Matthew Tyler, and Aydoğan Yanılmaz.

I have to bring Katie Wallace back again and thank her again — but that's okay, she deserves it. So thanks, Katie, for introducing me to so many great people (mostly via board games for which I'm super thankful particularly)! Jeanna Choo, Jonathan Coveney, Sam Gross, Cassia Martin, Tim Trueman and Zack Yang, your support was something really important during this last year both in finishing my dissertation and in figuring out my future — I'm really grateful for all.

My friends from way back home did give me strength throughout the years. What would I be without my co-Elfonists, Márton Sóskuthy and Péter Rácz anyway? Much sadder and less published for sure! Sylvia Blaho was also there for me personally and professionally for years before and into grad school, thanks for everything! My hugs to all "linguist friends" back home, Bálint Biczók, Amanda Cardoso, László Fejes, Bálint Feyér, Ágnes Füle (nem! igen!), Anna Gazdik, Ágnes Gyarmati, Zsófi Gyarmathy, Lilla Magyar, Cecilia Molnár, Kristóf Molnár, Márta Peredy, András Schepácz, Balázs Surányi, Ádám Szalontai, Viktor Trón, Barbara Ürögdi, Károly Varasdi and János Zsitvai. Thanks to the folks I tried to learn so much NLP smartness from, Péter Halácsy, Gábor Recski, András Rung, István Szakadát and Dániel Varga. And of course, cheers to everyone in VPG 94.b, the best high school class I've been part of!

You can navigate your life looking for friends and advisors who are the best for you, but there are people in your life, whom you do not choose. I was never asked what kinds of parents I would like to have. The fact that my parents are who they are and how they loved me and supported me every year, month, day, hour is the closest thing I have as a proof for God's existence in my life. Anyu, Apu, ez a siker nem csak az enyém, ez mi hármunké közösen. Különben is, ha Ti velem, ki ellenem?

Abstract

Vowel harmony has been of great interest in phonological research (eg. Clements and Sezer 1982, Kaun 2004, Nevins 2004, Hayes et al. 2009). It has been widely accepted that vowel harmony is a phonetically natural phenomenon (Fowler 1983, Kaun 2004, Linebaugh 2007), which means that it is a common pattern because it provides advantages to the speaker in articulation and to the listener in perception.

Exceptional patterns proved to be a challenge to the phonetically grounded analysis (Benus and Gafos 2007, Hayes et al. 2009) as they, by their nature, introduce phonetically disadvantageous sequences to the surface form, that consist of harmonically different vowels. Such forms are found, for example in the Finnish stem *tuoli* 'chair' or in the Hungarian suffixed form *hiːd-hoz* 'to the bridge', both word forms containing a mix of front and back vowels. There has recently been evidence shown that there might be a phonetic level explanation for some exceptional patterns: (Benus and Gafos 2007) have shown the possibility that some vowels which participate in irregular stems (like the vowel [i] in the Hungarian stem *hiːd* 'bridge' above) differ in some small phonetic detail from vowels in regular stems. The main question has not been raised, though: does this phonetic detail matter for speakers? Would they use these minor differences when they have to categorize a new word as regular or irregular?

A different recent trend in explaining morphophonological exceptionality by looking at the phonotactic regularities characteristic of classes of stems based on their morphological behavior (Becker et al. 2011, Linzen et al. 2013, Gouskova et al. 2015). Studies have shown that speakers are aware of these regularities, and use them as cues when they have to decide what class a novel stem belongs to. These sublexical phonotactic regularities have already been shown to be present in some exceptional patterns vowel harmony, but many questions

remain open. One such question is how exactly learning the static generalization can be linked to learning the allomorph selection facet of vowel harmony? Also, how much does the effect of consonants on vowel harmony matter, when compared to the effect of vowel-to-vowel correspondences?

This dissertation aims to test these two ideas — that speakers use phonetic cues and/or that they use sublexical phonotactic regularities in categorizing stems as regular or irregular — and attempt to answer the more detailed questions, like the effect of consonantal patterns on exceptional patterns or the link between allomorph selection and static phonotactic generalizations as well. The phonetic hypothesis is tested on the Hungarian antiharmonicity pattern (stems with front vowels consistently selecting back suffixes, like in the example *hiːd-hoz* 'to the bridge' above), and the results indicate that while there may be some small phonetic differences between vowels in regular and irregular stems, speakers do not use these, or even enhanced differences when they have to categorize stems.

The sublexical hypothesis is tested and confirmed by looking at the pattern of mixed stems in Finnish. In Finnish, stems that contain both back and certain front vowels are frequent and perfectly grammatical, like in the example *tuoli* 'chair' above, while the mixing of back and some other front vowels is very rare and mostly confined to loanwords like *amatøøri* 'amateur'. It will be seen that speakers do use sublexical phonotactic regularities to decide on the acceptability of novel stems, but certain patterns that are phonetically or phonologically more natural (vowel-to-vowel correspondences) seem to matter much more than other effects (like consonantal effects).

Finally, a computational account will be given on how exceptionality might be learned by speakers by using maximum entropy grammars available in the literature to simulate the acquisition of the Finnish mixed stem and disharmonicity pattern. It will be shown that in order to clearly model the overall behavior on the exact pattern, the learner has to have access not only to the lexicon, but also to the allomorph selection patterns in the language.

Table of contents

List of figures

List of Tables

Chapter 1

Introduction

This dissertation focuses on irregularity patterns in vowel harmony, and discusses phonetic and lexical explanations on how speakers acquire exceptional patterns in their language.

This chapter will first introduce the general pattern of vowel harmony and discuss the main terms used in this dissertation about vowel harmony (Section 1.1). This will be followed by a discussion of the exceptional patterns (Section 1.2) found in vowel harmony systems that are going to be important throughout the dissertation. Then Section 1.3 will outline the ways speakers' biases and the sources of these biases can be tested going forward. Section 1.4 will review the literature on two sources of explanation for exceptional processes in vowel harmony: phonetic effects (Section 1.4.1) and sublexical phonotactic regularities (Section 1.4.2). The core claims of this dissertation will be laid out in Section 1.5 with the outline the main topics of the rest of the dissertation.

1.1 Vowel harmony

Vowel harmony is typically described as a principle found in many languages that all vowels within a word must share the same value for a language specific vocalic feature (Baković 2000). This principle can show up as two quite different patterns on the surface, which do not necessarily depend on each other synchronically. These two facets are "dynamic" harmony: affixes that alternate their surface form depending on the root, and "static" harmony: cooccurence restrictions on what vowels can be present within one root (Nevins 2004).

1.1.1 Static and dynamic harmony

1.1.1.1 Backness and rounding harmony

A classic example for vowel harmony is found in Turkish, where generally all vowels in a stem share their [±back] and [±rounded] feature values. For example (from Zimmer 1967):

(1) *somun* 'loaf'

(2) *arɯ* 'bee'

(3) *tytyn* 'tobacco'

(4) *iʎeti* 'message'

The stem forms above in (1–4) show perfectly grammatical forms, while front and back; or rounded and unrounded vowels cannot mix: **aly*, **pikø*. This grammaticality distinction is not categorical, though, and languages with vowel harmony tend to have exceptional stems which do not adhere to the general restrictions. These stems are usually either loanwords or compounds, as the following Turkish examples illustrate (Clements and Sezer 1982):

(5) *takvim* ‘calendar’

(6) *oteʎ* ‘hotel

(7) *mebus* ‘member of parliament’

The exceptionality and the instability of these stems can be seen in their tendency to regularize (Polgárdi 1998), especially if a vowel harmony violating vowel is not in the five vowel system /a,e,i,o,u/ (Clements and Sezer 1982):

(8) *komynizim* ∼ *kominizim* ‘communism’

(9) *pyro* ∼ *puro* ‘cigar’

(10) *biscyvit* ∼ *byscyvyt* ‘biscuit’

The dynamic aspect of vowel harmony is also visible in Turkish: suffixes have back, front, and rounded, unrounded allomorphs. The first person plural possessive and the accusative case suffixes have all four allomorphs with the four high vowels of the language [u], [ɯ], [y] and [i]:

(11) *somun* ‘loaf’ ∼ *somun-umuz-u* ‘loaf-1Pl-Acc’, **somun-ymyz-y*, **somun-imiz-i*, **somun-ɯmɯz-ɯ*

(12) *arɯ* ‘bee’ ∼ *arɯ-mɯz-ɯ* ‘bee-1Pl-Acc’, **arɯ-muz-u*, etc.

(13) *tytyn* ‘loaf’ ∼ *tytyn-ymyz-y* ‘tobacco-1Pl-Acc’, **tytyn-umuz-u*, etc.

(14) *iʎeti* ‘message’ ∼ *iʎeti-miz-i* ‘message-1Pl-Acc’, **iʎeti-mɯz-ɯ*, etc.‘

However, Turkish does show some differences between the static and dynamic facets of vowel harmony: only *high* vowels in suffixes alternate according to the [±rounded] feature. The low vowels in the suffixes do not alternate according to rounding, as illustrated below with the locative suffix that has two allomorphs -*da* and -*de*, but no *-*do* or *-*dø*:

(15) *somun* 'loaf' ~ *somun-da* 'loaf-Loc', **somun-do*

(16) *arɯ* 'bee' ~ *arɯ-da* 'bee-Loc'

(17) *tytyn* 'tobacco' ~ *tytyn-de* 'tobacco-Loc', **tytyn-dø*

(18) *iʎeti* 'message' ~ *iʎeti-de* 'message-Loc'

The differences between the static restrictions and the allomorphy pattern are language specific. While, As seen above, Turkish does not involve non-high vowels in rounding harmony in the case of alternating suffixes, the related Turkic language Kirghiz differs in this exact regard: non-low vowels participate in rounding harmony as well (Korn 1969, Kaun 2004), although rounded /o/ in suffixes only occurs after another /o/ (Mailhot and Reiss 2007):

(19) *taʃ* 'stone' ~ *taʃ-ka* 'stone-Dat'

(20) *iʃ* 'job' ~ *iʃ-ke* 'job-Dat'

(21) *konok* 'guest' ~ *konok-ko* 'guest-Dat'

(22) *køz* 'eye' ~ *køz-kø* 'eye-Dat'

(23) *yj* 'house' ~ *yj-kø* 'house-Dat'

(24) *but*: *uʧ* 'tip' ~ *uʧ-ka* 'tip-Dat', not **uʧ-ko*

1.1.1.2 ATR harmony

There is another vocalic feature besides [±back] and [±round], which frequently occurs in vowel harmony patterns: the feature commonly referred to as advance tongue root, represented as [±ATR]. This feature whose acoustic effect is a difference between more ([+ATR], like /i, e, ə, o, u/) and less ([-ATR], such as /ɪ, ɛ, a, ɔ, ʊ/) peripheral vowels is the harmonizing feature in many Nilo-Saharan and Niger-Congo languages in Africa (Casali 2008), as well as in some East Asian language families like Mongolic and Tungusic, and ATR harmony can also be observed in many other languages not centered on a certain geographical location. A case of an ATR harmony with seemingly no neutrality is found in Kasem (Callow 1965). The same value of the [±ATR] feature is found in stems, and affixes do show allomorphy according to the stem's harmonicity:

(25) *ʧan-a* 'moon' ∼ plural *ʧan-ɪ*, not **ʧan-i*

(26) *bəkəd-ə* 'boy' ∼ plural *bəkəd-i*, not **bəkəd-ɪ*

(27) *mɪmɪn-a* 'thin' ∼ plural *mɪmɪn-ɪ*, not **mɪmɪn-i*

(28) *kukud-ə* 'dog' ∼ plural *kukud-i*, not **kukud-ɪ*

Khalkha Mongolian also shows ATR harmony. The [+ATR] class consists of /u, o, e/, while the [-ATR] class is /ʊ, ɔ, a/, vowels which are consistently lower and more retracted than [+ATR] vowels and are articulated with a stronger pharyngeal constriction (Svantesson et al. 2005). The phoneme /i/, while phonologically behaving as a neutral vowel (see Section 1.1.3), has a [+ATR] allophone [i], and a [-ATR] allophone [ɪ], an alternation that is conditioned by the previous syllable. In initial syllables, /i/ always counts as [+ATR]:

(29) *saːrmag* 'neutral'

(30) *xɔjɔr* 'two'

(31) *atʲɪ* 'like'

(32) *mini* 'mine'

For allomorphy in suffixes, high vowels show ATR vowel harmony, so /u/ alternates with /ʊ/:

(33) *it-* 'eat' ∼ *it-u* 'eat-CAUS'

(34) *uts-* 'see' ~ *uts-u* 'see-CAUS'

(35) *og-* 'give' ~ *og-u* 'give-CAUS'

(36) *χʊnʲ-* 'pleat ~ *χʊnʲ-ʊ* 'pleat-CAUS'

(37) *jaw-* 'go' ~ *jaw-ʊ* 'go-CAUS'

(38) *ɔr-* 'enter' ~ *ɔr-ʊ* 'enter-CAUS'

The two allophones of /i/ naturally are also conditioned based on the stem vowel quality:

(39) *suχ-* 'axe' ~ *suχ-ig* 'axe-ACC'

(40) *sʊm-* 'arrow' ~ *sʊm-ɪg* 'arrow-ACC'

Non-high vowels also alternate according to rounding, additionally to [±ATR]. Rounded allophones are only triggered by other non-high rounded vowels (/o, ɔ/), the high rounded vowels /u, ʊ/ select unrounded allophones:

(41) *it-* 'eat' ~ *it-e* 'eat-DPAST'

(42) *uts-* 'see' ~ *uts-e* 'see-DPAST'

(43) *og-* 'give' ~ *og-o* 'give-DPAST'

(44) *χʊnʲ-* 'pleat ~ *χʊnʲ-a* 'pleat-DPAST'

(45) *jaw-* 'go' ~ *jaw-a* 'go-DPAST'

(46) *ɔr-* 'enter' ~ *ɔr-o* 'enter-DPAST'

The goal of this section so far has been to illustrate how the most typical vowel harmony systems work. It has also been shown that in most cases, as in Khalkha seen above, more vocalic features interact to produce the surface pattern. Below, the effect of consonantal participation will be discussed.

1.1.2 Consonantal participation in harmony

In some languages with vowel harmony some consonants alternate according to the harmonicity of the stem they are found in. For example, besides the exceptions discussed in Section 1.2.4, Turkish /k/, /g/ and /l/ have two allophones: [k], [g] and [l] (or [ɫ], Levi 2002) in back word forms and [c], [ɟ] and [ʎ] in front word forms:

(47) *kɯz* 'girl' *vs.* *cøj* 'village'

(48) *gar* 'railway station' *vs.* *ɟeʧ* 'late'

(49) *lastik* 'rubber' *vs.* *ʎeʃ* 'carcass'

Similarly, in Khalkha Mongolian, the velar consonant phonemes /ŋ, x, ɡ/ alternate with their uvular allophones [ɴ, χ, ɢ] in [-ATR] contexts:

vowel	+ATR		-ATR	
consonant	velar		uvular	
/ŋ/	*xaːŋ*	'Khan'	*nɔɢɔɴ*	'green'
/x/	*xeːʓ*	'to decorate'	*χʊn^j*	'to pleat'
/ɡ/	*og*	'to give'	*tʊɢʊi*	'circle'

Table 1.1: Velar-uvular consonant alternation corresponding to [±ATR] harmony in Khalkha Mongolian

The explanation by Svantesson et al. (2005) for this alternation is based on a [±pharyngeal] feature that both [-ATR] vowels and uvular consonants share. Alternatively, if ATR harmony evolved from frontness harmony in Mongolian, then the velar-uvular alternations corresponded to front-back contexts, which would have been a phonetically natural coarticulatory process between consonant tongue root and vowel tongue root articulations.

1.1.3 Neutral vowels

Many vowel harmony systems involve a subclass of vowels that are treated as *neutral* in the pattern. Neutrality means that a certain vowel is not limited to cooccur with only one harmonic class in the language, but can be freely mixed with other vowels, some of which can possibly have a different phonetic value for the feature underlying the harmony in the given language. Neutral vowels can be classified as opaque or transparent. A vowel is *opaque* in a vowel harmony system if its own harmonicity takes precedence over the harmonicity of the preceding (for left-to-right spreading) or following (right-to-left) vowel when triggering allomorphy further outward in the phonological word. Take the following example from Chicheŵa (Rhodes 2010, originally from Mtenje 1985, Harris 1994), a language with height harmony:

(50) *put-its-a* 'provoke-CAUS-V'

(51) *bal-its-a* 'give birth-CAUS-V'

(52) *konz-ets-a* 'correct-CAUS-V' (not **konz-its-a*)

(53) *put-an-its-a* 'provoke-RECIP-CAUS-V'

(54) *konz-an-its-a* 'correct-RECIP-CAUS-V' (and not **konz-an-ets-a*)

These examples show that /a/ is opaque in Chicheŵa: the language has a height harmony system, with the suffix *-its*∼*ets-* surfacing with a high vowel after high and low vowels (50–51) and with a mid vowel after mid vowels (52). Suffixes with [a] like the verbal *-a* and the reciprocative *-an-* above do not alternate, and suffixes after these morphemes always use the allomorph used after non-mid vowels (53–54) even if there is a mid vowel in the preceding stem (54).

A vowel is treated as *transparent* in a vowel harmony system if it does not influence the harmonicity of the word it occurs in for further allomorph selection. In harmony that targets the [±back] feature, high front unrounded vowels tend to behave this way. This phenomenon can also manifest both as a static phonotactic generalization or as an effect on allomorph selection. For example, Finnish contains many stems where the non-low front unrounded vowels /e,i/ mix with back vowels:

(55) *tuoli* 'chair'

(56) *aarre* 'treasure'

(57) *kerho* 'club'

(58) *lintu* 'bird'

The transparency of /i/ and /e/ can be illustrated by the following examples, where Finnish stems below take back suffixes despite the intervening phonetically front vowels:

(59) *tuoli* 'chair' ∼ *tuoli-lla* 'chair-ADESS', not **tuoli-llæ*

(60) *ovi* 'door' ∼ *ove-lla* 'door-ADESS', not **ove-llæ*

There are similar patterns in other vowel harmony types as well. As it is the case with frontness harmony and non-low unrounded front vowels, the low vowel /a/ is very frequently neutral in many languages with ATR harmony (Casali 2008). So similarly to Finnish /i/, in Akan, the vowel /a/ occurs both with [+ATR] as well as [-ATR] vowels in the stem (O'Keefe 2004):

(61) *adɪ* 'thing'

(62) *adi* 'outside'

(63) *mog*ʲ*a* 'blood'

(64) *asɔrɪdan* ‘chapel’

While affixes with vowels other than /a/ have a [+ATR] and a [-ATR] allomorph (65–66), /a/ does not alternate and behaves as an opaque vowel triggering [-ATR] harmony to further suffixes (67–68):

(65) *di* ‘eat’ ∼ *mi-di* ‘1Sg-eat’

(66) *dɪ* ‘be called’ ∼ *mɪ-dɪ* ‘1Sg-be called

(67) *tu* ‘dig up’ ∼ *ɔ-a-tu* ‘1Sg-Perf-dig up’

(68) *tɔ* ‘bake’ ∼ *ɔ-a-tɔ* ‘1Sg-eat’

Khalkha Mongolian, on the other hand does not treat /a/ as neutral, rather it is the high front vowel /i/ which behaves as transparent. This vowel has two allophones, [i] and [ɪ] according to the ATR feature value of the preceding vowel, but its non-phonemicity and transparency becomes apparent in the following examples, where the /i/ does not influence the allomorph selection of subsequent affixes (Svantesson et al. 2005):

(69) *piːr* ‘brush’ ∼ *piːr-ig-e* ‘brush-Acc-Refl’

(70) *poːr* ‘kidney’ ∼ *poːr-ig-o* ‘kidney-Acc-Refl’

(71) *mʊːr* ‘cat ∼ *mʊːr-ɪg-a* ‘cat-Acc-Refl’

(72) *χɔː* ‘food’ ∼ *χɔː-ɪg-ɔ* ‘food-Acc-Refl’

The reason for /i/ to be transparent in the ATR harmony system of Khalkha has been argued that the ATR harmony developed from an earlier frontness harmony system with the backing of /y/ and /ø/ to /u/ and /o/, and the pharyngealization of original /o/ and /u/ to /ɔ/ and /ʊ/ respectively (Svantesson et al. 2005). This would mean that in Classical Mongolian, the front harmonic vowels were /e, y, ø/, the back harmonic vowels were /a, o, u/ and a neutral vowel was /i/, a system very similar to the Finnish pattern, and found in the related Oirat Mongolian dialects. However, based on typological and historical reasons Seongyeon (2011) argues that the language change was in the opposite direction (original ATR harmony turned into frontness harmony in Oirat, and stayed unchanged in Khalkha).

The section above summarized how most vowel harmony systems look like. Having set the table, the following section will discuss the most typical exceptionality patterns.

1.2 Exceptionality patterns

There are several types of exceptionality patterns found in vowel harmony systems. Having catalogued the main facets of vowel harmony in the previous section makes categorizing the exceptionality patterns easier. Exceptionality from the static facet of vowel harmony, which can be analyzed as two related phenomena called **mixed stems** and **disharmonicity** based on how well exceptional stems fit in the grammar will be discussed in Section 1.2.1. Exceptionality from the allomorph selection side can also be further categorized into two different phenomena: exceptional stems that select an allomorph that does not match their harmonicity (**antiharmony**, Section 1.2.2); and exceptional suffixes that do not alternate according to vowel harmony, while similar suffixes have different harmonic allomorphs (**suffixal exceptionality**, Section 1.2.3). Finally, Section 1.2.4 will illustrate some patterns where **consonantal** alternations have an **effect** on stems that make them either disharmonic or at times antiharmonic.

1.2.1 Mixed stems and disharmony

Most, if not all languages with vowel harmony do have stems where vowels from different harmonic classes are found. Such mixed stems can be perfectly harmonic, and not count as exceptional if the violating vowels are neutral, such as in the following stems in Finnish:

(73) *kaksi* ‘two’

(74) *tule*- ‘to come’

The grammaticality of these *mixed stem* forms is usually analyzed by calling some of the vowels regularly occuring in such stem neutral with regards to vowel harmony. In Finnish, for example, *i* and *e* are both neutral, and mixed stems with back vowels and these neutral vowels are as grammatical as any back-only or front-only stem.

Stems that can surely be called disharmonic — that is, stems with non-neutral vowels from different harmonic classes — have a gradient level of ungrammaticality. Disharmonic stems are usually either loanwords from languages without vowel harmony that had not adapted to the phonotactic regularities of the loaning language (75–77), or compounds (78–79) made from stems with different harmonicities (Campbell 1980, Ringen and Heinämäki 1999):

(75) *amatøøri* ‘amateur’

(76) *analyysi* ‘analysis’

(77) *tyranni* ‘tyrant’

(78) *silmæ+puoli* ‘eye+half = one-eyed’

(79) *avo+kæt-inen* ‘open+hand-ADJ = open-handed’

Khalkha Mongolian also contains several disharmonic stems, which are mostly recent Russian, Chinese or English loanwords; and also disharmonic compounds (Puthuval 2013):

(80) *signaːʤ* ‘car horn’

(81) *k^hursɔːr* ‘cursor’

(82) *awʲaː+tsui* ‘sound+ology = phonetics’

(83) xox+xOt^{h} ‘blue+city = Hohhot’

Speakers seem to be influenced by the highly exceptional nature of disharmonic stems. This is evidenced by the frequent process of eliminating such violations in faster or more casual speech:

(84) Finnish *olympia-laise-t* ‘Olympic-ADJ-PL’ (meaning ‘Olympic Games’) ∼ *olumpia-laise-t* or *ølympiæ-læise-t* (Campbell 1980)

(85) Finnish *polityyri* ‘shellac varnish’ ∼ *pulituuri* (Kiparsky and Pajusalu 2003)

(86) Finnish *dynamiitti* ‘dynamite’ ∼ *tinameetti* or *tynæmiitti* (Duncan 2008)

(87) Turkish *komynizim* ‘communism’ ∼ *kominizim* (Clements and Sezer 1982)

(88) Turkish *cysur* ‘fractions’ ∼ *cysyr* (Clements and Sezer 1982)

Other evidence for the unacceptability of disharmonic stems as grammatical comes from several psycholinguistic studies. Suomi et al. (1997) have shown that Finns use vowel harmony as an important cue for word segmentation: without any cues for word stress, Finnish speakers are more likely to perceive *hymy* as a word in a disharmonic nonce sequence *puhymy* than in a harmonic sequence *pyhymy*. Vroomen et al. (1998) showed that this effect is language-specific: vowel harmonicity cannot be used as a cue in languages which do not have vowel harmony, such as Dutch and French. The preference for harmonicity has also been shown by Arik (2015) for Turkish stems and stem suffixation.

While there is much evidence for the psychological reality of harmony being preferred to disharmony by the speakers of vowel harmony languages (Clements and Sezer 1982, Suomi et al. 1997), the exact details on how different kinds of disharmonic stems are evaluated by the speakers and how much consonantal effects interact with mixed stems and disharmony deserves further investigation. Chapter 3 will discuss Finnish mixed stems from a lexical point of view and will try to assess how much Finnish speakers learn the phonotactic regularities of mixed stems. Then Chapter 4 will attempt to build a learning model using Maximum Entropy grammars that would be able to both replicate Finnish speakers' judgments on different kinds of mixed and disharmonic stems and would be able to judge fully disharmonic stems as ungrammatical.

1.2.2 Antiharmony

Several vowel harmony systems also have an exceptional subpattern, where certain stems consistently select suffixes of a different harmonicity than what the vowels in the stem would predict. This might be limited to only certain suffixes, like the following two stems in Finnish, which despite their phonetically front stem vowels (neutral /e/) take the back allomorph of the partitive suffix, but the front allomorph of all other suffixes:

(89) *meri* 'sea' ∼ *mer-ta* 'sea-PART', but *mere-ssæ* 'sea-INESS'

(90) *veri* 'blood' ∼ *ver-ta* 'blood-PART', but *vere-ssæ* 'blood-INESS'

In other languages, like Hungarian, there are many more antiharmonic stems, containing only neutral, phonetically front vowels (/ɛ, eː, i, iː/), yet still consistently selecting back allomorphs of all suffixes:

(91) *ʃiːr* 'grave' ∼ *ʃiːr-hoz* 'grave-ALLAT', *ʃiːr-ok* 'grave-PL', *ʃiːr-om* 'grave-1SG', not **ʃiːr-hɛz*, **ʃiːr-ɛk*, **ʃiːr-ɛm*

(92) *dɛreːk* 'waist' ∼ *dɛreːk-nɒk* 'waist-DAT', *dɛrɛk-ɒʃ* 'waist-ADJ', *dɛrɛk-ɒm* 'waist-1SG, not **dɛreːk-nɛk*, **dɛrɛk-ɛʃ*, **dɛrɛk-ɛm*

This Hungarian antiharmonicity pattern has been quite stable. Even the earliest records show established antiharmonicity:

(93) Old Hungarian *biːr-ʃaːg+nop* 'judge-NOM+day = judgment day' (1195, Funeral Sermon and Prayer) (not **biːr-ʃeːg+nop*)

Most earlier literature (eg. Vago 1980, E. Abaffy 2004a) proposes that antiharmonicity is derived from the fronting of an earlier /ɯ/ phoneme. This would mean that stem like *biːr-* would have originally been **bɯːr-*, therefore taking regularly back suffixes, and the later merger or /ɯ/ with /i/ would have lead to the current antiharmonic pattern. Kis (2005) has shown that the arguments for the existence for an earlier /ɯ/ are weak and circular. Several antiharmonic stems have front /i/ correspondents in relative or in the originating languages (eg. *hiːd* 'bridge' ∼ *hiːd-nɒk* 'bridge-DAT'; original Alanic *χiːd*), and while the existence of the pattern has been stable, there have been stems that switched category in either way diachronically (eg. *hiːm* 'male' is harmonic today, but was antiharmonic up until the 17th century; *ʃiːr* 'grave' is antiharmonic today, but was harmonic originally). The conclusion is that antiharmonicity is an old phenomenon, possibly already present in Proto-Finno-Ugric, and while the membership of the antiharmonic subset of the lexicon can change diachronically, the existence of this phenomenon is very stabile.

Similarly, the following Turkish stems take back suffixes in certain more archaizing idiolects (Clements and Sezer 1982) despite the front vowels in them, mostly owing to the original /q/ in Arabic that was borrowed to Turkish as [k], or the presence of a pharyngealized consonant, like /tˤ/ triggering back harmony:

(94) *ʃevk* 'desire' ∼ *ʃevk-ɯ* 'desire-ACC', not **ʃevk-i*

(95) *haːʎik* 'creator' ∼ *haːʎik-ɯ* 'creator-ACC', not **haːʎik-i*

(96) *utarit* 'Mercury' ∼ *utarid-ɯ* 'Mercury-ACC', not **utarid-i* (cf. Arabic *ʕutˤaːrid*)

Conversely, the following Turkish stems take *front* suffixes despite their stem vowels triggering back suffixation by default, owing to the non-pharyngealized consonants present in the Arabic originals of these stems triggered the vowels to be more fronted. Some of these include the marginal phoneme /rʲ/:

(97) *harʲp* ∼ *harʲb-i*, not **harʲb-ɯ*

(98) *saat* 'hour' ∼ *saat-i*, not **saat-ɯ*

(99) *kɯraat* 'reading' ∼ *kɯraat-i*, not **kɯraat-ɯ*

(100) *rab* 'God' ∼ *rabb-i*, not **rabb-ɯ*

The stability of the antiharmonicity in Hungarian contrasts with the instability of this pattern in Turkish. The antiharmonic phonetically front stems that take back suffixes seem to have been regularized for younger speakers (Clements and Sezer 1982). The phonetically back stems with front suffixes either have a palatal consonant [rʲ] in them, or for some speakers a different consonant can also become palatalized (eg. *saac* 'hour'), resulting in a segmental level trigger for the antiharmonic behavior.

While antiharmonicity does not seem to be found in contemporary Khalkha Mongolian, there were stems with /i/ in Old Mongolian that took back (or [-ATR]) suffixes (Svantesson et al. 2005):

(101) *nis-* 'to fly' ∼ *nis-qa-ba* 'fly-Caus-Pst'

(102) *niʃi-* 'to beat' ∼ *niʃi-tʰʊχaj* 'beat-Impf'

These roots are regularized as harmonic in Khalkha Mongolian:

(103) *nis-ʤe* 'fly-DPst', not **nis-ʤa*

Spoken Oirat Mongolian, which has [±back] harmony instead of [±ATR] also has antiharmonic stems. In these stems, an original back vowel was palatalized by a following /i/ which was subsequently lost (Birtalan 2003):

(104) *æːl* 'camp' ∼ *æːl-ar* 'camp-INSTR'

(105) *øːrt-* 'to come closer' ∼ *øːrt-ul* 'come closer-CAUS'

Antiharmonicity presents a significant challenge for all explanations of vowel harmony. The frequent occurrence of this pattern poses important questions to analyses that would treat vowel harmony as a direct consequence of the phonetic properties of words. Antiharmonicity creates surface forms that are dispreferred acoustically and articulatorily, and go against the majority of stems and the productive patterns in the lexicon. Chapter 2 will discuss antiharmonicity in depth, focusing on the Hungarian pattern.

1.2.3 Suffixal exceptionality

Several vowel harmony languages have suffixes that do not participate in vowel harmony. For example, While the vast majority of Turkish suffixes do alternate according to the harmonicity of the stem, some suffixes do not harmonize. One type of such suffixes are borrowings from other languages, like the suffix *-istan* to form country names:

(106) *ermen-istan* 'Armenia', **ermen-isten*

(107) *mool-istan* 'Mongolia', **mool-ɯstan*

There are also native non-participating suffixes in Turkish. The second vowel of the present progressive suffix *-ujor/-ɯjor/-yjor/-ijor* does not participate, and the /o/ of the suffix blocks the spreading of harmony, forcing subsequent suffixes to select their back round allomorphs (Kardestuncer 1983, Polgárdi 1998):

(108) *ɟeʎ-* 'to come' ∼ *ɟeʎ-ijor-um* 'I am coming'

(109) *koʃ-* 'to run ∼ *koʃ-ujor-um* 'I am running'

(110) *ɟyʎ-* ‘to laugh’ ∼ *ɟyʎ-yjor-um* ‘I am laughing’

(111) *bak-* ‘to look’ ∼ *bak-ɯjor-um* ‘I am looking’

This peculiarity can be explained by positing that there is a different kind of boundary before this suffix, making it akin to a clitic or a compound stem (Kardestuncer 1983). Arguments for this hypothesis involve the unstressability of the suffix (unlike other, non-clitic suffixes) and the history of this suffix as a recently cliticized verb (Lewis 1967).

The Khalkha Mongolian negative suffix *-guj* does not participate in the ATR vowel harmony of the language either (Binnick 1991), as seen below in (112–113). However, it is not opaque, like the Turkish continuous suffix, but it is transparent, meaning that the suffixes that follow it alternate according to vowel harmony (114–115):

(112) *javsan-guj* ‘did not go’

(113) *ogson-guj* ‘did not give’

(114) *javsan-guj-gaːr* ‘by not going’

(115) *ogson-guj-goːr* ‘by not giving’

Similar, diachronically recently cliticized suffixes also appear in Hungarian, for example the temporal suffix *-kor* ‘at’ (Polgárdi 1998):

(116) *hɒt* ‘6’ ∼ *hɒt-kor* ‘at 6 o’clock’

(117) *heːt* ‘7’ ∼ *heːt-kor* ‘at 7 o’clock’, not **heːt-kør*

Diachronic evidence suggests that recently cliticized suffixes tend to regularize eventually. In Old Hungarian (ca. 10–11th centuries) many postpositions cliticized onto nouns, forming new cases. These suffixes originally did not participate in vowel harmony (118–120), but in

Modern Hungarian (besides some peripheral Western dialects, Kiss 2001) they adapted to the vowel harmony system (121–123):

(118) Old Hungarian *paraditʃum-bɛn* ‘in paradise’ (1195, Funeral Sermon and Prayer)

(119) Old Hungarian *myɲi urusaːg bɛleː* ‘into heaven’ (id.)

(120) Old Hungarian *halaːl-nɛk* ‘death-DAT’ (id.)

(121) Modern Hungarian *pɒrɒditʃom-bɒn* ‘in paradise’, not **pɒrɒditʃom-bɛn*

(122) Modern Hungarian *mɛɲorsaːg-bɒ* ‘into heaven’, not **mɛɲorsaːg-bɛ*

(123) Modern Hungarian *hɒlaːl-nɒk* ‘death-DAT’, not **hɒlaːl-nɛk*

Finally, some suffixes, like the first person singular indefinite conditional verbal suffix *-neːk* (eg. *neːz* ‘to look’ ∼ *neːz-neːk* ‘I would look’) in Standard Hungarian (Rebrus and Törkenczy 2005), do not alternate while there is no reason to analyze them as clitics, as is evident from the comparison with similar suffixes as seen on Table 1.2. A speculative reason for the non-alternating nature of this suffix is that it would be homophonous with the 3rd person plural *definite* conditional suffix for all verbs, while the lack of alternation at least helps homophony avoidance for back stems:

	back stem	**front stem**
	laːt	*neːz*
	'to see'	'to look'
1Sg.Indef	*laːt-**neːk*** (!)	*neːz-**neːk***
2Sg.Indef	*laːt-naːl*	*neːz-neːl*
3Sg.Indef	*laːt-nɒ*	*neːz-nɛ*
3Pl.Indef	*laːt-naːnɒk*	*neːz-neːnɛk*
	...	
1Sg.Indef	*laːt-naːm*	*neːz-neːm*
2Sg.Indef	*laːt-naːd*	*neːz-neːd*
3Sg.Indef	*laːt-naː*	*neːz-neː*
3Pl.Indef	*laːt-**naːk***	*neːz-**neːk***

Table 1.2: Non-harmonizing suffix in the indefinite conditional verbal paradigm in Standard Hungarian

Exceptionality on the suffixal level is not as stable as within stems or in antiharmonicity. There are Turkish varieties where the *-(i/ɯ/u/y)jor* suffix assimilates to the rest of the suffix system yielding forms like *ɟel-ijer-im* 'I am coming' (Baski 2009, Demirci and Kuanysbaeva 2016). Similarly, the exceptionality of the first person indefinite conditional suffix in Hungarian is regularized in most spoken varieties of Hungarian by aquiring a back *-naːk* allomorph. The Hungarian case suffixes provide another evidence for the diachronic instability of suffixal exceptionality, as they started alternating according to vowel harmony soon after cliticizing on the noun (118–123).

1.2.4 Vowel harmony patterns influenced by consonants

As has been discussed in Section 1.1.2, consonants also participate in the vowel harmony driven alternations in some languages like Turkish and Khalkha Mongolian. This subsection will show that certain exceptionality patterns in vowel harmony systems are driven by, or they condition exceptional behavior of these consonants.

In Turkish, the phonemes /k, g, l/ alternate with their palatalized allophones [c, ɟ, ʎ] according to the frontness of the stem. However, due to loanwords mostly from Arabic and Persian, the palatal allophones can be treated as marginally contrastive. The following examples show the palatalized consonants surface in back harmonic stems (from Clements and Sezer 1982):

(124) *boʎ* 'cocktail, drink' (cf. *bol* 'abundant')

(125) *car* 'profit' (cf. *kar* 'snow')

(126) *ɟavur* 'infidel' (cf. *gaz* 'gas')

There are some heuristics to predict the occurence of [c] in place of [k]: in most loanwords, the [±back] feature value of the *tautosyllabic* vowel determines whether the consonant is palatalized or not. However, there are still exceptions from this rule (eg. *becar* 'bachelor', *caːtip* 'clerk').

The limited predictability of the [k]∼[c] alternation also influences the harmonic behavior of a given stem. The following Arabic loanwords show a pattern where the word final velar is not palatalized when not affixed. However, this velar alternates with a palatal before suffixes and influences the allomorph selection of the stem, so that the suffix is also front, regardless of the preceding back vowel:

(127) *idrak* 'perception' ∼ *idraːc-i* 'perception-ACC', not **idraːk-ɯ* or **idraːc-ɯ*

(128) *heʎak* 'exhaustion' ∼ *heʎaːc-i* 'perception-Acc', not **heʎaːk-ɯ* or **heʎaːc-ɯ*

The distribution of [l]∼[ʎ] is even more complicated: for example, word initial laterals are predictably palatal, with no regard to the harmonicity of the stem (eg. *ʎaf* 'expression'). The consonant also surfaces as [ʎ] invariably, if it has a front vowel on *either* side, even if there are intervening consonants:

(129) *kaʎem* 'pen'

(130) *iʎatʃ* 'drug'

(131) *aʎbym* 'album'

A stem internal palatal [ʎ] can also influence the allomorph selection of the stem, as it triggers front suffixation despite an otherwise back stem harmonicity even if it is not the stem final consonant (Clements and Sezer 1982, Levi 2002):

(132) *usuʎ* 'system' ∼ *usuʎ-y* 'system-Acc', not **usuʎ-u*

(133) *roʎ* 'role' ∼ *roʎ-y* 'role-Acc', not **roʎ-u*

(134) *kaʎp* 'heart' ∼ *kaʎp-i* 'heart-Acc', not **kaʎp-ɯ*

(135) *goʎf* 'golf' ∼ *goʎf-y* 'golf-Acc', not **goʎf-u*

As seen in Section 1.1.2 on Table 1.1, Khalkha Mongolian velars and uvulars are in complementary distribution, which is conditioned by the harmonic value of the stem they are found in. The voiced velars and uvulars /g/ and /ɢ/ partially form an exception from this (Svantesson et al. 2005), as while they do alternate in suffixes and word-initially (136–139), there are minimal pairs found stem-finally (140):

(136) *sux-ig* 'axe-Acc'

(137) *aχɪɢ* ‘elder brother-Acc’

(138) *gir* ‘house’

(139) *ɢar* ‘hand’

(140) contrast: *pag* ‘team’ *vs.* *paɢ* ‘small’

(141) upheld before suffixes as well: *pag-as* ‘team-Refl’ *vs.* *paɢ-as* ‘small-Refl’

1.2.5 Summary of exceptional patterns

This section discussed different kinds of exceptionality patterns in vowel harmony. These exceptionality patterns generally introduce word forms in the language that contain vowels that do not agree in the feature that is targeted in the given language. The previous subsections discussed four main ways this could happen: due to irregular allomorph selection driven by the stems (antiharmonicity), due to loanwords or other processes introducing some irregular forms to the lexicon (disharmonicity), due to irregular suffixal behavior and due to the influence of some consonantal patterns. It was also seen that the presence of certain vowels does not imply irregularity in the static lexical phonotactics, as these vowels can be classified as *neutral* and the stems they occur in are regular *mixed* stems, and not strictly disharmonic.

This dissertation will focus on antiharmonicity, disharmonicity and mixed stems. Antiharmonicity is a problematic phenomenon because it entails that there are certain lexical items, which apparently select allomorphs of subsequent affixes that go against all phonetic and phonological cues in the language. It is worthwile asking: how do speakers learn that these stems behave in an opposite way? Chapter 2 will discuss this question, and will falsify the hypothesis that there is something different in these stems in a segmental level.

Disharmonicity is problematic because speakers have to learn the highly non-local pattern that vowels, which share their values in the harmonizing feature are legal to mix in a word, and that mixing vowels of different values for this feature is not fully grammatical. The distinction between ungrammatical disharmonic forms and grammatical mixed stems also raises questions: how do speakers learn what vowels are allowed to mix with vowels having different harmonic values and which vowels are not? As this distinction is not necessarily categorical, since truly disharmonic stems are usually attested in the language, the question is whether speakers are able to distinguish between these two groups, and if they are, how do they learn this gradient pattern? Chapters 3 and 4 will address these questions.

1.3 Testing speakers on vowel harmony and exceptionality

1.3.1 Psychological reality of vowel harmony

The existence of dynamic harmony in a language is easily testable: speakers are able to create affixed word forms that obey the requirements of vowel harmony. Whenever a speaker wants to use an affixed form for a stem they have to decide on what allomorph to use, a decision that can be straightforwardly analyzed when describing the generalizations (as done by Clements and Sezer 1982, Hayes et al. 2009 amongst others). There are occasions where there is variation or speakers cannot form a categorical judgment but this variation is generally limited to a minor subset of the lexicon (Ringen and Heinämäki 1999).

The easy testability of dynamic harmony stands in contrast with static harmony: it is not immediately clear that speakers' grammars even contain the generalizations over the phonotactic distributions of the stem. As already discussed in Section 1.2.1, there are several types of evidence found for the psychological reality of disharmony. Speakers often regularize disharmonic stems by changing some vowel features (84–88), and vowel harmony helps speakers segment words (Suomi et al. 1997, Vroomen et al. 1998). It has also been supported with evidence that speakers replicate the lexical frequencies characteristic to certain subsets of the lexicon when tested on these patterns (Ernestus and Baayen 2006), and this has been shown to be true when these subsets of the lexicon are based on the harmonic behavior of stem vowels as well (Hayes et al. 2009).

1.3.2 Wug testing

The main experimental tool used in this dissertation to discover what kinds of biases restrict the search space for learners of exceptionality in vowel harmony, and what influences drive speakers' judgments and their allomorph selection is going to be wug (or nonce-word) studies. Nonce word studies have originally been designed to investigate productivity, like the productive application of the English plural suffix by 5–7 year old children (Berko 1958). Using this paradigm has been extended to investigate speakers' decisions in zones of variation in their language, including exceptionality in past tense formation (Bybee and Moder 1983, Prasada and Pinker 1993, Albright and Hayes 2003), allomorph selection (Gouskova et al. 2015, Gouskova and Ahn 2016) and underlying form identification for stems (Becker et al. 2011) amongst others. Studies like Finley (2012), Finley and Badecker (2012) and Baer-Henney et al. (2015) expanded the paradigm towards artificial grammar language learning experiments, where speakers of a language without a given pattern, like vowel harmony, are exposed to alternations of nonce words based on this pattern, in order to identify innate biases that speakers pay attention to when exposed to the given phenomenon.

Wug studies will be used in Chapters 2 and 3 to find out how much phonetic cues and sublexical information matter when speakers have to decide on the acceptability of a form. Both possible sources of bias will be tested by devising a set of stimuli where the crucial conditions separate the test items based on the tested source of information. To test how well phonetic cues are used by Hungarian speakers when differentiating between exceptional and regular stems in the experiment discussed in Section 2.6, a stem word condition will differentiate the nonce word stem stimuli based on the phonetic difference predicted to affect allomorph selection. Finding how this condition affects speakers' judgments will indicate how much they are using the phonetic cue given to them in this decision. Similarly, when assessing the effect of sublexical phonotactic regularities on the acceptability of stems with different harmonicities in Finnish in the experiment discussed in Section 3.3, the stimuli with be prepared to test different trends in the lexicon so that the results can assess what kinds of lexical biases assist Finnish speakers in their decisions.

1.3.3 Computational study

The goal of the wug studies in this dissertation is to get a better understanding on what kinds of information speakers of a vowel harmony language use when learning vowel harmony. Based on these results, this dissertation will attempt to give a computational account on how the acquisition of exceptional patterns in vowel harmony can be modeled in Chapter 4. The goal of this study will be to identify how the lexical biases found in the wug studies are actually able to influence learning. This goal is different from what Goldsmith and Riggle (2012) intend to model when analyzing Finnish vowel harmony: their goal is to present an information theoretic account to better understand phonological patterns like vowel harmony, and they do not propose any hypotheses on the acquisition of this phenomenon, or of the exceptional patterns within vowel harmony.

This computational modeling will employ two different learners that have been proposed in the literature before. The UCLA Phonotactic Learner (Hayes and Wilson 2008) is able to model how learners, given a feature set and an input of training data discover phonotactic regularities in the data in the form of violable constraints. This learning simulates the acquisition of static phonotactic generalizations, one of the two sides of vowel harmony. To investigate the learning of the other facet, allomorph selection, the Sublexical Learner (Allen and Becker 2015; 2016) was employed, which is designed to discover the nature of a given alternation and the sublexical phonotactic regularities for the stems that select different allomorphs.

The two models will be shown to work together better than separately. Learning the allomorphy helps the speakers understand what kinds of feature combinations are the most relevant to exceptional behavior and the static phonotactic learner establishes the boundary between grammatical and ungrammatical forms in the lexicon. Both of these learners rely on Maximum Entropy grammar, so that the constraints one learns is adoptable to the other one. Using Maximum Entropy means that the outputs of these grammars is stochastic (Jäger 2007, Hayes and Wilson 2008), but is predictive of the probabilities a certain form will be accepted as an output, therefore its results are going to be easy to align with the wug experiment results. The main goal of the computational models will be, therefore, to approximate the speakers' behavior as attested in the nonce word studies.

1.4 Ways to analyze exceptionality in vowel harmony

Vowel harmony has been quite frequently discussed in the phonological literature, as it is a linguistic phenomenon that appears to be connected to several interfaces to phonology (Nevins 2004, Kaun 2004, Hayes et al. 2009). The fact that vowel harmony is a phonetically natural phenomenon, providing preferable input and output for speakers' articulation and perception processes (Öhman 1966, Fowler 1983, Beddor et al. 2001), means that it is a straightforward to analyze using phonetically grounded approaches (Kaun 2004). The existence of transparency shows that vowel harmony is hard to explain using purely phonetic tools, though, as a simple analysis of vowel harmony as plain coorticulation between vowels would fail to account for forms with transparency. The differentiation between non-participating transparent and non-participating opaque vowels, which is language-dependent and not fully predictable also calls for an analysis to involve more abstraction than pure phonetic signal.

While surely phonetic and lexical factors are both involved in learning vowel harmony, the relative importance of these two when examining exceptional behavior is not well understood yet. In this dissertation I will focus on these two main sources of cues provided to speakers for exceptionality in vowel harmony: the phonetic surface form and phonotactic generalizations over exceptional sublexica. Both explanations offer testable predictions on what speakers will rely on when asked to categorize or rate novel items with regards to their harmonic behavior.

1.4.1 Phonetically grounded accounts

Learning the grammar is influenced by several kinds of biases, and phonetic detail is one of the important ones (Hayes et al. 2004). Phonetic motivation has been argued to be a crucial element of grammar: learners have biases towards phonetically natural processes, while they disprefer unnatural ones (Kaun 2004, Hayes et al. 2009, Becker et al. 2011).

1.4.1.1 Vowel harmony as a phonetically natural process

Vowel harmony in general has clearly been shown to be a phonetically grounded process (Kaun 2004, Linebaugh 2007). While in all speech production there is coarticulation to be found between vowels across intervening consonants (Öhman 1966), vowel harmony accommodates to this well (Fowler 1983, Beddor et al. 2001), minimizing articulator movement in a specific dimension (eg. tongue height or pharyngeal constriction, Boyce 1990, Gafos 1999). Vowel harmony also eases perception: shared acoustic traits help the perception of a word (Kaun 2004, Kimper 2011) partially by improving the predictability of vowels within a word (Suomi 1983), and vowel harmony has also be shown to be used as a cue for word segmentation (Suomi et al. 1997, Vroomen et al. 1998).

The bias speakers have towards vowel harmony patterns when compared to majority rules (patterns enforced by the majority of the input) or disharmony patterns (patterns not conforming to vowel harmony) has also been shown in artificial grammar learning experiments, where speakers of a language with no vowel harmony (like American English) learned harmonicity patterns better (Finley and Badecker 2012). Artificial grammar learning has also helped expose some more subtle preference predicted by phonetically based accounts in vowel harmony (Finley 2012).

1.4.1.2 Transparency as a problem

Neutral vowels already cause some challenges for a phonetic analysis of vowel harmony (Beddor et al. 2001, Benus and Gafos 2007), as neutrality entails words with mixed harmonicities, which negates the articulatory and perceptual advantages of vowel harmony. Opacity is not as much of a problem, because the advantages of vowel harmony can be attested at least on one side of the phonological word, where the opaque vowel's harmonicity is spread on. In transparency, however, the phonetic advantages that vowel harmony brings to the table are not utilized, conversely, disadvantageous sequences are created. Consider the perfectly grammatical adessive of the word *tuoli* 'chair' in Finnish:

(142) *tuoli-lla* 'on the chair'

Phonetically this word is problematic for the exact reasons why vowel harmony is preferred in harmonic words. In articulation, the tongue must protract for the [i] and then retract again for the final back vowel. The coarticulation that results might be detrimental to perception, and the mixing of front and back vowels might lead to the misperception that there is a word boundary in the form.

Transparency is also a second-order nonadjacent dependency (Finley 2015): it is non-local even if the vowels are extracted onto a vocalic tier. Therefore transparency has presented a problem not only for phonetically grounded analyses but for any kind of phonological account of vowel harmony, requiring rules or constraints with more complex computations (Ringen and Heinämäki 1999, Baković and Wilson 2000, Kiparsky and Pajusalu 2003) or a more complex structure of the phonological representation (Goldsmith 1985).

1.4.1.3 Antiharmonicity: the problem and possible solution

Antiharmony causes an even more serious blow to a direct phonetical analysis of vowel harmony (Benus and Gafos 2007). Antiharmonicity means that a given stem selects a suffix with the exact opposite phonetic value of the harmonizing feature than what is found in the stem. Therefore the phonetic motivation for antiharmonicity (if such a motivation exists) must be the opposite of the motivation for the whole vowel harmony pattern. This would mean that the analysis would rely on contradictory assumptions: both harmony and antiharmony cannot be optimal at the same time.

Motivating antiharmonicity has frequently been done by diachronic explanations (Vago 1980, Ringen and Vago 1998). For example, the Hungarian antiharmonicity pattern introduced above in examples (91–92) has been consistently explained in a historical way, claiming that those irregular stems that contain /i/ yet still take back suffixes originally had a *back* vowel /ɯ/ in them (E. Abaffy 2004a) (although even this historical argument is probably invalid, see Kis 2005 again). The development of Oirat Mongol antiharmonicity pattern is more clear: back vowels became front by the influence of a following neutral /i/, which after fronting the preceding vowel disappeared to make the pattern opaque and the stems antiharmonic.

This kind of explanation inspired phonological analysis as well, with analyses proposing an abstract underlying /ɯ/ phoneme that triggers back-suffixation and is opaquely fronted to [i] on the surface (Vago 1980). Similarly, later analyses proposed a floating [+back] feature to cause the peculiarities of affix allomorph selection, while a constraint like *ɯ forces the opacity on the surface (Nádasdy and Siptár 1994, Ringen and Vago 1998).

1.4.1.3.1 Retraction in mixed and antiharmonic stems While the phonological analyses above operate on a more abstract level, some of the findings of Benus and Gafos (2007) might be interpreted as evidence for a segmental analysis to be valid on a phonetic level, and therefore as a solution for the antiharmonicity issue for phonetically grounded analyses.

The main claim of Benus and Gafos (2007) is that there is an articulatory difference between the /i/ vowels found in mixed and antiharmonic stems and /i/ vowels in harmonic stems: in the former, the vowels are articulated with a more retracted tongue root. Their findings for transparency should be expected: in such stems /i/ is surrounded by back vowels, whose coarticulatory pressure should cause the front vowel to be somewhat more retracted. However, Benus and Gafos have also found this effect, the retraction of /i/ in antiharmonic stems standing in isolation as well. In such cases, there are no back vowels around the front vowels that could exert their coarticulatory effect on it. This could indicate that there is a segmental level difference between vowels in regular and in irregular stems underlyingly, and also in a limited way on the surface. This difference could eliminate the need for more complicated computations or representations: all affixes would attach with the harmonicity corresponding to the stem.

The results of Benus and Gafos (2007) were widely interpreted as a case showing that transparency and antiharmonicity might be illusory phenomena (Hansson 2010), and as a serious question about the assumption that transparent vowels are not participanting in vowel harmony (McCarthy 2009). This interpretation suggests that exceptionality can be explained based on the phonetic facts in these minor details of segments occuring in exceptional morphemes.

An alternative interpretation of the Benus and Gafos results is that the retraction found in the antiharmonic and transparent stems is not the cause of these exceptional patterns, but rather a side effect of their application. This interpretation, endorsed by Benus and Gafos (2007) themselves and discussed further by Hansson (2008), is based on exemplar theory and coarticulatory effects. In the case of antiharmonic stems, this interpretation attributes the slight tongue retraction during the stem vowel in a word like *hiːd* 'bridge' in isolation with the fact that the stem appears very frequently before back vowels in suffixes, in word forms like *hiːd-nɒk* 'bridge-Dat', where coarticulation retracts the production of the [i] vowel. As these exemplars of suffixed forms in general have a more retracted vowel than the average exemplar of a harmonic stem like *viːz* 'water', the unsuffixed production will also be more shifted towards more retraction as the speaker samples their production based on their previously encountered exemplars for all word forms. This explanation, therefore, derives the retractedness from exceptionality and does not attempt to explain exceptionality based on the phonetic detail.

1.4.1.3.2 Incomplete neutralization-based analysis of antiharmony The relationship between /i/ and the slightly retracted /i̠/ would be similar to cases of incomplete neutralization (Port and O'Dell 1985, Port and Crawford 1989, Warner et al. 2004), a case of suspended contrast(Yu 2007). In incomplete neutralization contrast is lost between two categories in certain phonological environment, therefore the distinction between the two categories is not as clear cut as between other, fully contrastive environments. For example voicing contrast is lost at word-final position in German, Dutch and many Slavic languages. In German, for example /ʁaːd/ 'wheel' and /ʁaːt/ 'advice' are both uttered with a voiceless final obstruent, but the duration of the preceding vowel and stop closure and burst durations are significantly different between the two words. The distinction between /d/ and /t/ is full in environments other than word-final, as attested by the plural forms *ʁeː**d**ɐ* 'wheels' and *ʁeː**t**ə* 'advice', corresponding to the incompletely neutralized *ʁaː**t*** singular for the stems introduced above.

A similar phenomenon is near merger, where speakers of a certain variety fail to distinguish between two categories in perception tasks, however, speakers maintain consistent differences in production, suggesting the possibility of a distinction on some representational level (Labov et al. 1991). Examples include several vowel mergers in English varieties in Norwich (Trudgill 1974), Essex (Labov 1971), Belfast (Milroy and Harris 1980) and New York City (Labov et al. 1972), as well as tone in Cantonese (Yu 2007).

1.4.1.4 Phonetics first hypothesis

To formalize the proposal that phonetic detail can explain the exceptional patterns, most notably antiharmony, the phonetics first hypothesis will be defined as such:

(143) *Phonetic first hypothesis*: Irregular patterns in vowel harmony can be explained by phonetic (possibly subphonemic) detail in the surface form.

The main prediction of this hypothesis is that speakers are able to use phonetic differences when categorizing stems. In the Hungarian pattern, this would mean that there would be some phonetic difference that speakers are able to use when they decide whether a stem is harmonic or antiharmonic. Under this hypothesis, if Hungarian speakers are tested on nonce words containing full [i] vowels and on words with somewhat retracted [i̠] vowels, they would prefer back suffixation more for the latter set of nonce stems. Chapter 2 of this dissertation will investigate this exact question, and will demonstrate that speakers do not use such cues when categorizing novel stems as harmonic or antiharmonic.

1.4.2 Sublexical phonology

A different source of information for the learner when acquiring exceptionality is the lexicon. Several studies have shown that speakers identify sublexica (subsets of the lexicon) based on stems' morphophonological behavior (Becker et al. 2011, Hayes et al. 2009, Linzen et al. 2013, Gouskova et al. 2015). The sections below will discuss the literature supporting this idea, formalize a sublexical hypothesis for exceptionality in vowel harmony and lay out a plan for testing this hypothesis.

1.4.2.1 Frequency matching

An important idea needed for testing how much speakers are aware of sublexical regularities in their language is the law of lexical frequency matching defined by Hayes et al. (2009) as below:

(144) *law of frequency matching*: Speakers of languages with variable lexical patterns respond stochastically when tested on such patterns. Their responses aggregately match the lexical frequencies.

This definition means that when speakers rely on lexical subpatterns when categorizing (possibly novel) items, they do not produce either categorical or random decisions, but the distribution of their results pattern really closely with the lexicon. For example, Ernestus and Baayen (2003), which investigated Dutch word-final devoicing, looked at what Dutch speakers base their decision on, when they have to categorize a novel stem as ending in either an underlyingly voiced or an underlyingly voiceless consonant, given a form that could entail either interpretation. The suffix for Dutch verbs in their first person singular present form is null, therefore the stem final consonant is word final, and subject to devoicing, while in the voicing distinction is present before the infinitive suffix. This phenomenon is shown in the neutralization pattern below:

(145) *vervɛi**t**-* 'to reproach' ∼ *vervɛi**t**ən* 'reproach-Inf', *vervɛi**t*** 'reproach-1Sg'

(146) *vervɛi**d**-* 'to widen' ∼ *vervɛi**d**ən* 'widen-Inf', but *vervɛi**t*** 'widen-1Sg'

Ernestus and Baayen gave nonce forms similar to the existing form *vervɛit* to speakers and produce a past tense form that disambiguates between the underlying voiced and underlying voiceless stem. They found that the responses were not categorical, and the proportion of the voiced responses closely matched the frequency of voiced stems with the same final consonant in the lexicon. For example the proportion of voiced responses for *p*-final inputs with long vowels like *ɪk daup* was only 3%, as the ratio of *b*-final verbs to *p*-final verbs in the Dutch lexicon is 0%. For *f* finals, the proportion of voiced responses was 64%, which corresponds to 100% in the lexicon. Intermediate lexical proportions lead to intermediate proportion of voiced responses.

Similar results were found for various morphophonologically variable environments as English past tense formation (Albright and Hayes 2003), and Korean stem-final consonant quality variation (Jun and Lee 2007).

1.4.2.2 Evidence for sublexical knowledge

The law of frequency matching and the nonce word experimental paradigm provides a useful toolset to investigate what kinds of sublexical regularities are speakers aware of and what determines whether speakers learn the sublexical phonotactics for a given phonological pattern. This has triggered several studies that have found that sublexical phonotactic regularities are used as cues by the speakers for many morphophonological alternations in several languages.

Becker et al. (2011) investigated Turkish speakers' judgments on a very similar pattern to the Dutch example discussed by Ernestus and Baayen (2003): as Turkish is also a language with word-final obstruent devoicing, the underlying voicing of the stem is ambiguous based on the suffixless form. Becker et al. have found that speakers did use sublexical regularities in deciding how to use a novel stem, but they did not use all information. The results indicated that speakers used phonotactic regularities based on the last consonant's characteristics, but did not use information based on the previous vowel. Their conclusion was that speakers are innately aware that the consonant quality can affect this locus of variation, but the preceding vowel is never relevant.

Looking at Hungarian neutral vowels both in mixed and antiharmonic stems, Hayes et al. (2009) found that Hungarian speakers also decide according to the law of frequency matching both when deciding on allomorph selection where variation is possible in the grammar, and also when forming judgments about novel stems. Similarly to Becker et al. (2011), they have also found that vowel to vowel correspondences in the lexicon matter much more in speakers' decisions than consonantal effects, however, these consonantal effects they called "unnatural" also played a small role in speakers' judgments.

Similar results were found in other investigations of the relationship between speakers' knowledge of sublexical phonotactic regularities and morphophonological behavior. For example vowel deletion phenomena in Russian, which has been long analyzed as a segmental level process (Lightner 1972) using abstract phonemes called *yers*, turned out to be heavily influenced by sublexical phonotactic knowledge (Gouskova and Becker 2013, Becker and Gouskova 2016). The selection of a diminutive suffix attached to the stem in Russian is also variable and can be best explained by sublexical phonotactic regularities (Gouskova et al. 2015). Sublexical influence has also been found for the selection of synthetic or analytic comparative in English (Gouskova and Ahn 2016), French and Portuguese exceptional plural forms (Becker et al. forth.) and nominal class categorization in Xhosa (Braver and Bennett 2016).

1.4.2.3 Lexicon first hypothesis

The hypothesis based on the proposal that sublexical phonotactic patterns are the most important factor for speakers when categorizing possibly novel lexical items as exceptional or can be summarized as below:

(147) *Lexicon first hypothesis*: Irregular patterns in vowel harmony can be explained by marking of irregular stems in the lexicon.

The predictions of this hypothesis, as seen in this subsection, can be tested by comparing the responses of speakers of vowel harmony languages to the lexical statistics. The study of Hayes et al. (2009) is already a step in exploring the ways how exactly lexical knowledge enables speakers to make decisions regarding vowel harmony. Chapter 3 will further investigate on what kinds of sublexical informations speakers use when deciding on the grammaticality of a novel form under the static facet of vowel harmony, by looking at the lexical characteristics of Finnish mixed stems, and Finnish native speakers' judgments on novel stems.

1.5 Outline of the dissertation

1.5.1 Main proposals

The main goal of this dissertation is to investigate the main source of information that speakers use when deciding whether a given input behaves exceptionally in their language. The two sources discussed above: phonetics and sublexical information will be investigated.

The results in the following chapters indicate that sublexical information seems to be the crucial information speakers of vowel harmony languages use to learn, understand and use exceptional patterns. Chapter 2 will first demonstrate that contrary to how the results of Benus and Gafos (2007) were often interpreted, phonetic cues are not used at all by Hungarian speakers in antiharmony, while sublexical cues are, as shown by Hayes et al. (2009).

The hypothesis that sublexical regularities inform speakers' judgments will be confirmed in Chapter 3, where the wug study results will align with the lexical statistics closely. However, similarly to Becker et al. (2011) and Hayes et al. (2009) a significant difference will be found between the effect of segments that phonologically and phonetically more naturally influence vowel harmony (vowels) and other segments (consonants): the former can explain the vast majority of the results.

The computational part (Chapter 4) of this dissertation will present how speakers behavior can be modeled by maximum entropy learning algorithms already used in the literature. The target of learning will be derived from the results of the lexical and the wug study. The results will show that learning allomorphy enhances phonotactic learning, as it helps the learner identify the crucial feature bundles used in the harmony pattern in the language.

1.5.2 General summary

The rest of the dissertation is divided into four chapters. The chapters differ at the phenomenon they investigate: Chapter 2 is looking at antiharmonicity in Hungarian while Chapters 3 and 4 focus on static phonotactic generalizations in Finnish and mixed stems as a static exceptionality pattern. The hypotheses tested in each chapter are also different: in Chapter 2 the plausibility of the phonetics first hypothesis is tested. As its tenability is questioned, Chapter 3 will follow up testing the predictions of the lexicon first hypothesis, and will investigate the effect of the phonetic or phonological naturalness of a constraint on surface phonotactics. Chapter 4 will use the results of the experiments run on lexicon first hypothesis (as the phonetic first hypothesis is rejected in Chapter 2) and will study the learnability of exceptionality in static phonotactic regularities using maximum entropy learners.

1.5.3 Chapter-by-chapter outline

Table 1.3 below summarizes the main hypotheses tested in each chapter in the dissertation:

hypothesis	**predictions**	**chapter**	**results**
phonetics first	phonetic differences will be present between regular and exceptional stems	Chapter 2	partial
	phonetic detail will be used by speakers to categorize novel stems as exceptional or regular	Chapter 2	×
lexicon first	sublexical differences will be used by speakers to categorize novel stems	Chapter 3	✓
	vowel interactions will influence speakers more than consonantal ones	Chapter 3	✓
learnability with maxent	maximum entropy learners are able to replicate speakers' preferences on exceptional stems	Chapter 4	✓
	learning alternations helps the learner figure out the static phonotactic generalizations	Chapter 4	✓

Table 1.3: Summary of main hypotheses tested in this dissertation

First, the case of a dynamic irregularity pattern will be investigated with the case of Hungarian antiharmonicity. The claim that exceptionality can be represented on a segmental level, and this exceptionality surfaces as a phonetic difference between regular and irregular stems will be tested in Chapter 2. This claim will be evaluated as part of a phonetic first hypothesis as discussed above in 1.4.1.4.

To test this hypothesis, three different experiments will be presented in Sections 2.4–2.6. The first two experiments will be plain phonetic studies testing the presence of a surface difference predicted by the phonetics first hypothesis. The goal of the acoustic experiments is to find out whether there is a phonetic correlate of antiharmonicity in the affected stems. Since the results indicate that there is a small difference between vowels in regular and irregular stems. The main question, though was to find out whether such a difference can even be perceived by the speakers and whether they are able to use such differences in categorizing novel stems as regular or irregular. The answer to both of these questions was negative, as evidenced by the results of the follow-up perception (Section 2.5) and nonce word (Section 2.6) experiments.

The untenability of the phonetics first hypothesis above leads to the proposal of an alternative hypothesis, where irregularity is indeed a property coded in the lexicon. The lexicon first hypothesis, as discussed above in Section 1.4.2.3

This hypothesis is also testable: the law of frequency matching as discussed by Ernestus and Baayen (2006) and Hayes et al. (2009) should hold for sublexica learned based on harmonic behavior as well. That this is true for Hungarian antiharmonicity has been shown by Hayes et al. (2009) themselves. The rest of the dissertation explores this hypothesis further and attempts to answer what exactly and how exactly speakers learn these sublexica.

Chapter 3 will investigate the lexical hypothesis, more specifically, how much sublexical patterns influence speakers when deciding whether they prefer mixedness or full harmonicity for a given nonce word. First, a thorough corpus analysis will be provided (Section 3.2) to discover whether a stem's harmonicity can be predicted based on the phonologically less natural effect of the preceding consonants. After some patterns do emerge, speakers are again tested with a wug study in Section 3.3 to see how much they use these kinds of lexical information in wug decisions. The results will show that speakers replicate lexical statistics, but only partially: their responses correlate closely with predictions made solely based on the sublexical frequencies of the vowels in the stem, but the consonantal effects discovered in Section 3.2 are not borne out.

Finally, the goal of Chapter 4 will be to build a computational model on how exceptional patterns in vowel harmony are learned. First, a full lexical analysis of harmonicity types in the Finnish lexicon will be discussed (4.1) to identify the optimal input to the learning algorithm, and then this input will be analyzed by the UCLA Phonotactic Learner (Hayes and Wilson 2008, Section 4.2), which focuses on discovering static phonotactic generalizations. It will be shown that the learner will need certain biases in order to be able to successfully approximate the patterns found in the lexicon and in the wug studiess. Some of these biases will look like language specific generalizations that could be only acquired during learning allomorph selection. Therefore, Section 4.3 will enroll the Sublexical Learner of Allen and Becker (2016), which approximates the process of learning allomorphy based on sublexical regularities. This learner will be able to identify the language specific settings that are needed by the static learner to successfully simulate speaker judgments. The main conclusion of this part of the dissertation will be that maximum entropy grammars can model learning exceptionality patterns in a complex vowel harmony system like Finnish, and that at least some part of learning the phonotactic alternations in the language has to occur simultaneously or after the speaker has figured out how the grammar treats allomorph selection.

Chapter 2

Antiharmonicity in Hungarian

This chapter discusses the issue of antiharmonicity in Hungarian – the phenomenon that while regularly /i/ is a front vowel triggering front suffixation, a subset of stems containing /i/ regularly selects back suffixes. This pattern poses a problem to most analyses of vowel harmony, especially to phonetically grounded theories, because the resulting suffixed forms are phonetically disadvantageous for the exact same reason why vowel harmony in general is phonetically advantageous to the speaker and the listener. The chapter evaluates a phonetically based and a lexically based solution for this problem. To test the predictions of a phonetically based and a lexically based hypothesis, three experiments will be presented, which will give empirical background to the preference of a lexical analysis over a purely phonetic one.

Section 2.1 will present Hungarian vowel harmony, and introduce the pattern of antiharmonicity in Hungarian. Section 2.2 will give a basic analysis of the facts, pointing out that the handling of antiharmonicity needs some extra machinery, and Section 2.3 introduces two competing hypotheses: a phonetic (2.3.1) and a lexical (2.3.2) hypothesis, and how analytic tools built on these hypotheses can fix the problem. Three experiments were done to test the plausibility of the phonetic hypothesis. A production experiment is presented in Section 2.4, which shows that there is indeed a very small phonetic difference between an /i/ in a regular stem and an /i/ in a stem that selects back suffixes. The perception experiment Section 2.5 will show that this small difference is not perceivable for Hungarian speakers. The wug stem rating study in Section 2.6 will show that even if it is augmented so that it is perceivable, this phonetic difference does not influence Hungarian speakers' decisions whether to categorize a novel stem as regular or antiharmonic. Finally, Section 2.7 will summarize the results and implications of these experiments.

2.1 The pattern

The Hungarian vowel system contains seven short and seven long vowels, divided into front rounded, front unrounded and back categories based on their phonetic properties and their behavior in vowel harmony (Siptár and Törkenczy 2000, Hayes et al. 2009).

	front				back	
	unrounded		rounded			
high	i	iː	y	yː	u	uː
mid		eː	ø	øː	o	oː
low	ɛ				ɒ	aː
category	**neutral**		**front**		**back**	

Table 2.1: The Hungarian vowel system

Vowel harmony in Hungarian is governed by the frontness of the stem. The vowel system is categorized into back (phonetically back), front (phonetically front rounded) and neutral (phonetically front unrounded) classes based on their behavior in harmony (Hayes et al. 2009) that will be described below. Most suffixes have a front and a back allomorph: stems with only back vowels select the back suffix and stems with only front vowels select the front suffix, as illustrated in (148–149). This is complicated by "mixed stems": polysyllables with both back and front vowels. In these stems the final vowel determines the identity of the suffix. Examples in (150–151) illustrate the behavior of mixed stems. The suffix attached in these examples is the dative, which alternates between the *nɒk* back and *nɛk* front allomorphs – this suffix is representative of the pattern; other suffixes that alternate between a back and a front allomorph follow the same pattern:

(148) back stem: *hɒjoː* 'ship' ∼ *hɒjoːnɒk*, **hɒjoːnɛk*

(149) front stem: *tɛtøː* 'roof' ∼ *tɛtøːnɛk*, **tɛtøːnɒk*

(150) mixed back stem: *tɛrɒs* 'terrace' ∼ *tɛrɒsnɒk*, **tɛrɒsnɛk*

(151) mixed front stem: *ʃoføːr* 'driver' ∼ *ʃoføːrnɛk*, **ʃoføːrnɒk*

The exception from the generalization that the last vowel of the stem predicts the frontness of the suffix is when the final vowel is front unrounded /ɛ, eː, i, or iː/, which can be transparent. Therefore, the stem takes back suffixes if they follow back vowels (152) and front suffixes is they follow front (153) or other neutral vowels (154):.

(152) back-neutral stem: *kɒbin* 'cabin' ∼ *kɒbinnɒk*, **kɒbinnɛk*

(153) front-neutral stem: *røvid* 'short' ∼ *røvidnɛk*, **røvidnɒk*

(154) neutral-neutral stem: *kitʃi* 'small' ∼ *kitʃinɛk*, **kitʃinɒk*

Most stems which contain one front unrounded (neutral) vowel take front suffixes, as illustrated in (155–157):

(155) front i stem: *siːv* 'heart' ∼ *siːvnɛk*, **siːvnɒk*

(156) front *eː* stem: *eːv* 'year' ∼ *eːvnɛk*, **eːvnɒk*

(157) front *ɛ* stem: *hɛj* 'place' ∼ *hɛjnɛk*, **hɛjnɒk*

However, certain stems with only neutral vowels take *back* suffixes. These stems will be labeled as antiharmonic to highlight the exceptionality (non-productivity) of this lexicalized pattern. Most antiharmonic stems are monosyllabic, as in (158–160), although there are some disyllabic antiharmonic nouns and verbal stems with two neutral vowels as well.

(158) antiharmonic *iː* stem: *ɲiːl* 'arrow' ∼ *ɲiːlnɒk*, **ɲiːlnɛk*

(159) antiharmonic *i* stem: *fiŋg* 'fart' ∼ *fiŋgnɒk*, **fiŋgnɛk*

(160) antiharmonic *eː* stem: *t͡seːl* 'target' ∼ *t͡seːlnɒk*, **t͡seːlnɛk*

Only front suffixing is productive for neutral vowels: new stems entering the language (such as loan words and acronyms) are always harmonic, confirming that antiharmonicity is not productive:

(161) *liŋk* 'hypertext link' ∼ *liŋknɛk* 'hypertext link.DAT', **liŋknɒk*

2.2 Basic analysis

This section describes the framework used to formalize phonological patterns in this chapter. The general framework used for the analysis of Hungarian vowel harmony, including transparency, will be described here, while the machinery needed for the explanation of the hypotheses on the analysis of antiharmonicity will be discussed below in Sections 2.3.1 and 2.3.2.

The framework used for these analyses is the system developed in Kimper (2011), which is based on a Harmonic Serialism grammar with positive constraints for feature spreading. While the serial nature of the derivation is crucial for Kimper's system to avoid pathological predictions (see Kimper 2011, pp. 13–29), none of the data discussed here will require more than a single derivational step. The constraints used in the analysis are defined in Table 2.2 below.

general form	form for Hungarian	definition
SPREAD(αF)	SPREAD(+BACK)	Assigns a positive weight for each segment linked as a dependent to F in a candidate.
SPREAD(-αF)	SPREAD(-BACK)	Assigns a positive weight for each segment linked as a dependent to a non-preferential value for F in a candidate.
IDENT(F)	IDENT(BACK)	Assigns a negative weight for different feature specification in corresponding segments
*(X,αF)	*(+ROUND,-BACK)	Assigns a negative weight for the co-occurrence of a feature X with the feature F responsible for harmony in the output

Table 2.2: Constraints in Kimper (2011) used by this analysis

The faithfulness constraint IDENT(F) is usually ranked low for systems with vowel harmony, as spreading almost always overrules faithfulness to achieve vowel harmony. The base cases of harmony with simple spreading can be now analyzed with these constraints. The relative weights between constraints are chosen so that the spreading constraints outweigh faithfulness, resulting in harmony. The ranking between SPREAD(+BACK) and SPREAD(-BACK) is not yet relevant. An example for the spreading of [αback] onto the suffix is illustrated below in (162) and (163). Both tableaux present a situation when the underlying form of the suffix is chosen to have a different [αback] specification than the stem to illustrate that spreading occurs nonetheless. In both tableaux, the faithful candidate receives a harmony score of 0, as no spreading is happening and the faithfulness constraint has no reason to assign violation. The harmonizing candidates (*nɒp-nɒk* and *hɛɟ-nɛk* manage to emerge as winners, because they get assigned a positive score by the SPREAD constraints, since the suffix vowels are dependents on the [+back] or [-back] feature of the stems. This positive score outweighs the negative score assigned by the faithfulness constraint, allowing harmony to emerge.

(162) Spreading back:

[+] [-] | | /n ɒ p-n ɛ k/	SPREAD(+BK) 4	SPREAD(-BK) 2	IDENT(BK) 0.5	$\mathcal{H}$
[+] ☞ nɒpnɒk	+4		-0.5	+3.5
[+] [-] | | n ɒ pnɛk				0.0

(163) Spreading front:

[-] [+] | | /hɛɟ-n ɒ k/	SPREAD(+BK) 4	SPREAD(-BK) 2	IDENT(BK) 0.5	$\mathcal{H}$
[-] ☞ hɛ ɟn ɛ k		+2	-0.5	+1.5
[-] [+] | | hɛɟn ɒ k				0.0

Note that the directionality of feature spreading is left-to-right: from stem to suffix. Like Kimper (2011), the examples above and the analyses below set modeling the directionality questions aside. There are several ways of handling this question, like using faithfulness constraints indexed to stems, or scaling factors that penalize spreading onto the stem. Example 164 below shows how directionality is handled with indexed faithfulness constraints:

(164) Spreading from stem to suffix:

[-] [+] /hɛɟ-n ɒ k/	IDENT(BK)$_{\text{Stem}}$ 8	SPREAD(+BK) 4	SPREAD(-BK) 2	IDENT(BK) 0.5	$\mathcal{H}$
☞ [-] hɛ ɟn ɛ k			+2	-0.5	+1.5
[-] h ɛ ɟn ɛ k	-8	+4		-0.5	-4.5

2.2.1 Transparency

The markedness constraint *(X,αF) is responsible for blocking harmony, as if its weight is high enough, it blocks spreading that results in marked forms, leading to transparency or opacity. In the case of Hungarian, *(+ROUND,-BACK) blocks spreading of [+back] to front unrounded vowels /ɛ, eː, i, iː/ as they lack their back unrounded counterparts */ʌ, ɤ, ɯ/ in Hungarian. The absence of these segments motivates blocked harmony, which can manifest as opacity (as is can be the case with /ɛ/ in Hungarian) or transparency – the exact consequences being determined by the ranking of other constraints in the grammar.

The analysis of transparency is presented below in (165). The faithful candidate (165b) has a score of 0 again, so the crucial *guminɛk* candidate to compare the winning one with is (165d), where [-back] spreads onto the suffix from the stem. The latter gains a +2 score from the spreading constraint, but that is still not enough to beat candidate (165a), where the [+back] feature spreads from the farther vowel in the stem onto the suffix, because spreading of [+back] is more preferred than [-back]. An important candidate to consider is

(165c), where the [+back] feature value spreads from the initial vowel to the second vowel of the stem as well, besides the suffix, therefore gaining two positive +4 scores from the SPREAD(+BACK) constraint. The victory of this candidate is avoided by the -8 negative score assigned by the markedness constraint that penalizes [+back,-round] segments on the surface.

(165) Analysis of transparent vowels

[+] [-] [-] | | | /g u m i n ɛ k/	*(-RD,+BK) 8	SPREAD(+BK) 4	SPREAD(-BK) 2	IDENT(BK) 0.5	$\mathcal{H}$
[+] [-] ☞ a. g u m i n ɒ k		+4		-0.5	+3.5
[+] [-] [-] | | | b. g u m i n ɛ k					0.0
[+] c. g u m ɨ n ɒ k	-8	+4+4		(-0.5)+(-0.5)	-1
[+] [-] | d. g u mi n ɛ k			+2		+2

2.2.2 Distance sensitivity

The analysis above is based on the possibility of crossing association lines. There is a reason to believe that this possibility is not unlimited: in Hungarian, more than one transparent vowel can behave as opaque, spreading front harmony on the suffix (Count Effect in Hayes et al. 2009):

(166) two transparent vowels as optionally opaque: *ɒnɒliːziʃ* ‘analysis’ ∼ *ɒnɒliːziʃnɛk* or *ɒnɒliːziʃnɒk*

(167) two transparent vowels as optionally opaque: *horːibiliʃ* ‘horrible’ ∼ *horːibiliʃnɛk* or *horːibiliʃnɒk*

To account for this distance sensitivity, the effect that transparency is gradient depending on the distance of the trigger and the target, a scaling factor is introduced for the spreading constraints. Spreading can still happen between non-local vowels, skipping one or more, but the positive weight of the constraint is multiplied by this factor. This factor is always lower than 1, so that non-local spreading is always rewarded less than local spreading:

(168) s_{nl}: a penalizing scaling factor. A SPREAD constraint is multiplied by s_{nl} as many times, as many segments are not dependent on the head of the spreading feature and the dependent.

The requirement for transparency is that the weight of spreading of the preferred feature value is still greater than the sum of weight of the spreading constraint for the other value and the weight of the faithfulness constraint, even if it is multiplied by this factor. This is schematized below:

(169) $w(\textsc{Spread}(\alpha F)) * s_{nl} > w(\textsc{Ident}(\mathrm{F})) + w(\textsc{Spread}(-\alpha F))$

A setting of $s_{nl} = 0.7$ is used in the tableaux below. Now we are able to describe transparency in Hungarian, almost exactly as it is shown in Kimper (2011) (p. 92, the tableau in (84)). The tableau below in (170) shows that for one intervening transparent segment, the analysis still works:

(170) Analysis of transparent vowels, with penalty for non-local spreading

[+] [-] [-] | | | /g u m i n ɛ k/	*(-Rd,+Bk) 8	Spread(+Bk) 4	Spread(-Bk) 2	Ident(Bk) 0.5	$\mathcal{H}$
[+] [-] ☞ a. g u m i n ɒ k		+4*0.7		-0.5	+2.3
[+] [-] [-] | | | b. g u m i n ɛ k					0.0
[+] c. g u m ɨ n ɒ k	-8	+4+4		(-0.5)+(-0.5)	-1
[+] [-] | d. g u m i n ɛ k			+2		+2

The tableau below shows that for a word like /*ɒspirin*/ 'aspirin', the winning candidate can be either [*ɒspirinnɒk*] or [*ɒspirinnɛk*], as their harmony scores are almost the same. Since there is variation in the system, very close harmony scores do predict a roughly 50%-50% variation, though the further analysis using a probabilistic grammar is outside the scope of this dissertation. Below, the transparent winning candidate is (171a), where the spreading of the [+back] feature from the stem onto the suffix is penalized by 0.8^2, since there are two intervening vowels. The candidate gets a +2 from Spread(-Back) since spreading can happen from the first [i] in the stem onto the second [i]. The opaque winner is (171d), which gets the +2 for the non-preferential spreading twice, as the [-back] feature from the first [i] is able to spread on two dependents: the second [i] and the suffix vowel too. This analysis is similar in spirit to how Kuhn (2012) handles multiple trigger effects in Kazakh consonantal harmony.

(171) Analysis of distance sensitivity

[+] [+] [-] [-] / ɒ sp i r i nn ɛ k/	*(-Rd,+Bk) 8	Spread(+Bk) 4	Spread(-Bk) 2	Ident(Bk) 0.5	$\mathcal{H}$
[+] [-] ☞ a. ɒspirinnɒk		$+4 * 0.8^2$	+2	-0.5	+3.96
[+] [-] [-] [-] b. ɒ sp i r i nn ɛ k					0.0
[+] [-] [-] c. ɒ spi r inn ɛ k			+2		+2.0
[+] [-] ☞ d. ɒ spi r i nn ɛ k			+2+2		+4

2.2.3 Quality sensitivity

To explain the fact that while transparency is the most common strategy with [i], but [ɛ] can, and frequently does, behave opaquely (the fact summarized as Height Effect by Hayes et al. 2009), another scaling factor is employed. This factor also works as a multiplier for the positive spreading constraints, however, this is a factor that always has a value greater than 1, rewarding spreading from more preferred triggers. Kimper (2011) discusses preferred triggers in a phonetic way, following Kaun (1995) and more recently Finley (2008): vowels that are perceptually weak in the feature responsible for the harmony are preferred as triggers. In a backness/frontness harmony system, this means that lower triggers are more preferred, so spreading from them is rewarded by this scaling factor.

Without the scaling factor, a stem like /*ʤuŋgɛl*/ 'jungle' would behave exactly like /*gumi*/ in the previous section: opacity would not be possible:

(172) Problem with the analysis of the opacity of the low front unrounded vowel without the quality sensitive scaling factor

[+] [-] [+] /ʤ u ŋgɛln ɒ k/	*(-RD,+BK) 8	SPREAD(+BK) 4	SPREAD(-BK) 2	IDENT(BK) 0.5	$\mathcal{H}$
☞💣 [+] [-] ʤ uŋgɛ lnɒk		4*0.8			+3.2
[+] [-] [+] ʤ u ŋgɛln ɒ k					0.0
[+] [+] ʤ uŋgʌln ɒ k	-8	+4		-0.5	-4.5
☞ [+] [-] ʤ u ŋgɛ ln ɛk			+2	-0.5	+1.5

The scaling factor introduced to solve this problem is defined as below:

(173) s_{trig}: A rewarding scaling factor. A positive weight for a SPREAD constraint is

multiplied by s_{trig} if the trigger (the segment associated with the head of the spreading feature) is a preferred one – so in a back-front harmony system, it is low.

Spreading from a preferred trigger can overrule transparency if the less preferred feature spreading multiplied by s_{trig} outweighs the preferred feature spreading weight penalized with s_{nl} (the factor responsible for non-local spreading). This condition is summarized as the following inequality:

(174) $w(\textsc{Spread}(-\alpha F)) * s_{trig} > w(\textsc{Spread}(\alpha F)) * s_{nl}$

This explains why opacity is a choice for stems like /*ʤuŋgɛl*/, as illustrated in the tableau below in (175). The setting for s_{trig} is 2. The opaque candidate (175d) is able to win over the transparent canditate (175a) because the spreading from the front vowel [ɛ] is rewarded by multiplying the reward for spreading by 2:

(175) Analysis of the opacity of the low front unrounded vowel

[+] [-] [+] /ʤ u ŋgɛln ɒ k/	*(-Rd,+Bk) 8	Spread(+Bk) 4	Spread(-Bk) 2	Ident(Bk) 0.5	$\mathcal{H}$
[+] [-] a. ʤuŋgɛlnɒk		4*0.8			+3.2
[+] [-] [+] b. ʤ u ŋgɛln ɒ k					0.0
[+] [+] c. ʤuŋgʌln ɒ k	-8	+4		-0.5	-4.5
[+] [-] ☞ d. ʤ u ŋgɛ ln ɛ k			+2*2	-0.5	+3.5

2.2.4 The problem with antiharmonicity

The system at this point is unable to describe antiharmonicity: there is no way to distinguish between a harmonic stem like /*viːz*/ 'water' (∼ [*viːznɛk*] 'water-Dat') and an antiharmonic stem like /*hiːd*/ 'bridge' (∼ [*hiːdnɒk*] 'bridge-Dat'). This is illustrated in the tableaux below. The stems in (176) and (177) cannot be distinguished by the constraints: a spreading candidate with a front vowel wins in both tableaux, as the spreading constraint outweighs the faithfulness constraint. There needs to be a way to reward non-spreading or to penalize spreading for /*hiːd*/, but not /*viːz*/, in order to reach a correct analysis, but there is no way to do that with the machinery introduced in this section.

(176) Harmonic stem

[-] [+] | | /viːzn ɒ k/	*(-Rd,+Bk) 8	Spread(+Bk) 4	Spread(-Bk) 2	Ident(Bk) 0.5	$\mathcal{H}$
[-] ☞ viːzn ɛ k			+2	-0.5	+1.5
[-] [+] | | viːzn ɒ k					0.0

(177) Problem with antiharmonic stem

[-] [+] | | /hiːdn ɒ k/	*(-Rd,+Bk) 8	Spread(+Bk) 4	Spread(-Bk) 2	Ident(Bk) 0.5	$\mathcal{H}$
[-] 💣 hiːdn ɛ k			+2	-0.5	+1.5
[-] [+] | | ☜ hiːdn ɒ k					0.0

The following section will discuss two ways of fixing this problem: a segmental level, phonetic explanation and a morpheme level, sublexical proposal.

2.3 Hypotheses for analyzing antiharmonicity

The analysis developed in the section above is able to account for most of the Hungarian vowel harmony pattern including most of the exceptional behavior seen in this system – but it lacks a way to handle antiharmonicity. The goal of this section is to introduce two ways of incorporating antiharmonicity in the analysis. The big question is whether this incorporation can be done in a phonetically grounded way, therefore, one of the proposals will be phonetic and segmental level in nature, while as an alternative, a morpheme-level solution will be shown.

In the phonetically based phonology framework (Hayes et al. 2004), the synchronic grammar is directly influenced by phonetic biases. In such a model of phonological knowledge phonetic motivation is directly encoded in the grammar, so rules or constraints that have phonetic motivation will be strongly preferred. This preference can be manifested in different ways under different further hypotheses. It can be thought that the only constraints allowed into the universal set of constraints are the ones which have phonetically based explanation, and these constraints can be used to create a typology of languages with respect to a subset of phonological phenomena (see eg. Kaun 2004). Another way of thinking about the preference of phonetically grounded constraints is that they introduce a strong bias to the grammar. The grammar can also be influenced by other analytic or structural biases that can be responsible for creating unnatural patterns (Moreton 2008, Pater and Moreton 2012), but phonetically grounded biases should have larger effects and phonetically natural patterns should be learnable more easily (Hayes et al. 2009).

Vowel harmony has been widely discussed as a phenomenon that fits in a phonetically based framework well as it has a clear phonetic motivation. From an articulatory point of view vowel harmony is a natural pattern as it accomodates coarticulation between vowels

(Fowler 1983, Beddor et al. 2001). This means that harmony is driven by constraints which disprefer tongue movement between vowels following each other leading to prefering similarity in an articulatory dimension (tongue body retraction, height, etc.) across the word. Tongue shape can stay the same across consonants in vowel harmony languages (Boyce 1990, Gafos 1999, p. 41), so that tongue movement can be minimal and articulatory ease is enhanced.

From a perceptual point of view, the goal of vowel harmony is that perception can be made easier when an acoustic feature is shared throughout a word (Suomi 1983, Kaun 2004, Kimper 2011). The constraints needed to achieve this goal are ones that disprefer vowels with different acoustic properties next to each other (like AGREE) or that explicitly prefer that an acoustically based feature is linked to as many segments as it can, like the positive constraint SPREAD used in Kimper (2011)), and in the analyses shown in this chapter.

These phonetic influences on vowel production and perception provide a good basis for the argument that vowel harmony can be based on phonetic information alone. A proposal that phonetic information might play the strongest role in the grammatical patterns entailing vowel harmony runs into some problems with antiharmonicity. If coarticulation is responsible for vowel harmony, it is not clear how the same vowel can elicit certain behavior (eg. back suffixing in antiharmonic stems: *[hiːd-nɒk]* 'bridge-DAT') when it is found in some stems and different behavior (eg. front suffixing: *[viːz-nɛk]* 'water-DAT) in in others.

2.3.1 Phonetics first analysis

2.3.1.1 A phonetic solution

A possible solution for the problem of handling antiharmonicity within a phonetically based framework would be to claim that the instances of vowels in antiharmonic stems actually differ phonetically from instances found in harmonic stems. Under such a hypothesis, there would be two phonetically and phonologically contrasting vowels so there would be no vowel in the system with a contradictory behavior.

Benus and Gafos (2007) provide data that can be interpreted to be a solution for this problem. In their articulatory study on Hungarian transparency and antiharmonicity, they report a correlation between stem type and the backness of [i(ː)]. The results indicate that [i(ː)] in antiharmonic stems is articulated with a more retracted tongue body than in a harmonic stem even in isolation. For example, the tongue body during the vowel in antiharmonic stems like *[viːv]* 'he is fencing' and *[iːr]* 'he is writing' is more retracted than in corresponding harmonic minimal pairs like *[iːv]* 'bow' and *[hiːr]* 'news'. Similarly, the preliminary results of Benkő and van de Vijver (2016) also show some *acoustic* differences between these two sets of stems.

2.3.1.2 Radical model

An extreme model of the effects of phonetic detail on phonology is that phonetic facts are exclusively responsible for the morphophonological behavior of stems. Under this model, there is no need for a more abstract phonological representational level: if the vowel is back phonetically (it has a low F2 or retracted tongue root), the suffix will be back as well. In order to explain antiharmonicity one has to posit under this theory that the retracted allophone of /i/ in antiharmonic stems in Hungarian is realized as an actual back vowel, like [ɯ]. The model behind this proposal is sketched out on Figure 2.1.

This hypothesis is not plausible as it is evident that Hungarian does not have an allophonic high back unrounded [ɯ] vowel. The findings of Benus and Gafos (2007) indicated subtle, subphonemic differences. They have also expressed doubt over the perceptibility of these differences, which would certainly be an issue for a proposal with fully back [ɯ] vowels. Benus (2005) discusses an analysis of Hungarian transparency where [i̠] evokes back

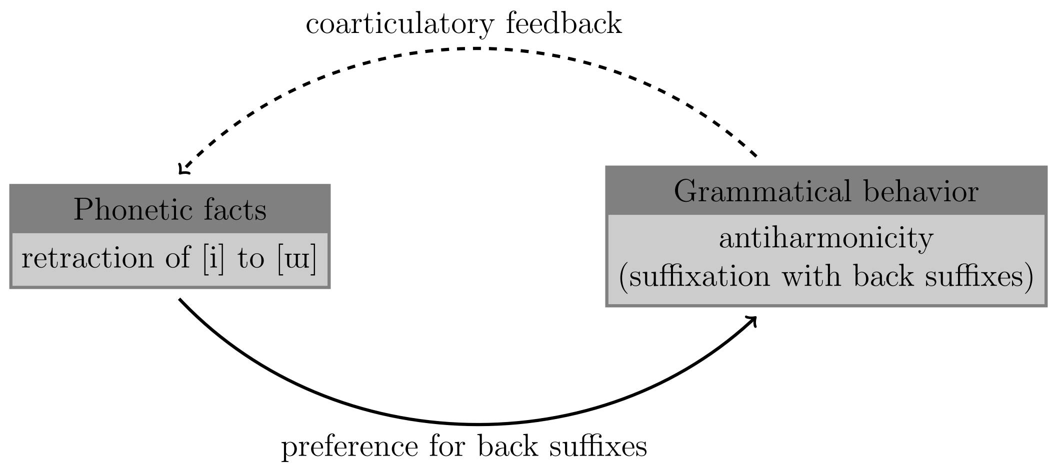


Figure 2.1: Implausible extremely phonetically based model of antiharmonicity

harmony, even though it is not back itself, because of the fact that it contrasts with a fronter [i] (although Kimper 2011, p. 215–223 shows that this analysis predicts unattested patterns as well, and presents problems with stem-internal harmony in Hungarian), so it is clear that positing a radical system with a back [ɯ] vowel is not viable for a phonetically based analysis.

2.3.1.3 Plausible model

These arguments illustrate that there needs to be an abstract phonological level that plays a role in antiharmonicity. The strongly phonetically driven analysis of antiharmonicity examined in this paper uses a less extreme model. This hypothesis claims that there is segmental level difference between the vowels in harmonic and antiharmonic stems, which explains the irregularity of antiharmonic roots. There is, however, an intermediate representational level of abstract grammatical knowledge. The understanding that a retracted (but still front) vowel [i̠] is phonologically [+back] is found on this level, and that is why a [+back] feature is spread onto the suffix. Phonetic retraction can introduce a bias for antiharmonic

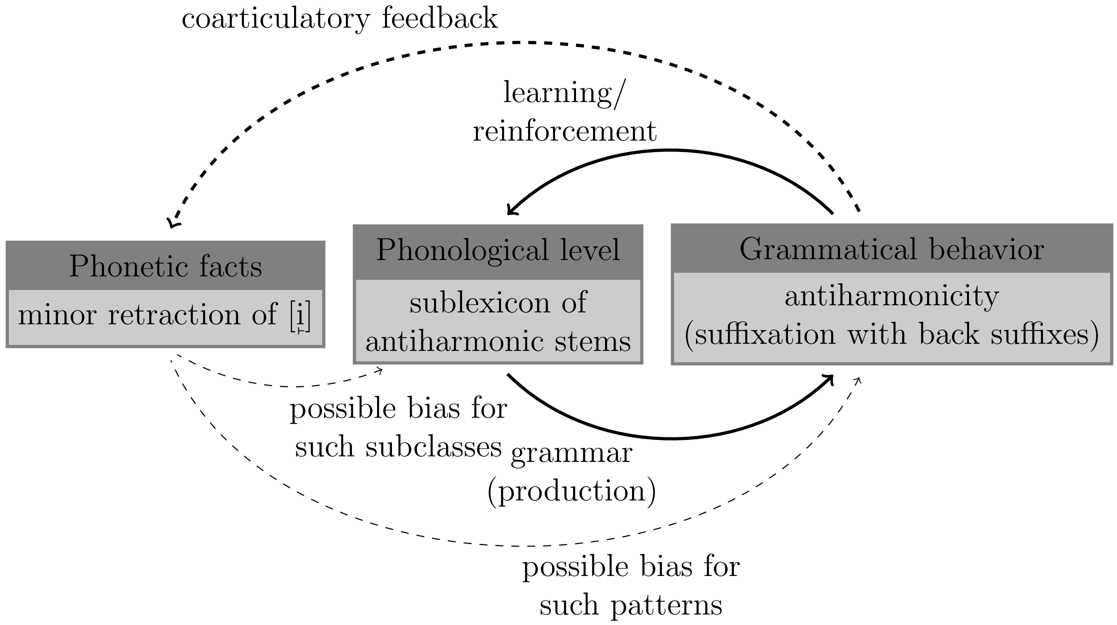


Figure 2.2: Phonetically biased lexical model of antiharmonicity

patterns to occur and would add a bias towards antiharmonicity for affected lexical items. This explanation is summarized below, and is illustrated on Figure 2.2:

(178) *phonetics first explanation for antiharmonicity*: Learning which stems are antiharmonic is assisted by stem vowels introducing a phonetic bias to the learner, being more retracted in antiharmonic stems than in harmonic stems.

In such an analysis, the grammatical behavior of stems is greatly influenced by phonetic substance. The tongue retraction of [i̙] leads to the stem having to be suffixed with a back vowel. There is a feedback in the system to reinforce the retraction for vowels in antiharmonic stems as these vowels, when the stem is suffixed, are coarticulated with the suffix vowel leading to retraction.

2.3.1.4 Analysis

If phonetic detail can explain antiharmonicity, there is no need to handle the stems as exceptional: these stems could be described as obeying vowel harmony too. Antiharmonic (and back-transparent) stems would have an an underlying retracted high front vowel, whereas harmonic stems would contain full, non-retracted [i] vowels. So under this hypothesis the representation of antiharmonicity would be marked on the vowel: there would be a retracted /i̠/ category on the phonological level, which is to be found in antiharmonic stems (a similar analysis has been employed by Vago (1980), who also handles antiharmonicity on the segmental level). On the surface, an underlying full vowel would appear as a peripheral vowel, but the surface form of underlying /i̠/ would be somewhat retracted:

(179) */siːv/* ∼ */siːvnɛk/* → *[siːvnɛk]*

(180) */ɲi̠ːl/* ∼ */ɲi̠ːlnɒk/* → *[ɲi̠ːlnɒk]*

This is different from the radical model as the retracted segment is not back phonetically. It is a segmental level solution, however, as the underlying segment /i̠/ is back representationally, which manifests itself as a small retraction on the phonetic level.

A possible analysis building on the foundation introduced in Section 2.2 is presented below.

Under a phonetics first model, the weight of the *(-Round,+Back) constraint needs to be decreased, so that actual retracted output *hi̠ːdnɒk* can emerge as the winner, as seen in the tableau in Table 181, where the weight of this constraint is 0. The key of this derivation is that the underlying form, just as the surface form, already contains the information on a segmental level that the stem is antiharmonic, as it contains a retracted front unrounded vowel.

(181) Analysis of an antiharmonic root under the phonetics first hypothesis

[+] [-] /h i̠ː d-nɛk/	Spread(+Bk) 4	Spread(-Bk) 2	Ident(Bk) 0.5	*(-Rd,+Bk) 0	$\mathcal{H}$
☞ [+] hi̠ːdnɒk	+4		-0.5	0	+3.5
[-] hiːdnɛk		+2	-0.5		+1.5
[-] [+] hi̠ːdn ɛ k				0	0.0

An important part of the analysis here is the [+back] feature value of the retracted [i̠] vowel. This retracted vowel obviously does not have an F2 as low or a tongue as retracted as phonetically back vowels like [ɒ], or [u]. However, for the phonological analysis it will be assumed that this value bears a [+back] feature, which is justified because under this analysis, /i̠/ stands in contrast with /i/ the same way /u/ and /y/, or /ɒ/ and /ɛ/ are contrasted in Hungarian.

Given the analysis in (181) a reanalysis for transparency is required, as *(-Round,+Back) has a lower weight than Ident(Bk), rendering the solution for transparency in the basic analysis in (165) ineffective. The new solution builds on an additional observation of Benus and Gafos (2007), that transparent front vowels surrounded by back vowels are also retracted. The winning candidate is therefore (182c), the one where spreading goes through. Again, the spreading of a [+back] feature happens on a phonological, not on a phonetic level, so the result of spreading on /i/ is not *[ɯ], but a phonetically front, but retracted [i̙]. Compare the tableau in (165) with (182) below:

(182) Analysis of transparency under the phonetics first hypothesis

	[+] [-] [+] | | | /g u m i n ɒ k/	Spread(+Bk) 4	Spread(-Bk) 2	Ident(Bk) 0.5	*(-Rd,+Bk) 0	$\mathcal{H}$
	[+] [-] a. g u m i n ɒ k	+4*0.8				+3.2
	[+] [-] [+] | | | b. g u m i n ɒ k					0.0
☞	[+] c. g u m i̙ n ɒ k	+4*2		-0.5	-0	+7.5
	[+] [-] | d. g u m i n ɛ k		+2	-0.5		+1.5

The opacity of a low front unrounded vowel can be handled by a more specific *(-Round,+Low,+Back) (=*ɛ̙) markedness constraint. If the weight of this constraint is greater than 0, spreading on the low vowel is avoided, and the result is opacity. The locus for the attested variation between opaque and transparent /ɛ/ is the weight of this constraint in this model, as if its weight is 0, the form with a back suffix is able to tie the winning candidate:

(183) Analysis of variation after [ɛ] under the phonetics first hypothesis

[+] [-] [+] /ʤ u ŋgɛln ɒ k/	SPREAD(+BK) 4	SPREAD(-BK) 2	IDENT(BK) 0.5	*(-RD,+BK) 0	*ɛ̠ 0	$\mathcal{H}$
[+] [-] ʤ uŋgɛ lnɒk	+4*0.8					+3.2
[+] [-] [+] ʤ u ŋgɛln ɒ k						0.0
[+] [+] ☞ ʤ uŋgɛ̠ln ɒ k	+4		-0.5	-0	-0	+3.5
[+] [-] ☞ ʤ u ŋgɛ lnɛk		+2*2	-0.5			+3.5

The consequence of this analysis is that under transparent conditions /ɛ/ is also retracted. Benus and Gafos did not investigate the behavior of the low front vowel, so this question is open for future research. They have pointed out that retraction of /ɛ/ is *not* expected, as a slight retraction of [i] is still perceived as the same category due to high front vowels being more quantal in nature: the perception of [i] is perceptually more stable given some articulatory variation than [ɛ]. Confirming this explanation, Szeredi (2012) has shown that a small retraction of [ɛ] to ≈230 Hz [ɜ] is often perceived as belonging to the /ɒ/ category, so such a variation has indeed more drastic effects on perception for a lower front vowel.

These examples illustrate the main advantage of a phonetics first analysis on the phonological theory level: there is no additional machinery needed to account for antiharmonicity.

The constraints needed for regular harmony and the models built on them are enough to explain antiharmonic behavior, if there is a phonetic difference on the segmental level.

2.3.2 Lexicon first analysis

2.3.2.1 Model

The concurrent hypothesis is that phonetic information is at most secondary in determining morphophonological behavior: antiharmonic stems are members of an irregular lexical class. Consequently, there is no underlying or surface representation where the vowel in an antiharmonic stem is different from vowels in harmonic stems. If there is any alignment between phonetic information (articulatory or acoustic) and antiharmonicity, that can be explained by secondary external factors. One of these explanations can be that coarticulation between the antiharmonic vowel and the back suffix leads to retraction which does not play any role in the grammar. Another explanation is diachronic: the result of an incomplete merger of two previously separate phonemes (such like an /i/ and an /i̙) can lead to different behavior of stems that contained one of these phonemes (like antiharmonicity for /i̙/) than stems containing the other one (regular harmonic /i/ stems). In either case, the synchronic grammar does not rely on these phonetic differences under any of these explanations – it marks antiharmonicity on the stem as a lexical class marker:

(184) *lexicon first explanation for antiharmonicity*: Antiharmonicity of stems is a property which has to be stored in the lexicon.

The relationship between phonetic facts and grammatical information is not very important in a model based on this hypothesis. There might be some effects of coarticulation on the stem vowels, but even if they are there, phonetic differences do not influence higher level behavior, as that is decided on the abstract phonological level:

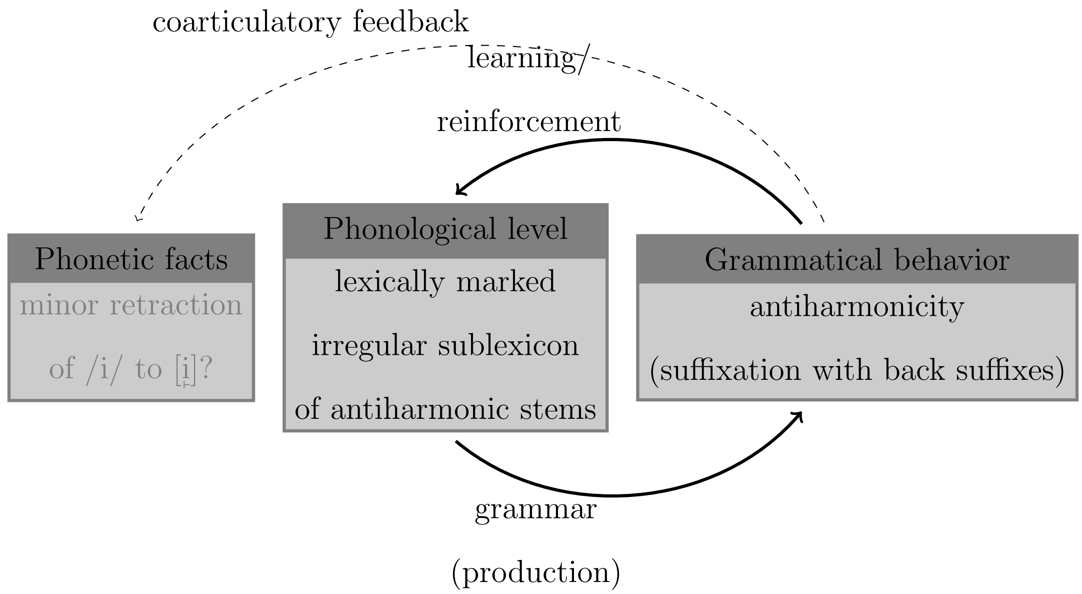


Figure 2.3: Lexically based model of antiharmonicity

2.3.2.2 Analysis

Traditionally Hungarian grammars have used this approach, and used lexical marking as in (185–186), either employing a floating [+back] feature or a morphological class marker (Nádasdy and Siptár 1994, Ringen and Vago 1998, Siptár and Törkenczy 2000). Contrast these representations, where the exceptionality is marked on the morpheme with (179–180 in the phonetics first approach, were exceptionality is marked on the vowel:

(185) */siːv/* ∼ */siːvnɛk/* → *[siːvnɛk]*

(186) */ɲiːl*$_{\mathrm{AH}}$*/* ∼ */ɲiːl*$_{\mathrm{AH}}$*nɒk/* → *[ɲiːlnɒk]*

A similar question of representation arises in other phenomena in other languages frequently. The question of Slavic 'yers' seems to be similar: the unpredictable vowel-zero alternation in certain Slavic morphemes is traditionally analyzed on the segmental level using abstract phonemes (yers) (eg. Lightner 1972), but Gouskova (2012) gives an analysis of the same phenomenon using exclusively morpheme-level marking of exceptionality.

The formal analysis that uses lexical marking must use some machinery that influences the evaluation of candidates that is sensitive to the lexical category morphemes belong to.

2.3.2.3 Scaling factor for antiharmonicity

To solve the problem of antiharmonicity, a tool is needed that conforms the Harmonic Grammar approach in this chapter, and is able to manipulate harmony scores assigned to candidates. There is such a tool that Kimper uses extensively: scaling factors. In the analysis below, a lexically marked scaling factor will be used, instead of lexically marked markedness or faithfulness constraints. The scaling factor approach fits seamlessly into the analysis having been developed in this chapter so far, and the lexically indexed constraint approach has more issues as it will be illustrated at the end of this section. To make the analysis fit in the system in Kimper (2011) the scaling factor will be multiplicative. This factor will be used in the following way:

(187) s_{AH}: a multiplicative scaling factor, which a Spread constraint is multiplied with if the trigger for the spreading is found in a morpheme that comes from the antiharmonic sublexicon, marked as AH

While Kimper's scaling factors are determined by vowel quality or distance between trigger and target, there is no disadvantage if scaling factors are to be determined by lexical marking. There is precedent for using such factors: Linzen et al. (2013) and Gouskova and Linzen (2015) use such additive scaling factors to account for variability in the deletion of vowels (analyzed earlier as 'yers') in the Russian prepositions *s(o), k(o), v(o)*. There are multiple scaling factors used in their analyses, some pertaining to only a given lexical item, others to certain sublexica (like religious terms). All of the relevant weights are added to the baseline weight of a constraint which they affect, like the faithfulness constraint motivating against deletion of the vowel (Max-V). So a scaling factor $s_{ecc} = 3$ that is added to the weight of the constraint for words whose meaning is ecclesiastic would mean that the vowel of the preposition before these terms is less likely to be deleted, which matches the observations of their experiments.

This explanation for antiharmonicity works as shown in the tableau below (Table 188). The antiharmonic stem /*hiːd*/$_{\text{AH}}$ 'bridge' comes from the AH sublexicon, therefore spreading from it incurs the penalizing scaling factor, which is set to 0.2.

(188) Analysis of antiharmonicity under the lexicon first hypothesis

[-] [+] | | /hiːd$_{\text{AH}}$-n ɒ k/	*(-Rd,+Bk) 8	Spread(+Bk) 4	Spread(-Bk) 2	Ident(Bk) 0.5	$\mathcal{H}$
[-] hiːdnɛk			+2*0.2	-0.5	-0.1
[-] [+] | | ☞ hiːdn ɒ k					0.0

In such an analysis, faithfulness and constraints enforcing vowel harmony come to a conflict with each other and a more faithful output is the winner for antiharmonic stems. Unlike other kinds of exceptionality patterns, this analysis suggests that antiharmonicity doesn't require exceptional stems to pick allomorphs differently, it just puts emphasis on the resistance for changing input-output correspondences.

The toolbox of scaling factors assembled at this point can also predict that the antiharmonic trigger cannot be low, since the s_{trig} multiplier for spreading from low triggers counteracts s_{AH}. This is illustrated below: even if /*hɛɟ*/ 'mountain' is marked with the antiharmonic marker AH, it still cannot *behave* antiharmonically:

(189) No low antiharmonic stems

[-] [+] /hɛɟ$_{\text{AH}}$-nɒk/	*(-Rd,+Bk) 8	Spread(+Bk) 4	Spread(-Bk) 2	Ident(Bk) 0.5	$\mathcal{H}$
[-] ☞ hɛɟnɛk			+2*0.2*2	-0.5	+0.1
[-] [+] hɛɟnɒk					0.0

Languages with antiharmonic low front vowels could be imagined (with a larger s_{AH}, or a smaller s_{trig}), but the typological prediction is that there are no languages where only low front vowels (and not high front vowels) are going to be antiharmonic. This prediction is clearly limited to backness harmony systems as the scaling factors for phonetic grounding would behave differently for a different harmonizing feature.

2.3.2.4 Problems with the analysis

This approach does admittedly have some problems. First, a crucial assumption in this analysis is that all alternating suffixes are back underlyingly, as the constraint which can incur any negative harmony score on the losing candidate is the faithfulness Ident(Bk) constraint. In this analysis, therefore, back harmony cannot be triggered, as back suffixation is the default mechanism and front spreading is the only unfaithful process, which the grammar can either let happen or block. Therefore an underlying front suffix would result in a necessarily positive harmony score for *[*hiːdnɛk*], causing it to beat the desired candidate [*hiːdnɒk*]. Making the scaling factor s_{AHf} negative could solve this problem, but that would lead to the SPREAD constraint assigning a negative score to a candidate, which would go against the theoretical point of SPREAD being a positive constraint.

The second problem is with transparency after an antiharmonic root. To illustrate this problem, consider the transparent behavior of the verbalizing/causative suffix /*-iːt*/ below. If another suffix follows /*-iːt*/, the [+back] feature is able to spread onto the following suffix after back stems, as in (190), while after front stems, the following suffix is naturally front, shown in (191). This behavior is clearly transparent, so /*-iːt*/ cannot be represented as an antiharmonic morpheme. After antiharmonic stems, however, the irregular stem is able to trigger back suffixation across the otherwise transparent causative /*-iːt*/ suffix, as in (192) below:

(190) after back stem: *hosːɒbː* 'longer' ∼ *hosːɒbːiːt* 'lengthen' ∼ *hosːɒbːiːtom* 'I lengthen it', **hosːɒbːiːtɛm*

(191) after front stem: *mɛlɛg* 'warm' ∼ *mɛlɛgiːt* 'warm up something' ∼ *mɛlɛgiːtɛm* 'I warm it up', **mɛlɛgiːtom*

(192) after antiharmonic stem: *ʃimɒ* 'smooth' ∼ *ʃimiːt* 'smooth something' ∼ *ʃimiːtom* 'I smooth it', **ʃimiːtɛm*

In [*ʃimiːt*], the first [i] is clearly antiharmonic, as it is responsible for back vowels following it all forms where possible, and the second [iː] is transparent. The lexicon first analysis introduced above fails to yield the correct output. First, the base *ʃimiːt* is derived in (193): spreading of [-back] can actually happen as the suffix has no back allomorph:

(193) Transparent suffix after an antiharmonic stem

[-] [-] | | /ʃ i m_{AH} iːt	*(-Rd,+Bk) 8	Spread(+Bk) 4	Spread(-Bk) 2	Ident(Bk) 0.5	$\mathcal{H}$
[-] [-] | | ʃ i miːt					0.0
[-] ☞ ʃi m iː t			+2*0.2		+0.2
[-] [+] | | ʃ i m ɨː t	-8			-0.5	-8.5

The derivation of *[ʃimiːtom]*, however, runs into the problem: the antiharmonicity of the stem is unable to be strong enough to rule out the candidate (194a) where frontness is spreading from the transparent suffix:

(194) Problem with antiharmonic stem followed by a transparent suffix

[-] [-] [+] /ʃ i m$_{\mathrm{AH}}$ iː t o m/	*(-Rd,+Bk) 8	Spread(+Bk) 4	Spread(-Bk) 2	Ident(Bk) 0.5	$\mathcal{H}$
[-] [-] a. ʃ i miː t ɛm			+2	-0.5	+1.5
[-] b. ʃi m iː t ɛm			+(2+2)*0.2	-0.5	+0.3
[-] [-] [+] c. ʃ i miːt o m					+0.0
[+] [-] d. ʃi m iːt o m			+2*0.2		+0.4

The solution for this problem is not covered in this chapter, as one might need some additional machinery to fix the issue. One such trick could be letting sublexical membership of a stem like /*ʃim-*/$_{\mathrm{AH}}$ to spread onto derived stems like /*ʃimiːt-*/. Another solution would need serial derivation, where the input for the tableau (194) is the output of (193). In this case, the candidate (195a) could be penalized for breaking up a pre-existing link, and the desired output (195) is able to emerge as the winner. A further exploration of the consequences for such a serial derivational analysis falls outside the scope of this chapter.

(195) Solution for the problem with serial derivation

[-] [+] /ʃi m iːt o m/$_{\text{AH}}$	*(-RD,+BK) 8	DEP(LINK) 6	SPREAD(+BK) 4	SPREAD(-BK) 2	IDENT(BK) 0.5	$\mathcal{H}$
[-] [-] a. ʃ i miː t ɛm		-6		+2	-0.5	-4.5
[-] b. ʃi m iː t ɛm				+(2+2)*0.2	-0.5	+0.3
[-] [-] [+] c. ʃ i miːt o m		-6				-6.0
☞ [-] [+] d. ʃi m iːt o m				+2*0.2		+0.4

2.3.2.5 Lexically indexed constraints

Lexically marked constraints from Optimality Theory analyses could be employed as an alternative to using scaling factors to distinguish how different stems weigh certain constraints differently. Pater (2010) analyzes the unpredictability of the behavior of Finnish *a*-final stems by assigning certain morphemes to an L sublexicon and using constraints like [ai]$_{\text{L}}$ that are specialized to assign violations only on morphemes belonging to this L sublexicon. Gouskova (2012) also uses a similar constraint, *MID$_{\text{L}}$ to account for deletion of mid vowels in a large subset of Russian roots. In Hungarian, this could mean again assigning antiharmonic morphemes to an AH sublexicon as in (185–186), and employing a lexically specified *negative* SPREAD(-BACK)$_{AH}$ constraint that penalizes spreading front from a morpheme in this sublexicon. For an antiharmonic stem like /hiːd/$_{AH}$, the analysis looks like in (196), where spreading is avoided by the lexically specified constraint.

(196) Analysis of antiharmonicity with lexically marked constraints

[-] [+] /hiːd$_{AH}$n ɛ k/	*(-RD,+BK) 8	SPREAD(-BK)$_{AH}$ 6	SPREAD(+BK) 4	SPREAD(-BK) 2	IDENT(BK) 0.5	$\mathcal{H}$
[-] hiːdnɛk		-6		+2		-4.0
☞ [-] [+] hiːdn ɒ k					-0.5	-0.5

There is a limited theoretical problem with this approach: letting a lexically specified SPREAD constraint assign negative weights goes against the definition and rationale behind this family of constraints. As defined in Kimper (2011), and used in this chapter, SPREAD constraints are able to *reward*, not to punish candidates where spreading happens. There is also a theoretical problem with using a lexically marked IDENT(BACK)$_{AH}$ constraint: it would be unable to assign violation marks to a [*hiːdnɛk*], even if the underlying representation for the suffix would be /*nɒk*/ as in the scaling factor analysis, because the faithfulness violation occurs in the suffix, not the lexically marked stem.

An alternative to this account would be a case where harmonic stems are the ones marked as special. In this case, the basic spreading constraint SPREAD(-BACK) would be ranked low with a weight of 0, and the lexically marked SPREAD(-BACK)$_{H}$ constraint is responsible for spreading in the harmonic cases. Assuming that the underlying form of the suffixes is again consistently back, we can get the desired result, because *_hiːdnɛk_ is avoided as spreading of [-back] does not happen in this system from an unmarked stem, while _viːznɛk_ is able to win because spreading is rewarded after a stem like /_viːz_/$_{\text{H}}$ marked with an H sublexical marker:

(197) No spreading from antiharmonic (unmarked) stems

[-] [-] | | /hiːdnɒk/	*(-RD,+BK) 8	SPREAD(+BK) 4	SPREAD(-BK)$_{\text{H}}$ 2	IDENT(BK) 0.5	SPREAD(-BK) 0	$\mathcal{H}$
[-] hiːdnɛk				-0.5	+0	-0.5
[-] [+] | | ☞ hiːdn ɒ k						0.0

(198) Front harmony after harmonic (marked) stems

[-] [-] | | /viːz$_{\text{H}}$nɒk/	*(-RD,+BK) 8	SPREAD(+BK) 4	SPREAD(-BK)$_{\text{H}}$ 2	IDENT(BK) 0.5	SPREAD(-BK) 0	$\mathcal{H}$
[-] ☞ viːzn ɛ k			+2	-0.5	+0	+1.5
[-] [+] | | viːzn ɒ k						0.0

While this solution fixes the problem without using negative SPREAD constraints, the implication that the regular and productive sublexicon of harmonic stems is marked as special in the grammar is problematic. Besides this problem, all stems with rounded front vowels or a low /ɛ/, which are never antiharmonic, have to be lexically marked by default. For these reasons, this explanation will not be pursued any further.

This section has shown a possible lexically based analysis of Hungarian vowel harmony, that fits seamlessly in the system developed in Section 2.2 based on Kimper (2011), by using a penalizing scaling factor on the spreading of a [-back] feature in a stem that is marked as antiharmonic in the lexicon. The rest of the chapter will present experimental evidence for the preferability of such a lexically based analysis over a phonetically based hypothesis.

2.3.3 Plan for experiments

The aim of the experiments in this chapter is to test the plausibility of a phonetic explanation for vowel harmony. The production experiment in Section 2.4 will follow up Benus and Gafos (2007) and test whether the articulatory differences have acoustic correspondences in Hungarian. The perception experiment in Section 2.5 will aim to determine the perceptibility threshold of F2 differences in high front vowels for Hungarian speakers to see if any differences found in the production experiment are available in perception to Hungarian speakers. Finally, a nonce word rating experiment will be described in Section 2.6 that was designed to determine if Hungarian speakers are able to use perceivable differences to classify a stem as harmonic or antiharmonic.

2.4 Production experiment

2.4.1 Goals and hypotheses

The concrete goal of the production experiment is to find out whether the acoustic correlates of the articulatory differences found by Benus and Gafos (2007) are present in the acoustics of Hungarian speakers. While Benus and Gafos argued that acoustic correlates are not expected to be present in the signal for the articulatory differences to be found, one would still expect acoustic differences to be present in the signal as well if a phonetically based explanation for antiharmonicity is to be accepted: how could a solely articulatory phenomenon be the basis of such a stable pattern?

If the explanation for antiharmonicity is to be mainly based on phonetic cues, these phonetic cues cannot present themselves exclusively in the articulatory domain. Presuming so, one would assume that the articulatory differences develop independently for each speaker during language acquisition. If the difference between a vowel in a harmonic stem and a vowel in an antiharmonic stem is only present in articulation, there would be no channel of communication where the knowledge of this difference could spread. During the acquisition of the language, a speaker has no access to any information that could tell them that articulatory differences should be used to set apart harmonic and antiharmonic stems. With no such information, a learner of the language is much more likely to attribute the different behavior to lexical classes (antiharmonic vs. harmonic), and the articulatory differences could only be secondary, as similar coarticulatory patterns can develop independently for different speakers. In this case, these articulatory differences could not play an explanatory role in the grammar.

To determine whether the articulatory differences found by Benus and Gafos (2007) have correspondence in acoustics, a production study was conducted. The main comparison was between the average F2 (the acoustic variable corresponding tongue frontness/backness) of [i(ː)] vowels in harmonic stems and the average F2 of the [i(ː)] vowels in antiharmonic stems.

There are three possible results of the experiment: a significant acoustic difference in the expected direction (vowels in antiharmonic stems having lower F2), a non-significant but somewhat systematic acoustic difference in the expected direction or no acoustic difference (F2 the same in harmonic and antiharmonic stems) in the expected direction. If significant differences are found, the phonetics first explanation for antiharmonicity in vowel harmony would become much more realistic, since it would be shown that the phonetic information on antiharmonic stems could propagate through the surface form. If no acoustic difference is found or a difference in the unexpected direction is found, the phonetics first explanation could be seriously questioned, as the articulatory differences could not be exclusively responsible for the antiharmonicity pattern. If the result is in the expected direction but not significantly, not for all speakers or not in all environments, further studies should be needed to see if Hungarian speakers could be more trained than average to trends of F2 differences in front vowels: if such effects are found, the phonetics first hypothesis could not be immediately discarded.

antiharmonic F2 < harmonic F2	significance	phonetics first explanation
✓	✓	✓
✓	×	further investigation
×	–	×

Table 2.3: Evaluation of the phonetics first hypothesis based on the results of the production study

This study was piloted with fewer stimuli, only in isolation and with only 7 speakers from Budapest and 5 from Párkány (Štúrovo, Slovakia) in Blaho and Szeredi (2013). This pilot did not find consistent significant differences between the acoustics of harmonic and antiharmonic stems in Hungarian. The experiment discussed in this section will expand the number of stimuli and subjects, improve the quality of recordings and investigate not only plain isolation, but suffixed forms and words in frame sentences as well.

2.4.2 Methods

2.4.2.1 Participants

There were 16 participants in the study, who were recruited through a shared Facebook post. All lived in Budapest, were over 18 years old at the time of the experiment, and reported no hearing or speech disorder. Two subjects failed to fill out the demographic form, so no further details are stored for them. The average age of the remaining 14 was 25.5 years (ranging 22–29), 9 females and 5 males. Three participants reported living 6 months or more in a different country than Hungary (in Netherlands, Germany and USA). Of these 14 participants, 13 reported some level of proficiency in English, followed by 8 in German, 2 in Italian and Norwegian, and 1 in Spanish, Norwegian, French, Hindi and Russian. The subjects were all from an urban background: 13 were from Budapest and 1 from another major city in Hungary.

2.4.2.2 Stimuli

The goal of the experiment was to determine the difference between the acoustics of a Hungarian high front vowel in a harmonic stem and in an antiharmonic stem. The set of target words therefore contained /i/ in harmonic and antiharmonic stems. Both long and short vowels were investigated, as the effect of vowel length on the possible difference between harmonic and antiharmonic stems is not known. Fillers with /ɛ/ and short and long /u/ were also included, partly so that a vowel space can be imaged for each speaker. The distribution of the stem vowels is seen in Table 2.4:

	i-stems				fillers		
vowel	/i/	/i/	/iː/	/iː/	/ɛ/	/u/	/uː/
harmonicity	harmonic	antiharmonic	harmonic	antiharmonic			
count	15	7	10	16	25	11	14

Table 2.4: Distribution of stem vowels in the production study

The participants read the target words of the experiment from a computer screen. There were three conditions on how the target word was presented in the experiment: in *isolation*, *framed* in a sentence and in a *suffixed* form. The isolation condition is designed to replicate the condition used by Benus and Gafos (2007), who recorded their stimuli for the articulatory study with the words uttered in isolation. In the frame sentence condition, test words were embedded in the sentence *[ɒ taːblaːrɒ ɒ* X *soː volt fɛliːrvɒ]* 'The word X was written on the board', to see if a more natural environment elicits a different pattern of separation between the harmonic categories. Finally, in the suffixed condition the dative marker *[nɒk/nɛk]* was attached on nouns, while the plural 3rd person suffix *[(ɒ)nɒk-(ɛ)nɛk]* was attached on verbs to find out the magnitude of how coarticulation influences the stem vowel in suffixed forms.

The target words were the same in the isolation and in the suffixed condition, but their order was differently randomized. In the frame sentence condition, a subset of the stimuli (52 out of 100) was used. All stimuli were monosyllabic stems (suffixed in the suffixed condition).

The stimuli contained both nominal and verbal stems, eg. [*his*] 'believe' ∼ [*his-nɛk*] 'they believe', [*ɲit*] 'open' ∼ [*ɲit-nɒk*] 'they exterminate', [*hiːr*] 'news' ∼ [*hiːr-nɛk*] 'news.DAT', [*ɟiːk*] 'lizard' ∼ [*ɟiːk-nɒk*] 'lizard-DAT', [*tɛʃt*] 'body' ∼ [*tɛʃtnɛk*] 'body-DAT', [*bus*] 'bus' ∼ [*bus-nɒk*] 'bus-DAT', [*luːd*] 'goose' ∼ [*luːd-nɒk*] 'goose-DAT'.

2.4.2.3 Procedure

The utterances of the participants were recorded using a Beyerdynamic DT 770 M head mounted microphone with a PMD 670 Marantz digital recorder. The subjects were instructed to read the target words and frame sentences the most natural way, without overarticulation and without speaking too casually. The task did not take more than about 10 minutes.

2.4.2.4 Analysis

The recordings were analyzed in Praat (Boersma and Weenink 2012). The vowels in the target words (including fillers) were segmented and the formant values of F1–F3 were extracted by a Praat script, looking for maximum 5 formants up to 5000 Hz for men and 5500 Hz for women with a time step of $0.01s$. The vowel space was imaged for each condition and each speaker to see if there were any problems with the measurements – if the vowel space was distorted, a new formant analysis was run with a higher setting for the maximum formant frequency, which solved most issues.

There were some items where an F2 formant was measured incorrectly due to some noise in the recording. This was important to fix for target words with an [i(ː)]. To eliminate this problem, certain [i(ː)] tokens were manually checked and measured again. These tokens included all [i(ː)] vowels that were below more than 2 standard deviations from the mean in F2 ($F2 < \mu(F2) - 2\sigma F2$) of all [i(ː)] tokens (F2<1625 Hz) were, and some measurements with F2<1900 of female speakers that were clearly not correct. This resulted in the reassessment of 47 tokens out of 1012 (4.6% of all [i(ː)] stimuli).

After the visual analysis of each speaker's behavior, the data was aggregated over speakers and the formant values were normalized using the Lobanov method in the `vowels` R package (Kendall and Thomas 2012). Average differences between conditions were taken at this point.

Statistical testing of significance of these differences was made by building a linear mixed effects regression model on the data containing only [i(ː)] vowels with the `lme4` package in R (Bates et al. 2014). The model predicts the non-normalized F2 with the vowel length and harmonic category (harmonic or antiharmonic) and their interactions as independent variables, and with random slopes for these variables by speaker and a random intercept by item. Normalized values were not used in the mixed effects models as they introduced a manipulation in the data for a problem that the subject-wise random effects could also handle and that possibly caused the algorithm of `lme4` to achieve singular convergence, corrupting the results. The optimization for the full random model structure with all the interactions present for the random effects did not converge, therefore random interaction slopes are not in the models analyzed below.

2.4.3 Results

The results of the experiment will be summarized and discussed separately by condition below. The expected difference in F2 between harmonic and the antiharmonic stems did show up in all conditions, but the effect only reached significance in the suffixed condition, where it was expected to be more emphasized based on coarticulation. The effect size of this difference was found to be quite small overall, at most 33.9 Hz in the suffixed condition for long vowels, which is much lower than the just noticeable difference of ≈110 Hz found in most studies.

To illustrate the results visually, Figure 2.4 shows the vowel spaces of the [i(ː)] vowels only for each condition. Each vowel chart shows the mean normalized and scaled position for long (marked as *ii*) and short (*i*) vowels in harmonic (simply *i, ii*) and antiharmonic (*i-, ii-*) stems.

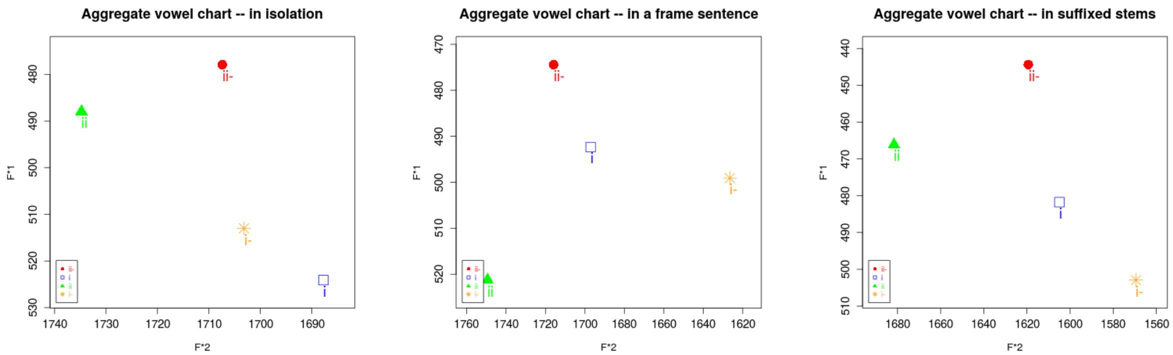


Figure 2.4: Vowel spaces for harmonic and antiharmonic [i(ː)] in each condition (the − suffix marks antiharmonicity)

2.4.3.1 Averages

The direction of the difference between the harmonic and antiharmonic conditions was found to be inconsistent both across speakers and conditions. Table 2.5 below summarizes the average F2 values of the [i] vowels in the two harmonicity values and conditions. It can be seen that the effect predicted by Benus and Gafos was borne out in all conditions, as vowels in antiharmonic stems had a lower F2 than the vowels in harmonic stems. However, the effect is very small, 13.6 Hz for the short /i/ and 18.2 Hz lower for the long /iː/ vowel. The difference between short and long vowels also seems to confirm the presence of the effect predicted by Benus and Gafos (2007), since the fact that the vowels are longer could minimize coarticulatory factors of neighboring consonants, so the underlying difference between harmonic and antiharmonic vowels could be more easily seen.

If the retraction effect found by Benus and Gafos is explained by the coarticulatory retraction in suffixed forms affecting paradigmatically the vowels in the isolated forms, a substantial retraction is expected in the suffixed condition. Although there was a retraction found in the suffixed condition for both short and long [i] vowels, the effect size was again minuscule (≈17–34Hz).

	[i]			[iː]		
	harmonic	antiharmonic	difference	harmonic	antiharmonic	difference
suffixed	1981.4	1963.5	17.9	2016.5	1982.7	33.9
isolation	2008.1	2000.6	7.5	2018.9	2010.2	8.7
frame	1987.8	1967.7	20.0	2004.4	1993.0	11.4

Table 2.5: Average normalized and scaled F2 values of high front vowels by stem type, vowel length and condition

To measure the effect of `harmonicity` more exactly, each condition was be examined with linear mixed effects models separately, and each condition investigates quite different hypotheses.

2.4.3.2 In suffixed forms

There is a small coarticulatory effect found in suffixed forms, as [i(ː)] vowels are more retracted in antiharmonic stems, when they are followed by a back vowel than in harmonic stems, when they are followed by a front vowel (the difference is 17.9 Hz for the short and 33.9 Hz for the long vowel). There were 12 subjects out of 16 that showed this effect, and 4 that showed an *opposite* effect, retracting harmonic [i(ː)] vowels more than antiharmonic ones. The range of harmonic-antiharmonic differences was −75.8 Hz – 177.71 Hz, positive values referring to the expected direction (M=33.78 Hz, SD=56.37 Hz).

These values were tested with a linear mixed effects model as well: the saturated model had `length` and `harmonicity` and their interaction as predictors, and subject-wise random slopes of these two factors and both subject-wise and item-wise random intercepts. The interaction could be eliminated from the saturated model, and both `length` and `harmonicity` turned out to be significant both with t test and with a likelihood ratio test, as seen in Table 2.6: short vowels and vowels in antiharmonic stems turned out to be more retracted

factor	β	SE	t	χ^2 (df)	p	
Intercept	2394.54	66.28	36.13			
`length=short`	-57.3	18.79	-3.05	8.39 (1)	0.004	**
`harmonicity=harmonic`	54.05	18.11	2.98	8.03 (1)	0.005	**

Table 2.6: Coefficients of the final model for the suffixed vowels

again.

These results indicate that there is indeed a significant retracting effect of coarticulation, although the size of this effect remains far below the untrained Just Noticeable Difference of $\approx$ 110 Hz (Mermelstein 1978, Kewley-Port 2001).

2.4.3.3 In isolation

The harmonicity of the stem turned out to play no significant role in the F2 of a word in isolation, contrary to the predictions based on Benus and Gafos (2007). While there was a 8.7 Hz difference in the expected direction for long vowels and a 7.5 Hz difference for the short [i], statistical analysis shows that this difference is not significant. The interaction of `length` and `harmonicity` was not significant here either, so the resulting model is the same as the one for vowels in suffixed words – but neither `length` or `harmonicity` turned out to be significant (see Table 2.7).

By-subject analysis shows that while most speakers (10 out of 16) did show a difference in the expected direction again, there were 6 speakers who differentiated in the unexpected direction. The range of by-subject differences was -106.6 Hz – 111.7 Hz, the overall mean was 10.1 Hz and the standard deviation was 50.85 Hz.

factor	β	SE	t	χ^2 (df)	p	
Intercept	2416.29	63.31	38.17			
`length=short`	-21.83	23.41	-0.93	0.849 (1)	0.35	n.s.
`harmonicity=harmonic`	17.26	17.73	0.97	0.931 (1)	0.33	n.s.

Table 2.7: Coefficients of the model with the non-significant factors for vowels in isolation

factor	β	SE	t	χ^2 (df)	p	
Intercept	2407.97	65.02	37.03			
`length=short`	-39.38	21.68	-1.82	3.102 (1)	0.078	.
`harmonicity=harmonic`	27.42	21.52	1.27	1.571 (1)	0.21	n.s.

Table 2.8: Coefficients of the model with no significant factors for vowels in frame sentences

Based on these results, the findings of Benus and Gafos are not replicated: there was no significant effect found for harmonicity of a stem in the F2 of the vowel in isolated words, although there was a minor trend towards the predicted results.

2.4.3.4 In frame sentences

When embedded in a frame sentence (before a word with a back vowel, explaining the overall lower F2), the effect of harmonicity fails to show up again: short [i] vowels are retracted by 20 Hz on average, while long vowels show a smaller backing of 11.4 Hz. The linear mixed effects models show that this difference is not significant again, as shown on Table 2.8. The `length:condition` interaction was not significant in this data set either, only the `length` condition approached significance.

By-speaker analysis shows that 5 speakers out of 8 have a difference in the unexpected direction and some of these might be somewhat outlying with a 135.4 Hz and a 104.7 Hz direction by two speakers. In the expected direction, there are 5 speakers with differences over 50 Hz. The overall range is -135.4 Hz – 79.1 Hz, with a mean of 15.28 Hz and a standard deviation of 62.27 Hz.

The two speakers with the differences over 100 Hz might be treated as outliers — their results might be more contaminated with noise due to the faster speech tempo found in this condition. If their data are discarded, the `harmonicity` condition becomes borderline significant ($\chi^2(1) = 3.708$, $p = 0.054$), and the effect size based on the coefficient would raise to $\beta = 48.77$Hz. This would not change the overall conclusion: there is a trend toward a more retracted realization for the /i/ vowels in antiharmonic stems, but this effect is at most borderline significant, highly variable across speakers and it is doubtful that it can be perceivable.

2.4.4 Discussion

The results above have have shown that the effect found by Benus and Gafos (2007) is not borne out consistently. It is very weakly found both in a frame sentence where speakers were able to be less artificially focused on the task, and in isolation, in the environment where Benus and Gafos found the articulatory effect. The results of the pilot experiment described in Blaho and Szeredi (2013) also showed a weak, inconsistent and non-significant effect to be present in isolation, while Benkő and van de Vijver (2016) found a small, yet significant effect of retraction in antiharmonic stems.

The results of the suffixed condition show that coarticulation with the following suffix vowel might have an effect on the stem vowel, giving some legitimacy to the hypothesis that this coarticulatory effect in suffixed forms drags the stem vowel of unsuffixed antiharmonic stems to a more retracted quality.

The differences by subject in the suffixed condition and in the frame sentence condition (r=0.15, p=0.59, n.s.) and in the isolated condition (r=0.31, p=0.25, n.s.) show a positive correlation, so speakers with higher coarticulatory effects show a bigger separation in the frame sentences. However, the correlation of the differences in a frame sentence and in isolation is negative (frame vs. isolation r=-0.41, p=0.11, n.s.), which might be assigned to a task effect of the careful speech pronunciation in the isolation condition.

The differences found above are mostly all below the JND based on previous literature of ca. 110 Hz. There are some speakers for whom the difference in frame sentences or in isolation approaches this figure, one participant even surpasses it (barely, though) with 111.72 Hz in isolation, but mostly these differences are around 10–15 Hz. The pattern is not consistent across speakers, as seen in the fact that these differences go in the unexpected direction for 21.4% of the speakers in the frame sentence condition and 42.8% of the speakers in the isolated condition. Therefore, the presence of a segmental level distinction based on the phonetics first explanation for antiharmonicity is questionable based on these results. The perceptibility of this small effect found is strongly questionable, but it is possible that Hungarians are attuned to such minor distances between high front vowel categories. The following experiment tests whether this is the case.

2.5 Perception experiment

The goal of the perception experiment was to determine the perceptibility threshold of a difference in the F2 of a high front vowel for Hungarian speakers. Based on earlier studies, the JND without training for such a vowel is at least $\approx$ 110–200 Hz in isolation for American English (Mermelstein 1978, Kewley-Port 2001) and Finnish (Aaltonen et al. 1997) speakers.

Certain studies on the JND of high front vowels, however, have shown a large effect of training. According to a study by Kewley-Port (2001), the discrimination threshold of ≈ 110 Hz can even be brought down to 46–62 Hz for vowels in syllables with 2-3 days of training. Based on these results, it could be argued that Hungarians are trained to be more attuned to smaller phonetic differences if these differences are meaningful in the grammar, as they are according to the phonetics first hypothesis. Therefore finding out the perceptibility threshold is necessary to evaluate the hypotheses about the explanation for antiharmonicity.

If the effect found by Benus and Gafos (2007) is indeed a segmental level phenomenon, then it is conceivable that the corresponding acoustic difference would be perceivable (or quite close to perceivable) for Hungarian speakers. In this case, the threshold for perception would fall on the lower part of the above range, at most around 110-150 Hz. If this low threshold is not found, then it is questionable if a pattern described by Benus and Gafos can be used by the Hungarian speakers in their representation of vowel harmony in the grammar.

perceptibility threshold	are Hungarians trained?	plausibility of phonetics first
$\leq$150 Hz	probably	✓
200 Hz	probably not	×
$\geq$ 250 Hz	definitely not	××

Table 2.9: Evaluation of the phonetics first hypothesis based on the results of the perception study

To find the threshold of perception, an ABX discrimination study was conducted. The initial difference tested was 200 Hz, as it was expected that this difference will be easily discriminated and smaller differences could be tested. This was not the case, though – the 200 Hz difference presented a lot of difficulties for the listeners, therefore the next condition employed a 250 Hz difference, which was more manageable for the listeners.

2.5.1 Methods

2.5.1.1 Participants

Participants were recruited through two means: a Facebook post that was shared by people in the author's social circle and a post on the microblogging site Tumblr that was reblogged by many.

There were 28 subjects in the perception experiments, 14 in the 200 Hz condition and 14 in the 250 Hz condition. One participant is not included in these summaries as they were excluded from the experiment having reported some hearing impairment. Gender was not perfectly balanced: 16 of the participants were female and 12 were male. The ages of the participants ranged from 18–33, with a mean of 25.7 years (SD=4.1). Of the participants, 7 reported having lived outside of Hungary for more than 6 months (these countries being Austria, Germany, Israel, Italy, Turkey, UK and USA), but none of them stayed in these countries for more than a year. They all lived in Budapest or within its metropolitan area.

All of the subjects reported knowledge of English on a more than novice level. The most frequent third language was German with 15 participants reporting some knowledge; followed by French with 7, Italian with 3 and Spanish with 2 speakers. Knowledge of Hebrew, Japanese, Polish, Russian and Turkish was reported by one speaker each. Familiarity with Polish, Russian and Turkish might help a Hungarian speaker discriminate between a full [i] and a retracted [i̠] as the phonemic inventories of these languages do include a retracted /ɨ/ or /ɯ/ phoneme. Speakers reporting some control of these languages will be analyzed further to see whether the knowledge of these languages did influence them.

vowel	stimuli
[i]	*tsit, ʧis, fiʃ, kir, lid, mit, niɟ, ʃiɟ, zip, zit*
[iː]	*biːp, fiːɲ, giːb, ɟiːb, hiːɲ, miːn, miːʃ, niːt, piːt, ziːm*
[ɛ]	*bɛz, tsɛg, dɛm, hɛd, hɛs, gɛl, lɛr, pɛv, ʃɛk, zɛm*

Table 2.10: Stimuli of the perception experiment

2.5.1.2 Stimuli

The stimuli of the experiment were monosyllabic nonce words made up of phonemes in the Hungarian phonemic inventory in a phonotactically legal way. All of these words had a CVC structure, where 10 stimulus words had a short [i], 10 had a long [iː], and there were 10 fillers with a short [ɛ] (as well as 10 fillers with [y(ː)], which were used only in training). The stimulus words were randomized using a Python script that automatically discarded C_1VC_1 shapes and words that exist in the Hungarian Webcorpus which contains 1.48 billion tokens (Kornai et al. 2006). The resulting set of words can be seen in Table 2.10.

The test words were recorded in a soundproof booth by the author, a male native Hungarian speaker, with a neutral tone and speech rate. To test how well the given difference was perceivable for Hungarian speakers, the vowels of the test words as recorded were taken as full vowels, and they were each manipulated using Praat to lower the F2 of the vowel with 200 or 250 Hz to create the stimuli with retracted vowels. The same procedure was used on the full vowel condition as well (lowering F2 by 0) so there was artificial manipulation on stimuli of both conditions. Lowering the F2 of the stimuli created words with retracted [i̠], [i̠ː] and a [ɜ] quality between [ə] and [ɛ] for the fillers (see Figure 2.5). Szeredi (2012) has shown that the [ɜ] vowel is not perceived by Hungarian speakers to be a realization of the /ɛ/ phoneme. This vowel quality seems to be on the border of the /ɛ/ and /ɒ/ categories:

speakers were as likely to accept it in place of an [ɛ] than in place of an [ɒ]. Therefore positive discrimination is expected for the fillers both in the 200 and 250 Hz conditions.

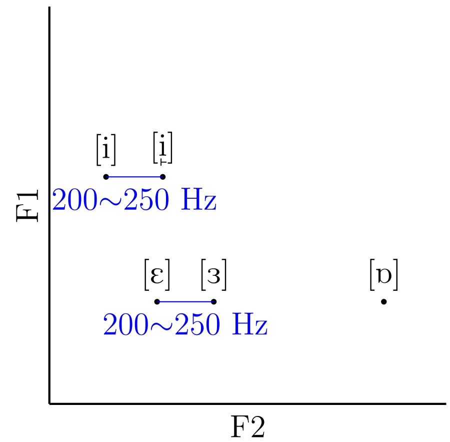


Figure 2.5: Design of the perception experiment

2.5.1.3 Procedure

The ability to discriminate between the two vowels in the stimuli was tested with an ABX design (Macmillan and Creelman 1991) – three sound files are played, where the first and the second one are the same and the third one is either the same as the first one (ABA) or the second one (ABB). The task of the participant is to decide which one it is the same as. For example, the stimulus pair [*biːp*] ∼ [*bi̞ːp*] was present four times in the experiment (ABA, ABB, BAB, BAA), as seen on table 2.11.

A	B	X	correct response
biːp	*bi̞ːp*	*bi̞ːp*	second one
biːp	*bi̞ːp*	*biːp*	first one
bi̞ːp	*biːp*	*bi̞ːp*	first one
bi̞ːp	*biːp*	*biːp*	second one

Table 2.11: Design of the ABX experiment for the example stimulus *biːp* ∼ *bi̞ːp*

The subjects were seated in a sound proof booth in Budapest, and they were all using the same laptop and headphones (Beyerdinamic DT 770 M closed dynamic headphones) for the presentation, which was run using the online JavaScript based experiment conducting framework Experigen (Becker and Levine 2010), hosted on Amazon AWS S3. Since the presentation was online, it was made sure in the extension code to Experigen which was responsible for handling ABX trials that all sounds be precached and loaded to avoid lags, and all efforts were made to ensure a consistent 500 ms ISI. Whenever there were some problems with the network and the playback was lagging, the subjects were instructed to listen to the trial again – while they were told not to replay the stimuli if there were no technical issues.

Before the presentation of the trial stimuli the participants were presented with 5 training stimuli to familiarize them with the task. These training stimuli did not contain [i(ː)] vowels to avoid unnecessary training on perception: subjects heard 3 trials with [ɛ]–[ɜ] and 2 harder ones with [y]–[y̘], the latter of which were only used in training.

The trials of the stimuli (with [i]–[i̘], [iː]–[i̘ː] and [ɛ]–[ɜ]) were randomized automatically within the Experigen script. All stimulus pairs were presented four times, so that the full vowel was the first in two trials and the second in the other two, and both ABA and ABB patterns were tested with both first vowels, as seen on Table 2.11. Finally, an anonymous demographic survey was filled out to get the participant's age, gender, check for other spoken languages and countries outside of Hungary where they lived for 6 months or more; as well as a question to filter out people with hearing disabilities.

2.5.1.4 Analysis

The analysis of the results was fairly straightforward. Each trial was first coded to show whether the participant had a correct identification or not. These results were then aggregated by subject and by vowel to find the percentages a given subject was correct by a given vowel. This was plotted for visual analysis. To determine whether a given subject was significantly better than chance in discriminating between the pairs, d' scores were calculated, as a confidence interval can be calculated around this measurement (Macmillan and Creelman 1991, Boley and Lester 2009). A d' score illustrates how well a participant is able to discriminate between two categories (Macmillan and Creelman 1991), and it and its standard deviance is calculated using the z-transformation of the hit and false alarm rates (the differencing model was used for this calculation). The following formulae define the calculation of d' and its standard error $SE(d')$ for ABX studies[1]:

(2.1) $p(c)_{\text{ABX,diff}} = \Phi\left(\frac{z(H)-z(F)}{2}\right)$

(2.2) solve the following equation for d':

$$p(c)_{\text{ABX,diff}} - \Phi\left(\frac{d'}{\sqrt{2}}\right)\Phi\left(\frac{d'}{\sqrt{6}}\right) - \Phi\left(\frac{-d'}{\sqrt{2}}\right)\Phi\left(\frac{-d'}{\sqrt{6}}\right) = 0$$

(2.3) $SE(d') = \sqrt{\left(\frac{H(1-H)}{\left(\frac{N}{2}\right)z^2(H)}\right) + \left(\frac{F(1-F)}{\left(\frac{N}{2}\right)z^2(F)}\right)}$, where N is the total number of trials (Boley and Lester 2009)

The hit rate corresponds to the proportion of correct responses to ABB and BAA stimuli, where the decision is made easier by the neighboring same stimuli. The false rate corresponds to the proportion of *incorrect* responses to ABA and BAB stimuli, where the crucial stimulus is next to a different sound, yet the participant still presses the button which would indicate that the last two sounds are the same.

If the false alarm rate is higher than the hit rate, the calculation of the d' for the ABX is not possible (since the equation in (2.2) to estimate d' cannot be solved for a root), therefore d' scores of subjects with $F > H$ will not be shown. Similarly, the variability of the d' cannot

[1] $p(c)_{\text{ABX,diff}}$ denotes the proportion of correct responses, Φ is the is the normal distribution likelihood function (`pnorm` in R), z is the z-transformation (`qnorm` in R), H is the hit rate and F is the false alarm rate.

be calculated if H=0.5 or F=0.5, since $z(0.5) = 0$, resulting in division with zero in (3). In these cases, error bars will not be shown in the following barplots. It is clear, though, that in these situations, the plotted contrast is not perceivable for the subject.

Since there were a few occasions where a subject was perfect (or perfectly bad) on a contrast, the by-subject by-contrast hit rates and false rates could be 0% or 100%. The z-transform has problems with these values, so 0 rates have been replaced with $\frac{0.5}{n}$ and 1 rates with $\frac{n-0.5}{n}$, following Stanislaw and Todorov (1999). It should be kept in mind, though, that the resulting scores are not entirely accurate because the small sample size exaggerates the effect of this manipulation (Miller 1996).

Conditions were analyzed by pooling the by-subject responses (whether they were correct or not) and building a logistic generalized linear mixed effects model to see if differences were perceived better than chance.

2.5.2 Results

The main results indicate that the [i]–[i̞] contrast is not well perceivable in the 200 Hz condition. Subjects in the 250 Hz condition fared better, but still did not reach the 80% benchmark for the perception of this contrast. The control stimuli with [ɛ]–[ɜ] were much easier to differentiate in both conditions. The results also suggest that long [iː]–[i̞ː] pairs were *harder* to discriminate compared to short pairs.

2.5.2.1 200 Hz condition

Figure 2.6 illustrates the percentages correct in the task for all 3 vowel conditions ([ɛ]–[ɜ], [i]–[i̞] and [iː]–[i̞ː]. It can be seen, that the difference in the control stimuli was much easier to perceive (ranging 75%–97.5%, on average 85.9%) than the high front vowels, and that long

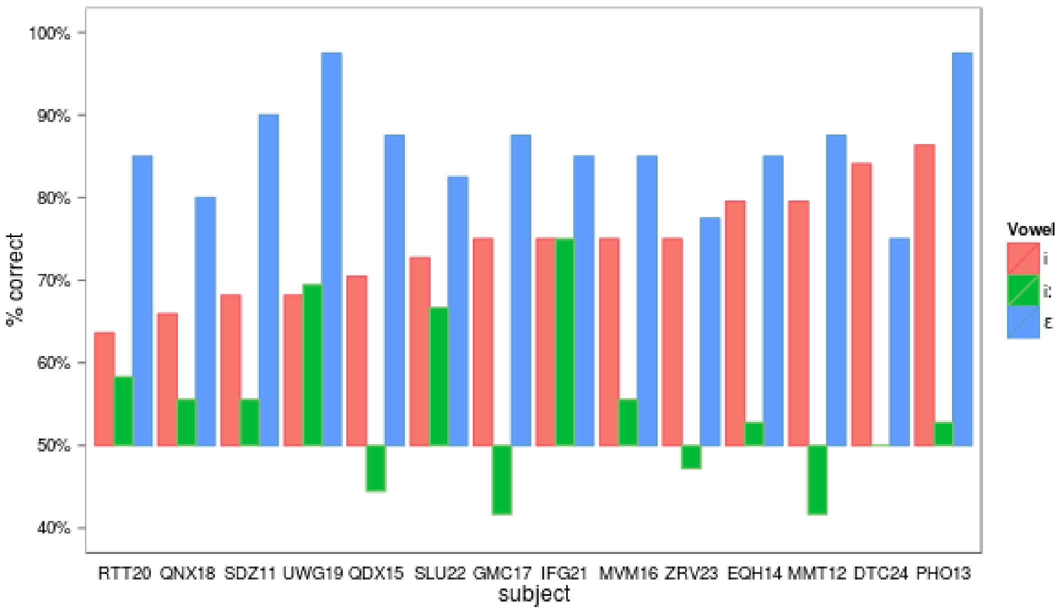


Figure 2.6: Perceptibility of 200 Hz contrasts for all vowel types, ordered by accuracy on the [i]–[i̠] contrast

[iː] vowels (41.6%–75%, on average 54.8%) were much harder to discriminate than short [i] (ranging 63.6%–86.3%, on average 74.2%).

There is no correlation between the performances on different vowels ([ɛ] vs. [i] $r = -0.01$, [i] vs. [iː] $r = -0.31$, [ɛ] vs. [iː] $r = 0.17$).

Figure 2.7 shows the d' scores of the participants in the 200 Hz condition, with standard errors represented by the error bars (except when variability is not calculable, cf. 2.5.1.4). It can be confirmed based on this figure, that the perceptibility of the 200 Hz difference in F2 is low for high front vowels, especially for the long [iː], where only 3-4 participants out of 14 could be argued to be able to discriminate between the full and the retracted vowel.

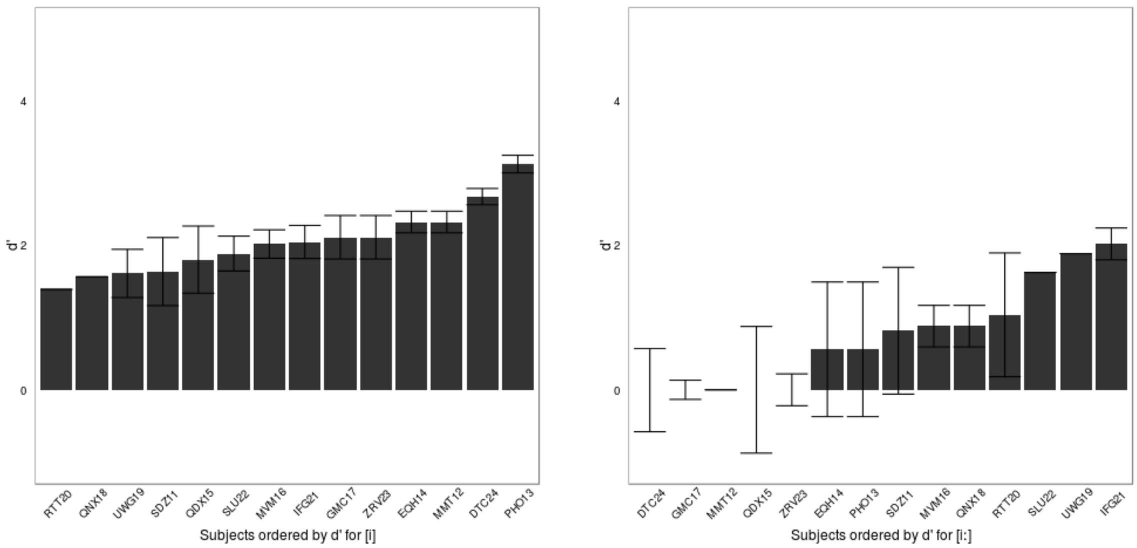


Figure 2.7: Perceptibility of [i]–[i̙] (left) and [iː]–[i̙ː] (right) contrasts as d' scores in the 200 Hz condition

2.5.2.2 250 Hz condition

The correct percentages for the 250 Hz condition are shown on Figure 2.8. Again, the control stimuli (82.5%–100%, average 95.7%) seem to be the most perceivable, as most of the subjects managed to discriminate between the full and the retracted stimulus easily. The distinction, however, was this time almost as easy for short [i] vowels (81.8%–100%, average 86.8%) as for [ɛ], while distinguishing between a full and a retracted [iː] still proved to be hard (38.4%–88.8%, average 69%).

The correlation between performance on short and long [i] was greater this time ($r = 0.42$), although still not significant ($p = 0.13$). There was no correlation between performance on the control stimuli and the test stimuli ([ɛ] vs. [i] $r = 0.11$, [ɛ] vs. [iː] $r = 0.23$).

It is also visible on Figure 2.9 with the d' scores, that the 250 Hz difference is now generally

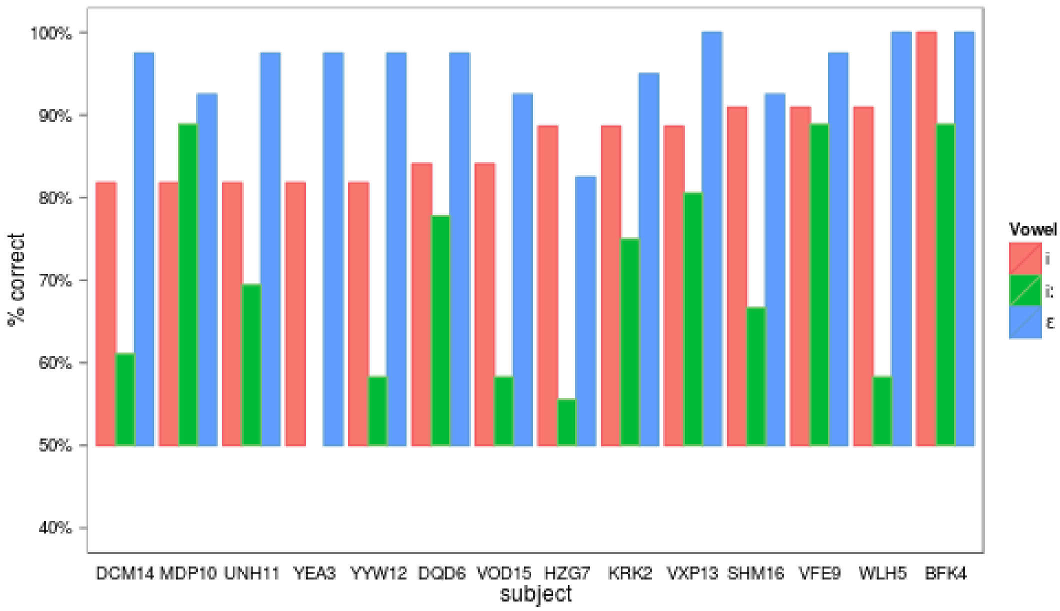


Figure 2.8: Perceptibility of 250 Hz contrasts for all vowel types, ordered by accuracy on the [i]–[i̤] contrast

perceivable for high front vowels, although the long [iː] still presents some problems for at least 5 subjects out of 14.

2.5.2.3 Regression model

A logistic mixed effects model was built with vowel, condition and their interaction as fixed effects; subject and item as random effects (with random intercepts and a random slope by vowel for subject, and a random slope by condition for item). The interaction of vowel and condition could be removed from the model without a significant decrease of log likelihood ($\chi^2(2) = 4.54$, $p = 0.1$).

The coefficients and the calculated odds ratios of the final model can be seen in Table

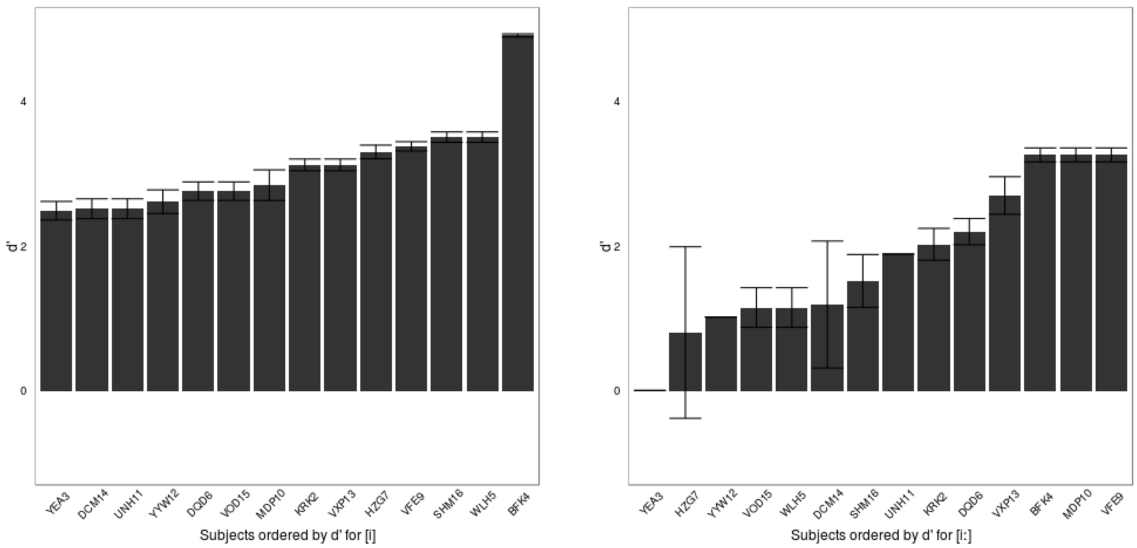


Figure 2.9: Perceptibility of [i]–[i̞] (left) and [iː]–[i̞ː] (right) contrasts as d' scores in the 250 Hz condition

2.12. It can be seen on the results that both condition and vowel quality influence the probability of a successful identification significantly. Specifically, if the difference is 250 Hz, the odds for success is 2.5 times higher than in the 200 Hz condition. Perception of the tested vowel is significantly harder than the control vowel, although the long [iː] seems to fare much worse than the short [i] again.

2.5.3 Discussion

Based on the results of the perception study, it can be claimed that the perception threshold of an F2 difference in a high front vowel for Hungarians is certain to be above 200 Hz. In the 200 Hz condition perception of this difference was unstable, especially for a long vowel. Discrimination in the 250 Hz condition seemed to be much easier based on the results, with several subjects reaching an 80% benchmark.

factor	e^{β} (odds ratio)	β	SE	z	p	
Intercept	7.67	2.04	0.22	9.29	< 0.001	***
`condition`=250 Hz	2.57	0.94	0.14	6.97	< 0.001	***
`vowel`=[i]	0.42 (1/2.38)	-0.87	0.27	-3.17	0.002	**
`vowel`=[iː]	0.14 (1/7)	-1.95	0.29	-6.74	< 0.001	***

Table 2.12: Model for the results of the perception experiment (default value for the `vowel` factor was [ɛ])

These results show that a difference of ≈ 15 Hz based on the size of the difference found in the production study is not perceivable for Hungarian speakers. For a vowel within a word, the perceptibility threshold seems to be around ≈ 250 Hz, which is much higher than the JND found isolated vowel sounds in other studies (with speakers of different languages). The conditions presented in this study are closer to real language use, as the vowels were heard in (nonce) words, between consonants.

The fact that even a 250 Hz difference was not perfectly perceivable yet implies that this difference cannot be contrastive. Suggesting that the vowel harmony pattern relies strongly on a phonetic difference that is not perceivable and cannot be contrastive is problematic. if speakers used this difference in their representation of vowel harmony, they should be more sensitive to these differences – and this is not the case based on this experiment.

A surprising result of the experiment was that a difference in a long vowel was much harder to discriminate than the same difference in a short vowel. The reasons for this are not yet clear – it might be attributed to the artificiality of the task or to memory issues, the time distance between the first (A) and the last (X) sound presentation being longer for long vowels.

2.6 Rating study

The question addressed in this study is whether Hungarian speakers are able to use a phonetically perceivable difference to categorize stems as harmonic or antiharmonic.

Since the production and perception studies have indicated that a strictly phonetic explanation for antiharmonicity is not plausible, a less strict definition of phonetic grounding has to be tested. This explanation would claim that despite the fact that antiharmonicity is not exclusively driven by phonetic differences, phonetic information can play a role in how exceptional patterns in vowel harmony like antiharmonicity are learned. The goal for this study is, therefore, to see if phonetic information that is perceivable to Hungarian speakers can be used by them to categorize a stem as harmonic or antiharmonic.

To test the phonetics first hypothesis, a nonce word rating study was conducted. In this study, Hungarian speakers rated suffixed nonce roots on how natural the word form sounded to them. Stem vowels were manipulated to two qualities: a retracted [i̠] and a full [i] vowel condition. All stems were suffixed once with a back suffix and once with a front suffix. The following table shows the full cross of the stem and suffix conditions – all four are in the stimulus set:

form	stem	suffix	fit phonetic bias?
ziptɛk	full	front	✓
ziptɒk	full	back	×
zi̠ptɛk	retracted	front	×
zi̠ptɒk	retracted	back	✓

Table 2.13: Example stimulus [zip] in all four conditions.

If phonetic information plays an important role in antiharmonicity besides lexical information, speakers should rate forms that are predicted by such a phonetic bias better (full:front and retracted:back) than forms that have a mismatch based on these predictions (full:back and retracted:front). The interaction of the stem:suffix vowel conditions is the term that shows the presence of such a phonetic bias: if a `stem=retracted:suffix=back` factor is significantly positive, that means that speakers are aware that a more retracted stem is more likely to be antiharmonic, therefore more likely to be suffixed with a back vowel. The possible interpretations of the results is shown on Table 2.14.

stem:suffix interaction	vowel quality used for classification	phonetic information has role in learning antiharmonicity
significant in the expected direction	✓	✓
not significant in the expected direction	?	more data needed
0	×	×
significant in the unexpected direction	✓	×

Table 2.14: Evaluation of a phonetic hypothesis based on the results of the rating study

2.6.1 Methods

2.6.1.1 Participants

The participants in this study were the same as in the perception experiment in 2.5. The two experiments were conducted on the same day in the same soundproof room. The rating experiment was conducted *before* the perception experiment, so that the training that might have occured during the ABX trials could not affect the subjects' judgments in the rating study.

In total, 28 subjects took part in the experiment, 16 female and 12 male, with all speakers living in Budapest, Hungary. No matter what experimental condition group the participants were in, the phonetic difference used was 250 Hz for all participants.

2.6.1.2 Stimuli

The stimuli for this study were monosyllabic nonce Hungarian words suffixed with either a back or a front suffix, as shown in Table 2.15. All of the test words had a CVC shape, where the V could be [i] (full vowel), [i̠] (retracted vowel), [iː], [i̠ː], and the fillers [u] (back filler) and [y] (front filler). The nonce stems were the same as the ones found in the perception experiment: 10 stems with long and 10 stems with short [i] vowels, and 10-10 stems in the same environments with [i̠]. The fillers were 10 stems with [u] and 10 stems with [y] – 5-5 short and long vowel stems for each. The stimuli were also controlled for the edit distance to the closest antiharmonic existing stem: 10 differed only in one consonant from a back i-stem, and 10 did not have an antiharmonic stem within 1 edit distance. No wug verb differed from an existing stem only in vowel length and edit distance is calculated ignoring vowel length.

All of the nonce words were suffixed with the past tense 3rd person plural suffix, which is *[-tɒk]* after back stems and *[-tɛk]* after front stems. Each stimulus word occurred once with the back suffix and once with the front suffix, so that the suffix and stem vowel conditions were perfectly crossed. Pairing the 60 wug stems with the 2 stem vowel conditions resulted in 120 stimuli in total.

The target words were embedded in a natural sentence with the following pattern (with syllable counts in parentheses):

(199) DET(1) N.Pl(3) **V.Past.3pl(2)** DET(1) N.Adv(3)

The following example illustrates the pattern:

vowel	N	stimuli
[i]	10	*ʦit, ʧis, fiʃ, kir, lid, mit, niɟ, ʃiɟ, zip, zit*
[i̘]	10	*ʦi̘t, ʧi̘s, fi̘ʃ, ki̘r, li̘d, mi̘t, ni̘ɟ, ʃi̘ɟ, zi̘p, zi̘t*
[iː]	10	*biːp, fiːɲ, giːb, ɟiːb, hiːɲ, miːn, miːʃ, niːt, piːt, ziːm*
[i̘ː]	10	*bi̘ːp, fi̘ːɲ, gi̘ːb, ɟi̘ːb, hi̘ːɲ, mi̘ːn, mi̘ːʃ, ni̘ːt, pi̘ːt, zi̘ːm*
all target	40	
[u]	5	*fug, lut, dum, kun, bup*
[uː]	5	*guːl, duːb, puːg, luːs, muːj*
[y]	5	*kyl, dyn, tyk, gyd, ʃyr*
[yː]	5	*zyːl, gyːp, jyːg, myːr, lyːd*
all filler	20	
TOTAL	60	

Table 2.15: Stimuli of the rating experiment

	A	tigrisek	fínytak	a	ketrecben.
(200)	[ɒ	*tigriʃɛk*	*fiːɲtɒk*	*ɒ*	*kɛtrɛʤbɛn*]
	the	tigers	*fiːɲ*+back suffix	the	cage.INESS

‘The tigers were *fíny*ing in the cages.’

The test sentences were read by the author in a soundproof booth with a neutral intonation. An acoustic analysis of the target vowels can be seen in Table 2.16. It can be seen that an effect of 41.2 Hz was seen before the different suffix vowel conditions, but this difference might be due to coarticulation, and it is not significant in this sample. The long vowels were significantly longer (by 34.4 ms) in duration than the short vowels. There is also a qualitative difference between the long and short [i] vowels: the long vowel is more peripheral, so its F2 is higher than the F2 of the short [i]. This be explained by vowel undershoot: there is not enough time for the short vowel to reach the most peripheral target. The findings of the production study in Section 2.4.3 are in line of these differences.

	F2 (Hz)	SD	t(39)	p
before front suffix	2170.8	135.2		
before back suffix	2129.6	115.1		
diff.	41.2		1.05	>0.05
short [i]	2072	89.6		
long [iː]	2231.2	105		
diff.	159.2		5.23	<0.001***
neighbor	2161.6	151		
non-neighbor	2138.3	97.7		
diff.	23.2		0.59	>0.05
	duration (ms)	SD	t(39)	p
short [i]	73.3	17.8		
long [iː]	107.7	14.7		
diff.	34.4		6.71	<0.001***

Table 2.16: Acoustic properties of the unmanipulated stimuli

The actual stimuli were artificially created from these utterances. The manipulation of formant values was made in Praat (Boersma and Weenink 2012). The whole utterance was resampled at 10 000 Hz, and the stem and suffix vowels were segmented from the word. The formants were filtered out based on the source-filter model of speech: the filter was created by an LPC analysis looking for one peak for every 1000 Hz. The source was filtered out and the formant analysis of the original source was altered using the target formant values. Finally the new formants were refiltered with the source, resulting in the final stimuli. The target formant values of the non-retracted vowels were the mean values from the acoustic data of the readings, and are presented in Table 2.17. The target F2 for the retracted [i̠] vowel was 250 Hz lower than the F2 of the full vowel.

The resulting sound files sounded a little noisy and artificial, as noted by several subjects, because the formant altering methodology and the downsampling of the input did introduce

vowel	F1	F2	F3
	stem vowels		
i	310	2150	2683
i̞	310	2050	2683
u	359	879	2281
y	338	1696	2142
	suffix vowels		
ɛ	578	1639	2358
ɒ	602	1406	2262

Table 2.17: Target formants for the formant manipulation process

some noise. This procedure was needed, however, as it was crucial for this experiment that the formant values for the stem and suffix vowels are controlled for. The importance of the effect of artificiality was lowered by the fact that the level of the introduced noise was the same for all stimulus items.

2.6.1.3 Procedure

The subjects were seated in a quiet room in Budapest, before doing the perception experiment. They listened to the stimuli using Beyerdinamic DT 770 M headphones. The experiment was presented using Experigen, and was hosted on Amazon AWS S3. The instructions for the rating experiment were shown on the screen and were reiterated by the experimenter to ensure that the participant understood the task well. The subjects were instructed to rate the sentences based on how natural they think the sentence sounds as a Hungarian sentence. They were encouraged to use the whole 1–7 scale, and there was a short training phase before the main task, where the subjects could familiarize themselves with the study and ask further questions about the task.

stem vowel	suffix vowel	stimuli seen	all stimuli
[i]	back	5	10
[i̞]	back	5	10
[i]	front	5	10
[i̞]	front	5	10
[iː]	back	5	10
[i̞ː]	back	5	10
[iː]	front	5	10
[i̞ː]	front	5	10
[u(ː)]	back	5	10
[u(ː)]	front	5	10
[y(ː)]	back	5	10
[y(ː)]	front	5	10

Table 2.18: Number of conditions seen by a subject in the experiment

The presentation was made again online with Experigen (Becker and Levine 2010). Because of the high number of stimuli, only a randomly selected half of them was presented to each subject (see Table 2.18), so instead of 120 judgments, only 60 were made. The length of the experiment was approximately 15 minutes.

The screen the subjects saw for each stimulus item is seen on Figure 2.10. The stimulus sentence was presented with the prompt *Íme a következő mondat:* ‘This is the next sentence:’, with the target word in boldface so that the subjects can pay more attention to it instead of the whole carrier sentence. To ensure this even more, the target word was repeated (*Erre a szóra figyelj:* ‘Concentrate on this word:’). A play button was placed below, which could be pushed several times, if the subjects were undecided after listening to the manipulated sentence once. After having played the test sentence at least once, 7 buttons for rating appeared below the prompt asking *Mennyire hangzott természetesen a kulcsszó?* ‘How natural did the key word sound?’, with labels next to 1 (*nagyon nem* ‘very much not’) and 7 (*teljesen* ‘totally’).

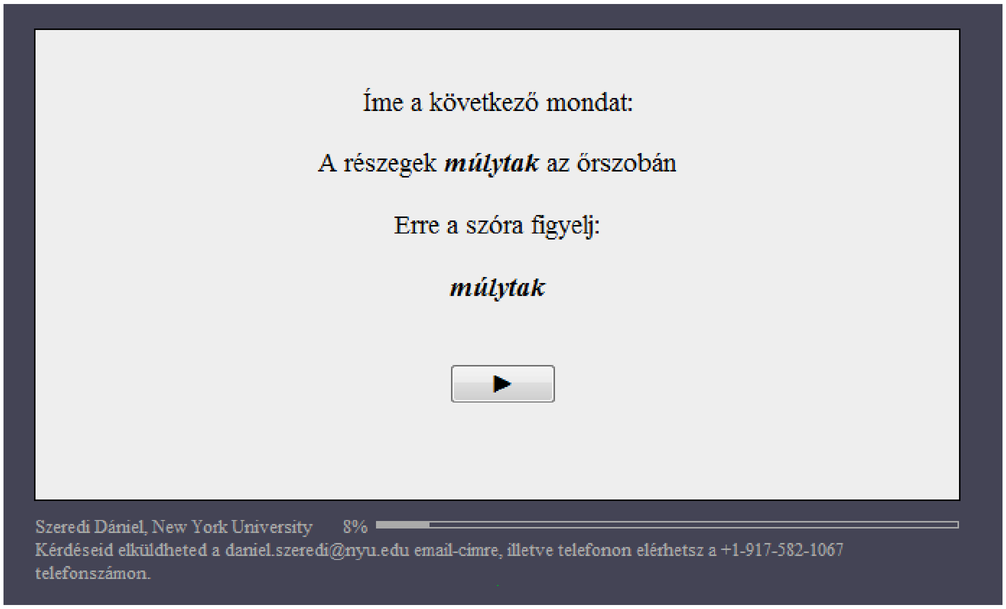


Figure 2.10: The screen shown to participants for each stimulus.

2.6.1.4 Analysis

The results were analyzed using linear mixed effects models on the numeric response as the dependent variable. The independent variables were the stem vowel and suffix vowel conditions, with their interaction being the key variable to the results. If speakers do use phonetic retraction in determining the harmonicity of the nonce verb, the stem:suffix interaction is expected to be significant: stems with retracted [i̠(ː)] stem vowels will tolerate back suffixes more than stems with full [i(ː)] stem vowels. Expectations of the effect of the non-interaction factors of stem vowel and suffix vowel are not relevant to the hypotheses.

The fillers are expected to show the same pattern in a more emphasized way: stems with [u] are expected to tolerate only back suffixes, while [y] stems are expected to be accepted only with front suffixes.

stem	suffix	total		i		iː	
		mean	SD	mean	SD	mean	SD
full	back	4.1	1.51	4.16	1.46	4.04	1.57
full	front	3.93	1.63	3.91	1.65	3.96	1.6
retracted	back	3.87	1.62	3.97	1.54	3.76	1.7
retracted	front	3.91	1.57	3.76	1.57	4.07	1.57

Table 2.19: Average acceptability scores for the target vowels

2.6.2 Results

2.6.2.1 Target words

The target words (with [i(ː)] or [i̙(ː)] do not show any effect of the stem vowel, the suffix vowel or their interaction at all. The mean acceptability score was 3.96 (with a standard deviation of 1.58), broken down by conditions in Table 2.19. The retracted condition shows lower averages than the full condition. Front suffix vowels had lower scores in the full stem vowel condition and higher in the retracted stem vowel condition, especially with a long vowel, which implies that even the sign of interaction is not in the expected direction.

The distribution of the scores were slightly bimodal (cf. Figure 2.11), with less 4 scores than 3 or 5. This implies that there was there was a separation of judgments as acceptable or less acceptable by the speakers – however, no subset of the data based on the available factors fails to show this bimodality (for the stem and suffix vowel factors, see Figure 2.12), meaning that the basis of this separation is not clear.

The significance for these differences was tested using a linear mixed effects model with item with a random intercept and subject with random intercept and all random slopes

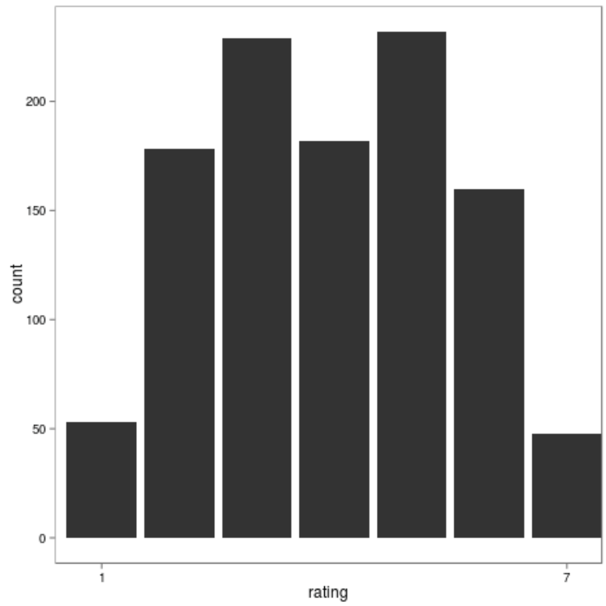


Figure 2.11: Histogram of the acceptability scores for all target vowels

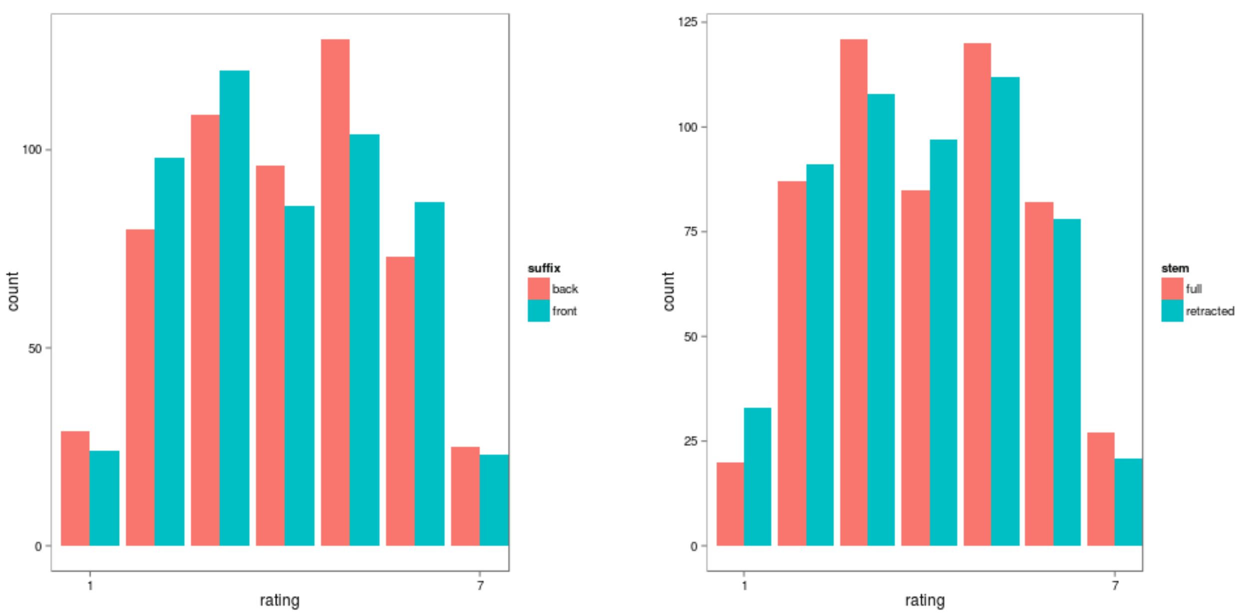


Figure 2.12: Histogram of the acceptability scores for different subsets of the data: suffix vowel (left) and stem vowel (right)

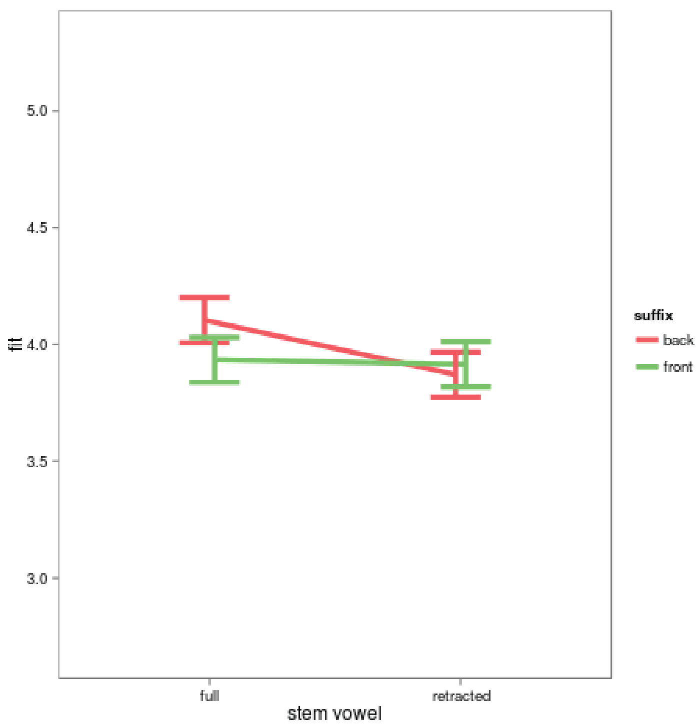
factor	β	SE	t	χ^2 (df)	p
Intercept	3.8	0.22	17.013		
`stem vowel = retracted`	-0.02	0.13	-0.12	*1.65 (1)*	*0.2*
`suffix vowel = back`	0.18	0.13	1.35	*0.68 (1)*	*0.41*
`length = long`	-0.01	0.23	-0.02	0 (1)	1
`neighbor:` 1 edit dist.	0.29	0.19	1.48	2.26 (1)	0.13
`stem = retracted : suffix = back`	-0.21	0.19	-1.11	1.22 (1)	0.27

Table 2.20: Coefficients of the saturated model on the response variable in the rating study. Because of the term elimination method, the χ^2 and p columns for `stem` and `suffix` variables refer to their elimination after the removal of their interaction, highlighted by italic formatting.

as random factors. The factors included in the model were the stem vowel, suffix vowel conditions, the length of the stem vowel ([i] or [iː]) and a binary variable on whether there is an antiharmonic root 1 edit distance away from the stem. Only the crucial stem vowel:suffix vowel interaction could be added to the model as more complex models failed to converge.

The coefficients of the saturated model are shown in Table 2.20. No factors turned out to be significant either by looking at t values of the coefficients or by eliminating terms and using likelihood ratio tests. The key interaction is not significant either, moreover, its sign points to the unexpected direction: stems with a full [iː] vowels actually seem to prefer back suffixes more than stems with a retracted [i̙ː]. The interaction is plotted on Figure 2.13 – it can be seen that back suffixes vowels are more preferred with full vowels. The confidence intervals are based on the standard error of a plain linear regression model on the stem:suffix interaction, as prediction from the mixed effects model is not straightforward.

Figure 2.13: Interaction of the `stem` and `suffix` vowel conditions with confidence intervals



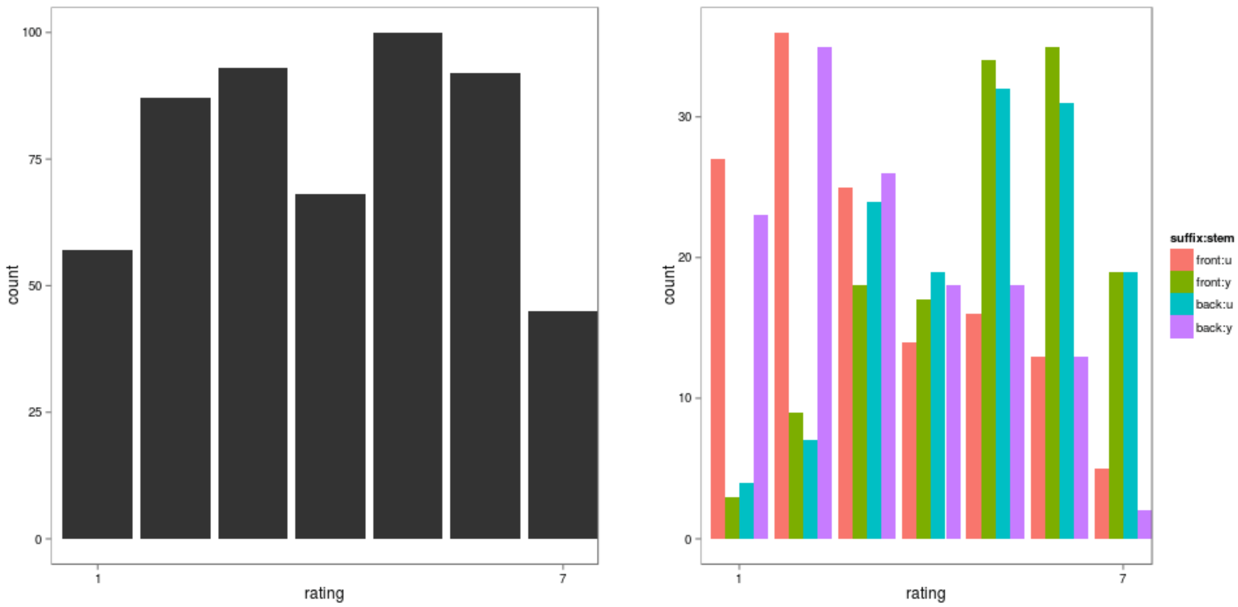


Figure 2.14: Histogram of the ratings of all fillers collapsed (left), and separated by stem and suffix (right)

2.6.2.2 Fillers

The fillers were used in this experiment to see if the intended interaction shows up when the contrast between a front and a back stem vowel is perfect, as both of them ([u(ː)] and [y(ː)]) are phonemic in the language. The distribution of ratings is bimodal again, but it is clear, that the stem:suffix interaction can explain the bimodality away: unharmonic combinations are rated low, harmonic ones are rated high (see Figure 2.14).

The linear mixed effects model on the filler data now shows that several factors are significant. The coefficients are seen on Table 2.21. The results indicate that while length did not play a significant role in the ratings, the sign of the interaction is in the expected direction and its magnitude is significant. This means that the participants did prefer back suffixes much more after [u] than after [y] – the excepted behavior based on the knowledge

factor	β	SE	t	χ^2 (df)	p	
Intercept	4.92	0.27	18.37			
`stem = [u]`	-1.7	0.38	-4.48			
`suffix = back`	-1.58	0.37	-4.26			
`length = long`	-0.21	0.26	-0.79	0.6 (1)	0.44	
`stem = [u] : suffix = back`	3.21	0.61	5.29	23.34 (1)	<0.001	***

Table 2.21: Coefficients of the saturated model on the response variable for fillers. Likelihood ratio could not be calculated for the `stem` and `suffix` factors, as their interaction was significant.

of vowel harmony. The interaction is plotted again on Figure 2.15: the effect is much higher than with the i-stems, and the direction of the interaction is in the expected direction.

2.6.3 Discussion

The results of the rating study indicate that contrary to the expectation based on a framework with phonetic bias, Hungarian speakers do not seem to employ subphemic phonetic differences to classify stems as harmonic or antiharmonic. The perception study indicated that a 250 Hz difference in the F2 of a high front vowel is distinguishable for Hungarian speakers. The 250 Hz difference employed in this experiment was therefore perceivable by the speakers – yet there was no indication in the results that Hungarian speakers use this difference to categorize the nonce stems.

The effect size of the interaction between stem vowel and suffix vowel qualities was negligible for [i(ː)]∼[i̞(ː)] stems, it actually pointed to the unexpected direction and it was not significant. Looking back to Table 2.14, the null result indicates that even the weak version of the phonetic explanation for antiharmonic behavior cannot be supported: phonetic information seems not to have a role in acquiring antiharmonicity.

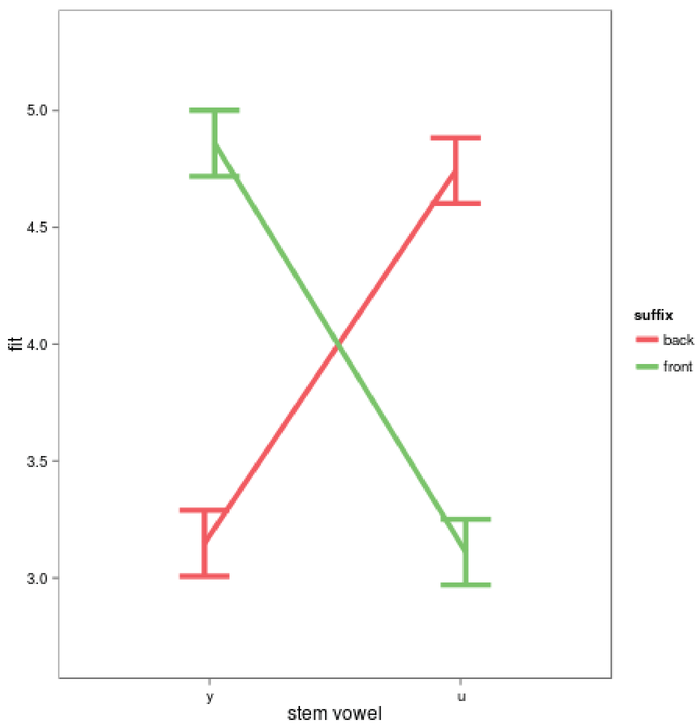


Figure 2.15: Interaction of the `stem` and `suffix` vowel conditions with confidence intervals for the fillers

The fillers in the data did show the expected interaction, illustrating that the methodology and the analysis of the experiment could not be to blamed for the lack of such a result. Therefore, task effects are unlikely to be responsible for the null finding.

The question of how much possible phonetic differences extend to novel stems was investigated by Benkő and van de Vijver (2016) as well and their results are compatible with the results of this study. They ran their acoustic analysis on Hungarian antiharmonicity not only using real harmonic-antiharmonic stem pairs, but using nonce stems as well. Their preliminary results show that nonce stems show the acoustic differences found for real stems was weaker in nonce stems, suggesting that these differences might be secondary as they are not extended to novel input.

2.7 Summary

The results of the three experiments on antiharmonicity in Hungarian have therefore hint that the role of phonetic detail in vowel harmony is smaller than what recent phonetically based analyses would suggest. While naturalness of the pattern could still play a role in the evolution of vowel harmony, exceptional subpatterns can emerge and be sustained with no strong phonetic grounding. The synchronic representation of antiharmonicity vowel harmony seems to rely exclusively on lexically marked subcategories of exceptional stems.

The possibility of phonetically grounded explanations is not deemed implausible by these results. It is very probable that phonetically natural processes do influence what kinds of grammatical patterns can emerge. However, based on the three experiments discussed above, there must be a more abstract intermediate representation between phonetics and speakers' behavior. The role of phonetic properties of the sounds involved in grammatical processes is to introduce a bias towards the plausibility and learnability of these patterns but it is not the only bias (there is structural and possibly other biases as well, Moreton 2008), and grammar is able to override the effects of a phonetic bias. This grammatical effect is responsible for why phonetic differences on higher level conditions in the rating experiment fail to influence speakers. Vowel harmony is grammaticalized in Hungarian, therefore phonetic differences can no longer influence the grammatical behavior on a synchronic level.

Chapter 3

Mixed stems in Finnish

Finnish vowel harmony, is similar to Hungarian vowel harmony, as it treats front non-low unrounded vowels [e,i] as neutral. Contrary to Hungarian, antiharmonicity — that is, stems exceptionally selecting unexpected allomorphs of subsequent affixes — is extremely rare, confined to only 2 forms in Finnish, as the prototypical stems in Finnish have two syllables, unlike prototypical Hungarian monosyllabic stems. The question in Finnish that is similar to Hungarian antiharmonicity is the quality of the vowel in the *second* syllable after a neutral first vowel. As [e,i] are neutral, the backness of the vowel after such vowels is not predictable: it could be either back (like [a] in *mixed* stems) or front (like [æ] in *harmonic* stems). This exceptionality, therefore, is not about allomorph selection, but about static phonotactic generalizations in the lexicon.

The goal of this study is to determine whether Finnish speakers use sublexical phonotactic patterns that underlie the distribution of mixed stems in Finnish. The sublexical hypothesis for exceptionality — as discussed in the previous chapter — states that speakers are aware of phonotactic patterns that set apart regular and exceptional stems in their lexicon. This implies that speakers will categorize novel stems according to these patterns, as their goal is to assess which sublexicon a new stem is more likely to belong.

Unlike discussions of sublexical regularities in studies like Albright and Hayes (2003), Becker et al. (2011) and Gouskova et al. (2015) and the previous chapter, a sublexicon in this discussion is not defined as a class of stems determined by their morphophonological behavior, such as allomorph selection. Sublexica in the Finnish mixed stem case will be defined as subsets of the lexicon based on their phonotactic characteristics: whether they conform to strict vowel harmony, they are mixed back-neutral or they are completely disharmonic. The results will show that the assumptions on sublexica based on allomorph selection are going to be valid for this interpretation of sublexicality as well.

In order to start finding supporting evidence for the sublexical hypothesis, some patterns that are characteristic of the mixed stem sublexicon or to the corresponding harmonic sublexicon were identified using the UCLA Phonotactic Learner on a web corpus. These patterns were stated over some phonotactic property of the stems: for example, one such pattern is that intervocalic *-p-* is dispreferred in the harmonic sublexicon. The goal of finding these patterns was to then test if Finnish speakers use these sublexical regularities in deciding whether a novel form is more likely to be harmonic or mixed. For example, it would support a sublexical hypothesis if Finnish speakers rated a novel form like *mipa* better than *mipæ*, as compared to an average preference for harmonicity. To find out whether this is true, a wug study was conducted based on forms corresponding to the patterns found in the corpus study.

The results of the study indicate that Finnish speakers are indeed aware of sublexical regularities of harmonic and mixed sublexica. However, there is a gradient nature of this awareness: Finns are using evidence based on regularities about the quality of the vowel(s) found in the stem exclusively much more than patterns based on consonantal effects when categorizing novel stems. This is based on the fact that the effects of medial or initial consonants were found to be much weaker than the effect of the stem vowel in the wug study. These results parallel the findings of Hayes et al. (2009) for Hungarian: certain patterns seem to be easier to learn that other patterns. Hayes et al. call the patterns that are better learned natural constraints, as these constraints involve vowel-to-vowel dependencies, while they call constraints that were underlearned as they involved consonantal patterns unnatural. The findings of this chapter match Hayes et al. (2009) in these details as well: consonantal patterns are clearly weaker than vowel-to-vowel patterns, but still present in Finnish speakers' grammars.

Section 3.1 will describe the pattern of mixed and disharmonic stems in Finnish, and will discuss earlier literature on the phonotactics of Finnish mixed stems (Section 3.1.2). Section 3.2 introduces the main corpus study to explore regularities in the harmonic and mixed sublexica and Section 3.3 will present a nonce word study on whether Finnish speakers use these regularities to decide on the well-formedness of novel items. Finally, Section 3.4 will summarize the results and their implications.

3.1 The pattern

The Finnish vowel system consists of eight vowel phonemes, all of which can be short or long. Similarly to the Hungarian system, it makes sense to divide the vowels in the three broader categories of front rounded, front unrounded and back vowels based on their behavior in vowel harmony. The vowel system is presented below. As the Finnish orthography is very phonetic, the only difference between IPA and the orthography being the notation of vowel length as a double vowel, for example <aa> for [aː], which will be adopted in this dissertation, and the spelling of [ø] as <ö> and [æ] as <ä>, which will not be adopted. For reference, the Finnish consonant system with rare phonemes mostly found in loanwords in parentheses is presented in Table 3.2.

	front				back	
	unrounded		rounded			
high	i	ii	y	yy	u	uu
mid	e	ee	ø	øø	o	oo
low	æ	ææ			a	aa

Table 3.1: The Finnish vowel system

	labial	dental	palatal	velar	glottal
stop	p (b)	t d		k (g)	ʔ
fricative	(f)	s	(ʃ)		h
nasal	m	n		ŋ	
approximants	ʋ	l r	j		

Table 3.2: The consonants of Finnish (Suomi et al. 2008)

The Finnish vowel harmony pattern is quite similar to Hungarian. The high and mid front unrounded vowels are considered neutral, and show similar patterns as in Hungarian too. In the examples below, stems are suffixed with the inessive case marker, which takes the form [-ssa] after a back stem and [-ssæ] after a front stem. Examples (201) and (202) illustrate the simple cases of the allomorphy:

(201) *back stem:* talo 'house' ∼ talossa, *talossæ

(202) *front stem:* møly 'noise' ∼ mølyssæ, *mølyssa

Mixed stems behave a little bit differently than in Hungarian. Just like with Hungarian, if the last vowel is back, the suffix will always be back (203–204). The consonant alternations found in these examples traditionally called gradation or grade alternation (Suomi et al. 2008) like [kk]∼[k] here or [t]∼[d] and [ks]∼[kt]∼[hd] in later examples; as well as the rule [e]→[i] / _# are not relevant to the discussion here, as gradation and vowel harmony do not interact, and the vowel raising rule applies only word-finally and the harmonic behavior of the vowel does not change.

(203) *mixed back stem:* nekku ‘a kind of confection’ ∼ nekussa, *nekussæ

(204) *mixed back stem:* visa ‘quiz’ ∼ visassa, *visassæ

However, if the last vowel is front, the suffix can vary based on the exact identity of this front vowel and the previous front vowel(s), if there are any. Final /i/ and /e/ is always transparent, letting the backness of the preceding vowel through:

(205) *transparent /i/:* koti ‘home’ ∼ kodissa, *kodissæ

(206) *transparent /e/:* kaksi ‘two’ ∼ kahdessa, *kahdessæ

Truly mixed stems, which tend to be overwhelmingly loanwords in the language, where back vowels mix with front rounded vowels in a way, where the back vowel is followed by a front rounded vowel, the agreement suffix tends to be front, but there is variation found (207). There is also variation in mixed stems where there is a front rounded vowel or /æ/ followed by a transparent /i/ after a back vowel in the stem, a situation common in loanwords (208–210). Ringen and Heinämäki (1999) show that the rate of back-selection or front-selection for such mixed stem words varies based on several factors, such as the placement of the primary and secondary stress, and the height of the front vowel.

(207) *vacillating back+front rounded vowel:* Malmø ‘city in Sweden’ ∼ Malmøssæ or Malmøssa

(208) *vacillating loanword with back+/y/+/i/:* analyysi ‘analysis’ ∼ analyysissæ or analyysissa

(209) *vacillating loanword with back+/ø/+/i/:* amatøøri ‘amateur’ ∼ amatøørissæ or amatøørissa

(210) *vacillating loanword with back+/æ/+/i/:* miljonææri 'millionaire' ∼ miljonæærissæ or miljonæærissa

Similar mixed words with more than one mid or high front unrounded vowels after the back vowel almost exclusively take back suffixes, with some stems vacillating (211–213, data from Ringen and Heinämäki 1999). These are also mostly loanwords, native stems either don't have such forms, or are compounds.

(211) *mixed loanword with neutral vowels:* artikkeli 'article' ∼ artikkelissa, *artikkelissæ

(212) *mixed loanword with neutral vowels:* dynamiitti 'dynamite' ∼ dynamiittissa, *dynamiittissæ

(213) *vacillating mixed loanword with neutral vowels:* arkkitehti 'architect' ∼ arkkitehtissa or arkkitehtissæ

3.1.1 Exceptionality

Exceptionality in Finnish vowel harmony is manifested differently than in Hungarian. Antiharmonicity is extremely rare, only two stems show a similar behavior. These two stems take the back *-ta* partitive suffix instead of the front *-tæ* (214–215), but they take regular front suffixes otherwise (216–217):

(214) meri 'sea' ∼ merta, *mertæ

(215) veri 'sea' ∼ verta, *vertæ

(216) meri 'sea' ∼ meressæ, *meressa, *merassa

(217) veri 'sea' ∼ veressæ, *veressa, *verassa

To understand why Finnish has no antiharmonicity like the related Hungarian language, it can be seen that in Finnish most stems are disyllabic in a (C)VC(C)V pattern, while the stem final vowel was deleted in Hungarian after the 10th century resulting in the frequent (C)VC(C) monosyllabic pattern of Hungarian stems (Sammallahti 1988, E. Abaffy 2004b). This is illustrated in Table 3.3 below: while the stem final vowel is intact in Finnish throughout its history, the main changes being the voicing of the stem final consonant in closed syllables and the delabialization of the accusative suffix, in Hungarian the stem final vowel, when word final, is reduced to /u/ and later deleted (the accusative marker /t/ is an innovation in Hungarian). The forms in Proto-Finno-Ugric are parallel, showing a /CiCa/ structure, the diachronic developments led to a /CiC/ antiharmonic stem in Hungarian and to a conservative /CiCa/ mixed stem in Finnish.

	Hungarian		Finnish	
	Nom	Acc	Nom	Acc
Proto- (∼ 2000 BC)	*ɲila	*ɲilam	kita	kitam
Old (∼ 1000 AD)	ɲilu	ɲilat	kita	kidan
Modern (∼ 2000 AD)	ɲiːl	ɲilɒt	kita	kidan
gloss	‘arrow’		‘maw’	

Table 3.3: Diachronic parallelism between Hungarian antiharmonic and Finnish mixed stems

Therefore stems which would be antiharmonic in Hungarian could be seen as corresponding to a disyllabic stem in Finnish with a neutral first vowel and a back final vowel. Thus a /Ci,eCa-/ stem like (218) corresponds to an antiharmonic Hungarian stem (219); but synchronically it looks identical to a Hungarian mixed stem (220). A /Ci,eCæ-/ stem like (221) corresponds to a harmonic Hungarian stem (222). These examples are schematically summarized in Table 3.4.

origin	**V-final**	**C-final**
mixed	Finnish *kita* ∼ *kitassa*	Hungarian *ɲiːl* ∼ *ɲiːl-nɒk*
harmonic	Finnish *isæ* ∼ *isæssæ*	Hungarian *siːv* ∼ *siːv-nɛk*
different	Hungarian *libɒ* ∼ *libaː-nɒk*	
	Hungarian *ʧibɛ* ∼ *ʧibeː-nɛk*	

Table 3.4: Comparison of Finnish mixed stems and Hungarian antiharmonicity

(218) *Finnish mixed i-stem:* kita ‘maw’ ∼ kida-ssa

(219) *Hungarian antiharmonic i-stem:* ɲiːl ‘arrow’ ∼ ɲiːl-nɒk

(220) *Hungarian mixed back stem:* libɒ ‘goose’ ∼ libaː-nɒk

(221) *Finnish harmonic i-stem:* isæ ‘father’ ∼ isæ-ssæ

(222) *Hungarian harmonic i stem:* siːv ‘heart’ ∼ siːv-nɛk

The question of whether antiharmonicity is predictable in Hungarian based on phonetic or lexical factors, therefore, is parallel to the question of whether the frontness of the final vowel of disyllables with a neutral first vowel is predictable based on similar factors in Finnish. This chapter will focus on this question, more specifically, on the distribution of *a*-final stems versus *æ*-final stems after a syllable with a neutral vowel *e, ee, ei, i, ii* or *ie*. This pattern for stems will be abbreviated as CIC(C)A in this chapter, with I representing any short, long or diphthongal non-low front vowel and A representing /a/ or /æ/.

3.1.2 Previous study (Anderson 1980)

This chapter is not the first corpus study on Finnish mixed stems. Anderson (1980) has investigated mixed stems and the complicated issue of neutral vowels in front-back harmony in general. His main claim is that spreading [-back] can be weakened by *grave* consonants, labials and velars (/p, k, m, ŋ, ʋ/ in the native Finnish inventory, see Table 3.2), which have the coarticulatory effect of pulling the F2 of neighboring vowels down to a lower frequency. According to Anderson, this phonetic effect can have the effect of a preference of a [+back] vowel in its environment.

To give evidence for grave consonants prefering back environments, Anderson provided lexical statistics of Finnish disyllabic verbal stems where the first vowel was a non-low front vowel and the last vowel of the stem was /a/ or /æ/ — that is, verbal CIC(C)A stems. His sources were earlier studies and a manual search in the dictionary of Alanne (1968).

The claim of Anderson (1980) is that the (lexically determined) "choice" of the final vowel is predictable based on the quality of the first vowel and the quality of the second consonant in the stem. He treats "directly inflected" verbs and verbs with a (productive or non-productive) derivational suffix lexicalized in them, and claims that there is only one underived mixed stem. For the derived disyllables, he predicts the final vowel to be back after a grave consonant or after a high front vowel /i, ii, ei/, but harmonicity is preferred after a mid front vowel and a grave consonant. He explains the antiharmonicity of exceptional partitives *merta* 'sea.PART' and *verta* 'blood.PART' (214–217) by the fact that the /rt/ cluster is realized as a retroflex [ʈ] in the neighboring Swedish language, which is not an acute segment phonetically, therefore it might motivate back suffixation following it. This is illustrated in Table 3.5 below: the actual counts of -*a* final and -*æ* final stems is also included.

last consonant	first vowel	preference	-a	-æ
grave, but not /p/	any	mixed	46	2
acute or /p/	/i, ii, ei/	mixed	14	1
	/e/	harmonic	0	10

Table 3.5: Anderson's study of CIC(C)A stems, reproduced from Anderson (1980, p. 285)

Anderson's study is an early peek into sublexical regularities of exceptionality in vowel harmony. His paper, however, relied on manual collection of the data which constrained lexical studies at the time. Also, his distinction of derived ("protected") and underived verbs is somewhat arbitrary, as several verbs that are in his derived class are analyzed as monomorphemic by the morphological parser used later in this chapter, so it is suspicious that most of these verbs are not stored as morphologically complex in Finnish speakers' lexicons. The studies conducted in this chapter below build on Anderson's findings, but will include more data, using a large webcorpus and automated classification to obtain more reliable quantitative data.

3.2 Corpus analysis

This corpus study investigates the distribution of harmonic and mixed stems with a CIC(C)A structure: having a mid or high front first vowel and a low (either front or back) vowel in its second syllable. The main goal of this study is to find the phonotactic patterns that differentiate the harmonic and mixed sublexicon the best. The purpose of finding these patterns is to be tested in the wug study with Finnish speakers (Section 3.2.2), but there is a question about the phonetic grounding of the distribution between harmonic and mixed stems that the corpus study can attempt to give an answer to (Section 3.2.1).

3.2.1 Effect of gravity

The first issue investigated here is the verification of Anderson's hypothesis that the phonological feature [±grave] of a consonant in a given stem can predict the harmonicity of that stem. If the data from the corpus study confirms that grave medial consonants prefer mixedness, that would be evidence for Anderson's hypothesis, and would confirm a link between a feature based on perceptionand the phonotactic properties of the two (harmonic/mixed) sublexica. If no such evidence is found, that would cast a shadow of doubt on such a claim. The evaluation of this hypothesis based on the results is shown below in Table 3.6.

stems with grave consonant	stems with acute consonant	plausibility of the hypothesis
mixed more likely	mixed less likely	✓
equally likely to be mixed		×

Table 3.6: Evaluation of the hypothesis on the phonetic effect of consonant gravity based on the results of the corpus study

3.2.2 Knowledge of sublexical phonotactics

The broader question addressed in this chapter is the role of sublexical phonotactic regularities in speakers' grammars. Several case studies have indicated that speakers are aware of sublexical regularities, so they are able to conclude the morphological behavior or well-formedness of a new lexical item based on these sublexical patterns (Hayes et al. 2009, Gouskova and Becker 2013, Becker and Gouskova 2016, Allen and Becker 2015). In learning or in a wug experiment, speakers can compare how well a novel item fits phonotactically to the different subclasses it could be a member of, and they decide on their well-formedness judgments, or on the stem's morphological behavior based on this comparison.

The hypothesis on how Finnish speakers handle mixed stems suggested by the sublexical learning theory is that they have knowledge of the phonotactic patterns within the sublexicon of harmonic stems and within the sublexicon of mixed stems. To test this, first a lexical study was conducted based on the same corpus as the one testing the phonetic hypothesis, so that some phonotactic regularities can be found in both sublexica. Second, a nonce word study was run with Finnish speakers to see whether they are using these regularities to decide on the well-formedness of the word. The evaluation of the sublexical hypothesis is schematized in Table 3.7.

different phonotactics between harmonic and mixed stems?	speakers use phonotactic difference to categorize?	evaluation of sublexical hypothesis
none	—	N/A
present	no	×
present	yes	✓

Table 3.7: Evaluation of the hypothesis on the phonetic effect of consonant gravity based on the results of the corpus study

3.2.3 Methods

To establish a database that can be used to assess the role of different phonological properties in predicting (dis)harmonicity in Finnish, the pipeline described on Figure 3.1 was established and executed.

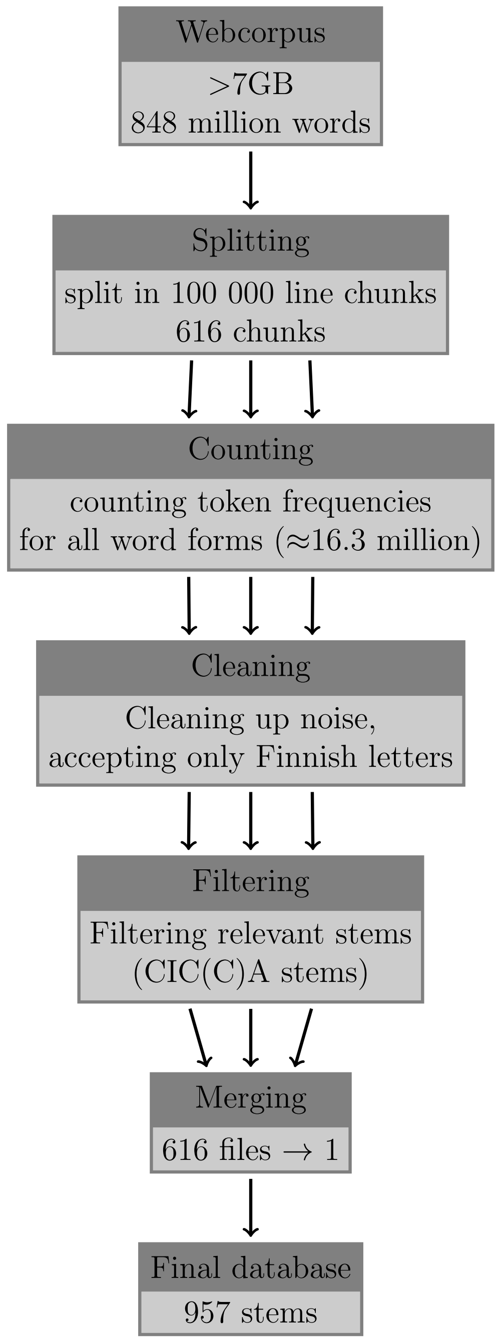


Figure 3.1: Flowchart of data processing for the corpus analysis

3.2.3.1 Corpus

The size of the corpus used for this study needs to be large enough to most likely contain all relevant Finnish stems. To ensure this, a webcorpus with $\approx$ 846 million tokens was used (Zséder et al. 2012). This corpus was collected by an automatized webcrawler, which gathered data from the `.fi` Finnish top-level domain. The crawler used an automatic language detector to filter out pages that might not have been in Finnish. This filter did not work perfectly, leaving in quite a lot of English, Swedish and other words in the corpus, which had to be dealt with later in the Filtering stage.

3.2.3.2 Splitting

The size of the corpus was over 7GB, and the tools in the latter parts of the pipeline were very inefficient or even failed while handling such a huge file. Therefore, the corpus was split so that a chunk contained 100 000 lines — splitting by line breaks was important so that words themselves do not get split. The following operations were done on these 616 chunks.

3.2.3.3 Counting

For each chunk of the corpus, a frequency distribution was generated. Such a distribution contained each token (case insensitively) that occured within the chunk with the frequency of its occurence.

3.2.3.4 Cleaning

The corpus itself was quite messy, therefore several rounds of cleaning and filtering was needed. The first operation was a simple cleaning which filtered out any characters that is not a letter in the Finnish alphabet. All 26 Latin alphabet letters from *a* to *z* were included, even though some of these occur only in loanwords, and the letters *ä* and *ö* were also allowed as these are frequent letters in Finnish. The letter *å* that is included in the Finnish alphabet was filtered out as it only occurs in Swedish or Fenno-Swedish names.

The words containing the removed characters were themselves kept in the database, allowing for punctuation marks to be ignored: a token like *"itachi-samaan"* turned into *itachisamaan*, letting the morphological analyzer decide on whether the word is legal in Finnish, or not, as described for the next step.

3.2.3.5 Filtering and stemming

The main filtering was done to ensure that a token is a Finnish word, and that its stem fits in the CIC(C)A pattern investigated in this study. As the study is only interested in the frequency of stems, tokens were grouped together by their stem, and further analysis was made on the stem.

Stemming was done using the open-source tool *Omorfi* (Pirinen 2014) available online. Omorfi, which is based on HFST (Helsinki Finite-State Technology, Lindén et al. 2009), provides a command-line toolbox to analyze Finnish words morphologically. Filtering was automatized by running the interactive tool provided by Omorfi, and requesting morphological information about every token. If the analyzer did not return an analysis at all, meaning that the word is unknown in Finnish, the token was discarded. If Omorfi did return morphological analysis, and the stem had a relevant CIC(C)A shape, the frequency of the token was added to the frequency to the entry of this stem. If there were more plausible and relevant analyses for a token, its frequency was split equally between the stems. For example, the token *voin* has 5 distinct analyses in Omorfi, 2 of which is an inflected form of the noun *voi* 'butter', another 2 is a form of the verb *voida* 'to be able to' and 1 is the plural instructive of the stem *vuo* ‘flow’. In a case like this, 2/5 of the frequency would be distributed to the stems *voi* and *voida* and 1/5 to the stem *vuo*. The output of the filtering process was a frequency distribution in the very same format as the result of the counting step, but this time, tokens were aggregated by their stem forms.

When stemming verbs, Omorfi analyzes the infinitive of the verb as the stem (eg. *puhua, antaa, kieltææ, mennæ*, etc.). This was not ideal for the purposes of this study as the infinitive contains a suffix *-a/æ/ta/tæ/(C)a/(C)æ* that is added to the stem morpheme (*puhu-, anta-, kieltæ-, men-*). To find the real base of the stem, the present indicative connegative form (*puhu, anna, kiellæ, mene*) was generated for each verbal morpheme using Omorfi, as this form is very close to the stem. The difference between the stem and the connegative form for vowel-final stems can be in the quality of the consonant before the vowel, which is subject to an alternation called gradation. This alternation does not influence the results of the analysis by much, as gradation mostly changes length or manner of a consonant and not its place of articulation, and has no effects on the vowels in the stem. For consonant-final stems, an *-e* is added in the connegative form (eg. *mene*), so these stems do not have a CIC(C)A shape, so (as intended) they will be excluded from the analysis.

Filtering required a lot of resources as the morphological analyzer had to go through each token. Therefore, it was conducted on a cloud-based server, provided by Amazon EC2. Filtering of all 616 chunks took about 2.5 days in total. These 616 chunks contained all the 957 relevant Finnish stems amongst each other.

3.2.3.6 Merging

Finally, the filtered frequency distribution chunks were merged into a single, 7.3 kB frequency distribution file, which contains all 957 relevant Finnish stems in the corpus with their frequencies. The most frequent stem was *ettæ* ‘that (complementizer)’, while the 6 least frequent stems had 1 occurence in the corpus in total.

3.2.4 Analyzing results

3.2.4.1 Verifying Anderson

The resulting frequency distribution file was first organized into harmonic (neutral-front, eg. *isæ*) and mixed (neutral-back, eg. *ihan*) categories. The first analysis was the verification of Anderson (1980). For this verification, overall type and token frequencies were aggregated by the category of the stem according to Anderson and harmonic category. Only stems with *one* consonant between the first and second vowel were included in both Anderson's work, and this verification study. Anderson differentiated between the following categories:

- *grave*: grave middle consonant except [p] (k, m, ŋ, ʋ)— predicted to be *mixed*
- *i-acute*: acute (t, d, s, n, l, r, j) or [p] middle consonant after [i(ː), ei] — predicted to be *mixed*
- *e-acute*: acute (t, d, s, n, l, r, j) or [p] middle consonant after [e(ː)] — predicted to be *harmonic*

The statistical analysis of the results was done using χ^2 statistics on the 2×3 table on the Andersonian $\times$ harmonicity conditions.

3.2.4.2 Discovering more patterns

There could be more patterns in the data, either of tendencies contradicting Anderson (1980), or — more likely — of tendencies not noted by earlier studies. In order to discover these patterns, the database was further explored using the UCLA Phonotactic Learner

(Hayes and Wilson 2008). The same methodology has been used by Hayes et al. (2009) to explore the properties of antiharmonic and transparent vowels in Hungarian, and by Becker and Gouskova (2016) to find generalizations on Russian stems containing mid vowels, which can be deleted (as they are "yers") based on the stem's membership in a sublexicon, or not. Very similarly, Gouskova et al. (2015) also use the UCLA Phonotactic Learner to discover what kinds of sublexical regularities are there in the lexicon for different Russian diminutive suffixes.

The main goal of using the Phonotactic Learner in this study is to find patterns that are less typical for one of the subsets of the lexicon. The constraints that the learner outputs are not considered to be more than a way of describing sublexica. Their main purpose will be to be used as an analytic tool to describe phonotactic regularities in the lexicon. As such, it will be tested in the wug study, whether Finnish speakers are able to generalize over these patterns, regardless of their phonetic or phonological naturalness, or not. The degree that they are found to be aware of these patterns will help assess how much the naturalness of a pattern affects grammatical generalizations based on the given pattern.

The database with words having a CIC(C)A structure was separated to front-final and back-final stems. The learner was trained separately on these two sublexica. The resulting 100 constraints each are therefore descriptive of the phonotactic properties of either CIC(C)a or CIC(C)æ words in Finnish. The strings disfavored based on these constraints were investigated manually, and 6 such patterns out of the 200 were chosen. Sequences which seemed to be relevant for the discrimination between the two sublexica, and which did not seem to state a generalization on the Finnish phonotactics in general were chosen.

The 5 selected patterns were tested with wug words. The reason for this was to find out whether these manually chosen patterns separate the two sublexica clearly. The degree of separation was measured by comparing the distribution (average, median, maximum) of harmony scores assigned by the UCLA Phonotactic Learner to the wug stems in both sublexica. The reason to compare the assigned scores between the two sublexica instead of measuring raw scores is to exclude possible interaction of other patterns that could influence the rating of the words. So if a given pattern disprefers forms in one of the sublexica, the corresponding forms in the other sublexicon could also be dispreferred, as other, more specific or more general dispreferred subpatterns could conspire to do that. The goal of this study is to find patterns that set harmonic and mixed stems apart the best, so such patterns have to be dismissed.

To provide a hypothetical example, a pattern could be found that penalizes medial -*k*- in the harmonic sublexicon. It could not be straightforwardly claimed that medial -*k*- prefers mixed stems: other patterns, like ones penalizing medial velars, medial stops, or more specific subpatterns like **ik* could result in a similar generalization against -*k*- in the mixed sublexicon as well. If the comparison between the scores in the harmonic sublexicon and the mixed sublexicon indeed shows that overall -*k*- is less preferred by harmonic stems, and no contradictory subpattern emerges, only then can the dispreference of a medial -*k*- in the harmonic sublexicon postulated.

There was a set of wug words generated for each pattern. These wugs were created to violate the pattern they were created for. Two versions were created for each wug: a front one (eg. *kejæ* for a pattern against *VjV sequences) and a back one (*keja*), no matter which sublexicon the pattern originated from. The UCLA Phonotactic Learner was retrained on the original corpus and tested with these wugs — the resulting harmony scores for the words were compared for the back and for the front version. Again, the reason for using the Phonotactic Learner was to have a phonologically sensitive way of discovering and testing patterns on a sublexicon, and not to devise constraint-based grammars that further the analysis. Because of this, the learner was trained 10 different times, to average out possible marginal effects of running the learner only once. The difference between scores indicates whether the two sublexica are indeed well separated based on the given pattern. This process is schematized in Figure 3.2.

An important detail to note is that during this process the number of stems in the lexicon which adhere to a given pattern was not taken into account. The job of discovering the usefulness to divide stems according to patterns was done entirely by the UCLA Phonotactic Learner, and the size of the sublexicon picked out of the pattern was ignored. This might have introduced some problems, as the membership sizes of these patterns were quite low, as explained below.

3.2.5 Results

3.2.5.1 Verifying Anderson

Out of the 957 total stems, only 240 (total token frequency 23 335 130) are harmonic, while 717 (14 853 211) are mixed. Only 345 stems had a plain, investigable CVCV structure — Table 3.8 shows the contingency table of these stems.

	harmonic		mixed		*total*
grave	14	(11.2%)	111	(88.8%)	125
i-acute	44	(28%)	113	(72%)	157
e-acute	25	(39.7%)	38	(60.3%)	63
total	83		262		345

Table 3.8: Distribution of CVCV stems in Andersonian categories

The difference between these conditions is significant, $\chi^2(2) = 21.08$, $p < 0.001$. Anderson's predictions are partially borne out: he claimed that the *grave* and *i-acute* classes prefer

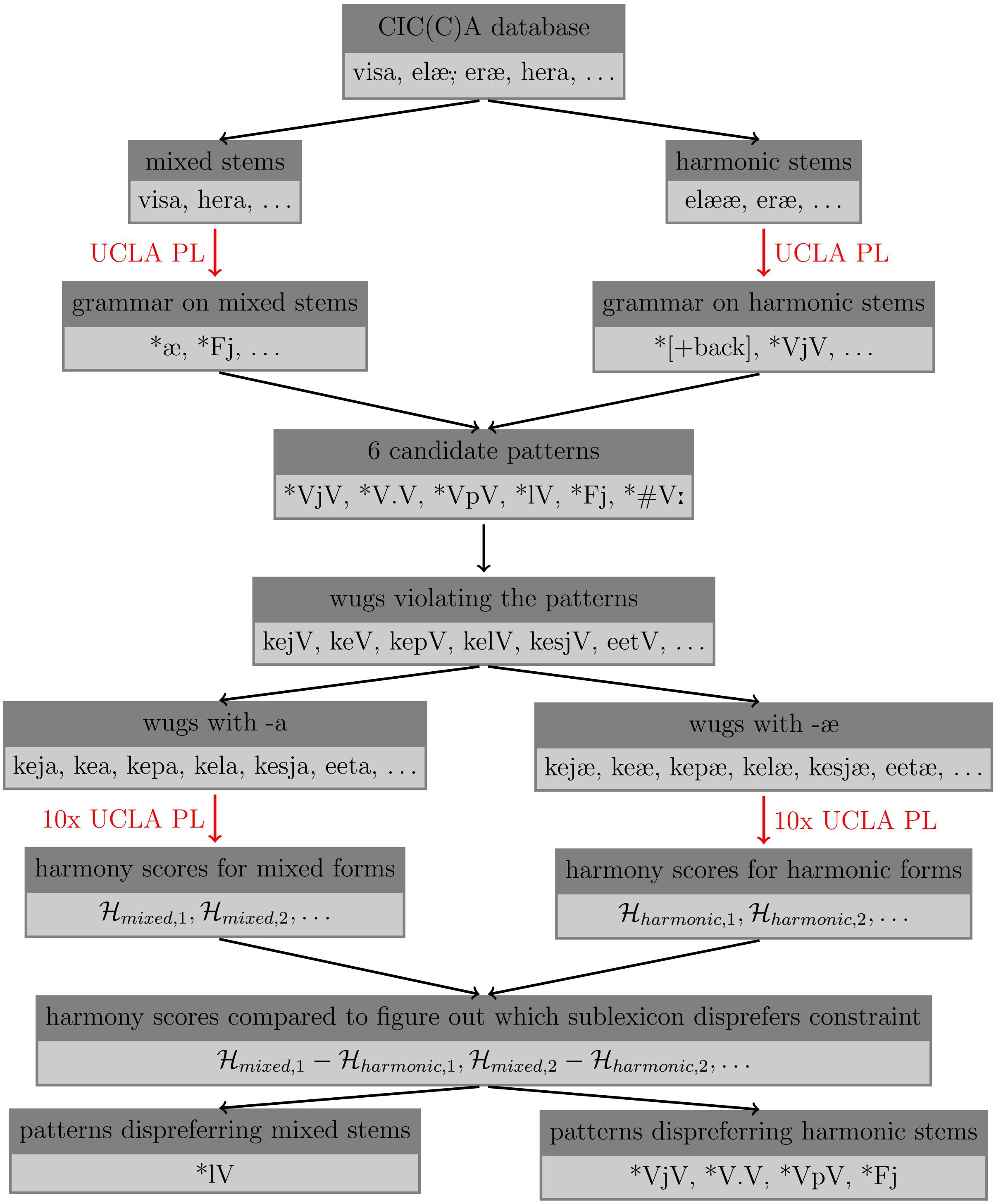


Figure 3.2: Flowchart of using the UCLA Phonotactic Learner to discover patterns that separate the front and the back sublexicon.

mixed stems, while *e-acute* stems prefer harmonicity. While mixed stems are the plurality within the *e-acute* class as well, contrary to Anderson's claims, it seems clear that this class prefers mixing the least by proportion.

Token frequency shows a bit different pattern, as illustrated on Table 3.9. It is visible that harmonic stems are more frequent on average, as harmonicity is much more frequent in token frequency than in stem type frequency. Now the majority of occurences are harmonic in all 3 categories, however, it is not *e-acute* stems that seem to prefer harmonicity the most, but *i-acute* stems, which prefer antiharmonicity according to Anderson (1980).

	harmonic		mixed		*total*
grave	2 154 241	(34.3%)	4 121 383	(65.7%)	6 275 624
i-acute	2 956 881	(36.8%)	5 088 785	(63.25%)	8 045 666
e-acute	744 013	(34.2%)	1 432 714	(65.8%)	2 176 727
total	5 855 135		10 642 882		16 498 017

Table 3.9: Distribution of CVCV stems in Andersonian categories by token frequency

3.2.5.2 Discovering more patterns

As described in Section 3.2.4.2, the UCLA Phonotactic Learner was used to identify phonotactic patterns that are dispreferred in either the back or the front sublexicon, but do not have such a role in the other one. As described on the flowchart in Figure 3.2, the database built from the corpus was split to two sublexica — one with all forms with a final back vowel and one with forms containing a front vowel.

The UCLA Phonotactic Learner was trained on both sublexica. The learner was set to look for 100 constraints based on at most three adjacent segments. The vocalic tier and the consonantal tier were *not* separated, as the possible interactions between consonants and vowels are relevant in this study. In other settings the default parameters were used.

Out of the 200 discovered constraints, the 5 described below were picked out for more investigation. The patterns these constraints motivate against do not seem to be very strict generalizations about Finnish phonotactics in general, and they separate the two sublexica clearly, based on the initial observations. To see whether these patterns really separated mixed and harmonic stems, wug stems were created that violated these patterns, as described in Section 3.2.4.2, and the Phonotactic Learner was tested on these stems.

The main goal of picking out these patterns was to find phonotactic regularities that are the most likely to be learned by Finnish speakers. The most important property of these patterns is their ability distinguish between the mixed and harmonic sublexica the best. The complexity of these patterns is therefore only secondary.

3.2.5.3 Pattern 1: Harmonic stems disprefer intervocalic *-j-*

The first results for the grammar for harmonic stems contained a pattern motivating against postvocalic *-j* in the form of a *[-consonantal][+palatal] restriction. This is a ban against CVjV, CVjCV and CVC(C)Vj forms as well, however, for the latter two structures, the mixed grammar provides patterns that motivate against them as well. Therefore, the only difference between the two grammars was that intervocalic -j- was discouraged in the grammar for harmonic stems.

constraint	mixed test words	harmonic test words
*-VjV-	keja, kija, keeja, kiija, keija, kieja, teja, tija, teeja, tiija, teija, tieja, peja, pija, peeja, piija, pieja, peija, meja, mija, neja, nija, leja, lija, reja, rija, seja, sija, heja, hija	kejæ, kijæ, keejæ, kiijæ, keijæ, kiejæ, tejæ, tijæ, teejæ, tiijæ, teijæ, tiejæ, pejæ, pijæ, peejæ, piijæ, piejæ, peijæ, mejæ, mijæ, nejæ, nijæ, lejæ, lijæ, rejæ, rijæ, sejæ, sijæ, hejæ, hijæ

Table 3.10: Test words for Pattern 1: intervocalic *-j-*

The second test was run with the test words presented in Table 3.10. The variety of the test words is needed to filter out any possible secondary effects of the other segments in the word. To do this, words with initial stops in all places of articulation (*p-, t-, k-*), two initial fricatives (*s-, h-*), initial nasals (*m-, n-*) and liquids (*l-, r-*) are included. Also, all relevant first vowel phonemes (*-e-, -i-, -ei-, -ie-, -ee-, -ii-*) are tested as well.

The results have confirmed the dispreference of intervocalic *-j-* for harmonic stems. The 30 test words had a harmony score of 1.45 assigned by the mixed grammar on average for the mixed stems and 3.45 assigned by the harmonic grammar for harmonic stems. After 10 iterations of the grammar, only six candidates received a higher mixed score than harmonic score (*kiej-, kiij-, piej-, piij-, tiej-, tiij-*), possibly indicating the presence of a more restricted subpattern after initial stops and long high vowels prefering harmonicity. Consequently, the harmonic stems were rated worse by their grammar, than the antiharmonic ones by theirs, for the other 24 stems. This difference was statistically significant, based on a two-sided paired t-test ($t(29) = 4.23, p < 0.001$). So the results confirm that a medial intervocalic *-j-* is dispreferred in harmonic stems, and not dispreferred in mixed ones.

3.2.5.4 Pattern 2: Harmonic stems disprefer hiatus

The second pattern picked out from the comparison of the harmonic and mixed grammar was that *mixed* stems disprefer hiatus, though it seemed to matter most after long vowels, and the least after [e]. For the test words, though (see Table 3.11), all possible first vowels were tested.

constraint	mixed test words	harmonic test words
*-V.V-	kea, keea, keia, kia, kiea, kiia, lea, leea, leia, lia, liia, mea, meea, meia, mia, miia, nea, neea, neia, nia, niia, sea, seea, seia, sia, siea, siia, tea, teea, teia, tia, tiia	keæ, keeæ, keiæ, kiæ, kieæ, kiiæ, leæ, leeæ, leiæ, liæ, liiæ, meæ, meeæ, meiæ, miæ, miiæ, neæ, neeæ, neiæ, niæ, niiæ, seæ, seeæ, seiæ, siæ, sieæ, siiæ, teæ, teeæ, teiæ, tiæ, tiiæ

Table 3.11: Test words for Pattern 2: hiatus

The results show that this is also a good separating constraint, however, it is going in the different direction than what the first pattern discovery found: *harmonic* stems disprefer hiatus more. There was no test word amongst the 32 analyzed that preferred harmonicity in the lexicon, even after 10 iterations of the learning algorithm. The average score for mixed stems was 2.15, while the average score for harmonic stems was 9.88, a significant difference (paired t-test: $t(31) = 25.82, p < 0.001$).

It might seem unnatural at first sight that hiatus in the word might influence the sublexical classification of a stem, but the effect of hiatus might have a phonologically reasonable reason for such a behavior. Hiatus in a harmonic word creates neighboring front vowels, a stronger OCP violation than the mixed vowel-vowel sequences. Hiatus seems to be a recurring indicator of sublexical differences: Gouskova et al. (2015) also found hiatus to be a strong predictor for the Russian diminutive allomorph selection.

3.2.5.5 Pattern 3: Harmonic stems disprefer intervocalic *-p-*

The third pattern found in the harmonic grammar indicated that intervocalic voiceless labial segments (*-p-, -f-*) are dispreferred in this sublexicon. However, certain patterns from the mixed grammar interact with this: intervocalic *-f-* and *-p-* after long high vowels *-ie-, -ii-* are dispreferred in the mixed stems as well. Testing the exact behavior of the pattern was therefore especially important here. Due to the common dispreference for *-f-*, only medial *-p-* was tested.

constraint	mixed test words	harmonic test words
*-VpV-	epa, eppa, ipa, ippa, keepa, keeppa, keipa, keippa, kepa, keppa, kiepa, kieppa, kiipa, kiippa, kipa, kippa, leepa, leeppa, leipa, leippa, lepa, leppa, liepa, lieppa, liipa, liippa, lipa, lippa, meepa, meeppa, meipa, meippa, mepa, meppa, miepa, mieppa, miipa, miippa, mipa, mippa, neepa, neeppa, neipa, neippa, nepa, neppa, niepa, nieppa, niipa, niippa, nipa, nippa, seepa, seeppa, seipa, seippa, sepa, seppa, siepa, sieppa, siipa, siippa, sipa, sippa, teepa, teeppa, teipa, teippa, tepa, teppa, tiepa, tieppa, tiipa, tiippa, tipa, tippa	epæ, eppæ, ipæ, ippæ, keepæ, keeppæ, keipæ, keippæ, kepæ, keppæ, kiepæ, kieppæ, kiipæ, kiippæ, kipæ, kippæ, leepæ, leeppæ, leipæ, leippæ, lepæ, leppæ, liepæ, lieppæ, liipæ, liippæ, lipæ, lippæ, meepæ, meeppæ, meipæ, meippæ, mepæ, meppæ, miepæ, mieppæ, miipæ, miippæ, mipæ, mippæ, neepæ, neeppæ, neipæ, neippæ, nepæ, neppæ, niepæ, nieppæ, niipæ, niippæ, nipæ, nippæ, seepæ, seeppæ, seipæ, seippæ, sepæ, seppæ, siepæ, sieppæ, siipæ, siippæ, sipæ, sippæ, teepæ, teeppæ, teipæ, teippæ, tepæ, teppæ, tiepæ, tieppæ, tiipæ, tiippæ, tipæ, tippæ

Table 3.12: Test words for Pattern 3: intervocalic -*p*-

The results indicate that harmonic stems indeed disprefer medial -*p*- more than mixed stems: the average harmony score was 3.736 for harmonic and 0.99 for mixed stems (paired t-test $t(75) = 6.33, p < 0.001$). The pattern is more restricted, however, as 14 pairs show equal preference and 13 pairs actually show a reversed pattern.

Neither harmonic or mixed stems show a dispreference for *-p-* after an *-e-*: *ep-, epp-, kep-, kepp-, lep-, lepp-, mep-, mepp-, nep-, nepp-, sep-, sepp-, tep, tepp-* have an average score of 0 in both grammars,

The inverse pattern, dispreference of mixed stems is present with 13 stems, as seen on Table 3.13. These also show that an intervocalic labial stop is not dispreferred in a harmonic stem if it is preceded by a high mid vowel (*-ee-* or *-ei-*). Only one non-mid vowel (*-ie-*) disprefers mixedness, but the harmonic pair of that test stem is also quite dispreferred. This subpattern is an issue not discussed further.

stem	mixed score	harmonic score
keep-	2.529	0
keip-	2.529	0
kiep-	4.015	3.428
leep-	2.529	0
leip-	2.529	0
meep-	2.529	0
meip-	2.529	0
neep-	5.39	2.867
neip-	5.39	3.536
seep-	2.529	0
seip-	2.529	0
teep-	2.529	0.992
teip-	2.529	1.661

Table 3.13: Dispreferred mixed stems with intervocalic *-p-*

The summary of the above is that intervocalic *-p-* is dispreferred in harmonic stems, especially if it is preceded by a high neutral vowel. A further comparison of all wug stems with *-e/ei/eep(p)-*, ie. mid vowels followed by a labial stop shows that mixed stems are still somewhat preferred, though the difference is not significant anymore (mean harmonic = 1.62, mean mixed = 1.16, $t(35) = 1.14, p = 0.27$). The dispreference of long high vowels before *-p-* for mixed stems as predicted by the initial grammar is not borne out at all (mean harmonic = 4.37, mean mixed = 1.39, $t(23) = 6.36, p < 0.001$).

Taking the complications with mid vowel stems into account, they are not included in the further study. Therefore, the main generalization for medial *-p-* is that harmonicity is highly dispreferred after high vowels (*-ipæ-*, *-iipæ-*, *-iepæ* ≪ *-ipa-*, *-iipa-*, *-iepa-*). This is the pattern with a medial *-p-* which distinguishes between harmonic and mixed stems the best, so this is the pattern Finnish speakers are expected to be sensitive to, given the hypothesis that they use sublexical regularities to categorize stems as harmonic or mixed. The high vowel+*-p-* pattern is the one that will be used therefore in the wug study.

3.2.5.6 Pattern 4: Harmonic stems disprefer intervocalic *-l-*

After the initial inspection of the grammars produced by the UCLA Phonotactic Learner it seemed that harmonic stems dispreferred medial singleton or geminate -l(l)-, but with further inspection the opposite result turned out to be the case: mixed stems disprefer *-l(l)-* more.

constraint	mixed test words	harmonic test words
*Vl(l)V	eela, eella, eila, eilla, ela, ella, iela, iella, iila, iilla, ila, illa, keela, keella, keila, keilla, kela, kella, kiela, kiella, kiila, kiilla, kila, killa, meela, meella, meila, meilla, mela, mella, miela, miella, miila, miilla, mila, milla, neela, neella, neila, neilla, nela, nella, niela, niella, niila, niilla, nila, nilla, reela, reella, reila, reilla, rela, rella, riela, riella, riila, riilla, rila, rilla, seela, seella, seila, seilla, sela, sella, siela, siella, siila, siilla, sila, silla, teela, teella, teila, teilla, tela, tella, tiela, tiella, tiila, tiilla, tila, tilla	eelæ, eellæ, eilæ, eillæ, elæ, ellæ, ielæ, iellæ, iilæ, iillæ, ilæ, illæ, keelæ, keellæ, keilæ, keillæ, kelæ, kellæ, kielæ, kiellæ, kiilæ, kiillæ, kilæ, killæ, meelæ, meellæ, meilæ, meillæ, melæ, mellæ, mielæ, miellæ, miilæ, miillæ, milæ, millæ, neelæ, neellæ, neilæ, neillæ, nelæ, nellæ, nielæ, niellæ, niilæ, niillæ, nilæ, nillæ, reelæ, reellæ, reilæ, reillæ, relæ, rellæ, rielæ, riellæ, riilæ, riillæ, rilæ, rillæ, seelæ, seellæ, seilæ, seillæ, selæ, sellæ, sielæ, siellæ, siilæ, siillæ, silæ, sillæ, teelæ, teellæ, teilæ, teillæ, telæ, tellæ, tielæ, tiellæ, tiilæ, tiillæ, tilæ, tillæ

Table 3.14: Test words for Pattern 4: hiatus

The set of test words is again seen on Table 3.14. The results show that harmonic stems indeed have a higher (less preferred) harmony score (mean = 1.84) than harmonic stems (mean = 1.487). This difference is not significant, paired $t(83) = 1.56, p = 0.122$, which can be attributed to the low number of such stems in the lexicon. This pattern is not as strong as the ones above, so further analysis was conducted on subsets of this set of stems.

stem	mixed score	harmonic score
iell-	6.1265	4.1979
iill-	5.9092	4.1979
keell-	3.1096	2.8257
kiel-	1.4861	0
kiell-	4.5956	0
kiil-	0.064	0
kiill-	3.1735	0
meell-	3.1096	2.8257
miell-	3.1096	0
miill-	3.1096	0
neell-	5.9707	5.6922
niell-	3.1096	0
niill-	3.1096	0
reell-	3.1096	2.8257
riell-	3.1096	0
riill-	3.1096	0
seell-	3.1096	2.8257
siell-	3.1096	0
siill-	3.1096	0
tiell-	3.1096	0
tiill-	3.1096	0

Table 3.15: The subset of stems with intervocalic *-l-* that dispreferred harmonicity.

Out of the 84 wug stems in total, 29 were more preferred by the mixed grammar, 21 preferred by the harmonic grammar, and a tie with 34 stems. The stems preferring harmonicity are presented in Table 3.15. These stems do seem to share some common properties, as two subpatterns can be observed. 17 out of 21 stems that prefer harmonicity contain a geminated *-ll-* after a long vowel — this subpattern seems to be responsible for most of the exceptionality in this data.

A further set of 34 stems have an average harmony score of 0 in *both* grammars. The list of these stems can be seen in Table 3.16 — no straight regular subpattern can be extracted from this list. Despite these problems, on average, medial *-l-* is still dispreferred in the mixed grammar, so this broad generalization is the one that will be tested in the wug study.

stem
el-, ell-, il-, ill-
keel-, kel-, kell-
meel-, mel-, mell-, miel-, miil-, mil-, mill-
nel-, nell-, niel-, niil-, nil-, nill-
reel-, riel-, riil-
sel-, seel-, sell-, siel-, siil-, sil-, sill-
tel-, tell-, tiel-, tiil-

Table 3.16: Stems with intervocalic *-l-* not dispreferred by either grammars

3.2.5.7 Pattern 5: Mixed stems disprefer voiceless fricative+-j- clusters — except for /fj/

In the initial grammar for mixed stems, there had been a constraint, that penalized clusters consisting of a fricative followed by /j/. This pattern of consonant clusters does exist in the language, monomorphemically frequently like /hj/, as in *pohja* 'bottom', *ehjæ* 'intact', etc.

The set of wug word forms used to test this pattern can be seen in Table 3.17. This set again contains a variety of frequent word initials (*k-, l-, t-*) and all investigated first syllable vowels, and the four fricative+/j/ combinations of *-fj, -sj-, -hj-* and *-vj-*.

constraint	mixed test words	harmonic test words
*Fj	keefja, keehja, keesja, keevja, kefja, kehja, keifja, keihja, keisja, keivja, kesja, kevja, kiefja, kiehja, kiesja, kievja, kifja, kihja, kiifja, kiihja, kiisja, kiivja, kisja, kivja, leefja, leehja, leesja, leevja, lefja, lehja, leifja, leihja, leisja, leivja, lesja, levja, liefja, liehja, liesja, lievja, lifja, lihja, liifja, liihja, liisja, liivja, lisja, livja, teefja, teehja, teesja, teevja, tefja, tehja, teifja, teihja, teisja, teivja, tesja, tevja, tiefja, tiehja, tiesja, tievja, tifja, tihja, tiifja, tiihja, tiisja, tiivja, tisja, tivja	keefjæ, keehjæ, keesjæ, keevjæ, kefjæ, kehjæ, keifjæ, keihjæ, keisjæ, keivjæ, kesjæ, kevjæ, kiefjæ, kiehjæ, kiesjæ, kievjæ, kifjæ, kihjæ, kiifjæ, kiihjæ, kiisjæ, kiivjæ, kisjæ, kivjæ, leefjæ, leehjæ, leesjæ, leevjæ, lefjæ, lehjæ, leifjæ, leihjæ, leisjæ, leivjæ, lesjæ, levjæ, liefjæ, liehjæ, liesjæ, lievjæ, lifjæ, lihjæ, liifjæ, liihjæ, liisjæ, liivjæ, lisjæ, livjæ, teefjæ, teehjæ, teesjæ, teevjæ, tefjæ, tehjæ, teifjæ, teihjæ, teisjæ, teivjæ, tesjæ, tevjæ, tiefjæ, tiehjæ, tiesjæ, tievjæ, tifjæ, tihjæ, tiifjæ, tiihjæ, tiisjæ, tiivjæ, tisjæ, tivjæ

Table 3.17: Test words for Pattern 5: fricative+/j/ clusters

The results show that the difference between harmonic and mixed stems is significant: the mean harmony score for mixed stems was 5.114, and for harmonic stems the mean score was 4.207 ($t(71) = 2.025, p = 0.046$). Further investigation is possible by investigating the behavior of the separate clusters, as seen in Table 3.18:

cluster	fj	sj	hj	vj
harmonic	10	0	15	15
mixed	8	18	3	3

Table 3.18: Number of preferred stems of different fricative+/j/ clusters

Based on this table, it looks like /sj/ clusters are preferred in the mixed grammar — stems with this cluster was never preferred by the harmonic grammar. Stems with /vj/ and /hj/ clusters are preferred in the harmonic grammar, while /fj/ clusters are not particularly preferred in either of the sublexica. The behavior of these subsets of the wug stems were further analyzed individually. For stems with /sj/, the mixed grammar did fare better (average score 3.536) than the harmonic grammar (6.976), $t(17) = 7.85, p < 0.001$. For stems with /vj/, the mixed grammar has an increased dispreference with an average score of 7.961, therefore the harmonic grammar is more preferred (average score 4.236, $t(17) = 4.176, p < 0.001$). Similarly, stems with /hj/ preferred harmonicity (3.511 score versus harmonic 0.978, $t(17) = 3.569, p = 0.002$). The statistical test confirmed that /fj/ is not underattested in either grammar, average mixed score being 6.534, harmonic score 7.264, $t(17) = 0.85, p = 0.41$.

As the only strong subpattern indicating dispreference against harmonic stems was due to wug stems with /sj/, it would seem to make sense to exclude stems with this cluster. However, it would be a slightly odd and rather complex generalization to look for noncoronal fricative + /j/ clusters, therefore /sj/ was included in the study, and the further breakdown of the analysis of stems with different consonants before the palatal approximant was done after the wug stem study.

3.2.5.8 A discarded pattern: Neither grammar disprefers initial long vowels

A pattern in the initial grammar indicated that word initial long vowels are dispreferred in the mixed grammar. This was tested with the wug stems seen in Table 3.19. All four possible word initial long vowels (*ee-, ei-, ie-, ii-*) were investigated, and a set of consonants, both singleton and geminate, with different places and manners of articulation (*-k-, -n-, -r-, -t-*).

constraint	mixed test words	harmonic test words
VV-	eeka, eekka, eena, eenna, eera, eerra, eeta, eetta, eika, eikka, eina, einna, eira, eirra, eita, eitta, ieka, iekka, iena, ienna, iera, ierra, ieta, ietta, iika, iikka, iina, iinna, iira, iirra, iita, iitta	eekæ, eekkæ, eenæ, eennæ, eeræ, eerræ, eetæ, eettæ, eikæ, eikkæ, einæ, einnæ, eiræ, eirræ, eitæ, eittæ, iekæ, iekkæ, ienæ, iennæ, ieræ, ierræ, ietæ, iettæ, iikæ, iikkæ, iinæ, iinnæ, iiræ, iirræ, iitæ, iittæ

Table 3.19: Test words for Pattern 6: initial long vowel

The results did not confirm the hypothesis based on the initial grammar — the mixed grammars treated the mixed stems similarly (average score 4.028) as harmonic grammars treated harmonic stems (3.769), $t(31) = 0.432, p = 0.67$. The contingency table in Table 3.20 below summarize the behavior of subpatterns based on the quality of the initial vowel.

initial vowel	ee	ei	ie	ii
harmonic	3	3	5	5
mixed	5	5	3	3

Table 3.20: Number of preferred stems of different initial long vowels

It seems that there is no clear pattern based on the quality of the initial vowel. A distinction of mid-onset (/ee/ and /ei/) and high-onset (/ie/ and /ii/) vowels could be made, and while mid-onset vowels on average do prefer mixed stems and high-onset vowels do prefer harmonic stems, neither of these preferences approach statistical significance.

3.2.5.9 Number of stems selected by the patterns

As indicated above in Section 3.2.4.2, the number of stems whose shape conforms a given pattern was not taken into account in the process of finding them. This might be a problem, as illustrated below in Table 3.21, which shows the number of stems in the lexicon conforming to the given pattern. These numbers are quite low, especially for 1, 2 and 5:

pattern	Pattern 1	Pattern 2	Pattern 3	Pattern 4	Pattern 5
n	6	5	12	33	1
harmonic	1	5	12	13	1
mixed	5	0	0	20	0

Table 3.21: Number of stems selected by a given pattern

These ns were not taken into account as the performance of the UCLA Phonotactic Learner was trusted, following similar studies, like Hayes et al. (2009). However, the effect of these low numbers might have influenced the results of the wug study discussed in Section 3.3.

3.2.6 Discussion

Section 4.2.3 introduced 5 patterns based on the investigation of the online corpus. These patterns are phonotactic generalizations over the harmonic and mixed sublexica of Finnish stems of a CIC(C)A shape. One of the patterns investigated (the one about stems with initial long vowels) turned out to separate the two grammars poorly, therefore it was not included in the further studies. The final form of the five remaining generalizations are shown in Table 3.22 below:

	pattern	disprefers	example
1	intervocalic *-j-*	harmonic	keja > kejæ
2	hiatus	harmonic	kea > keæ
3	intervocalic *-p(p)-* after high vowels (*-ip(p)-*, *-iip(p)-*, *-iep(p)-*)	harmonic	kippa > kippæ
4	intervocalic *-l(l)-*	harmonic	kilæ > kila
5	fricative + *j* clusters (*-Fj-*)	mixed	kihja < kihjæ

Table 3.22: Summary of patterns identified in the corpus

These patterns represent generalizations based the corpus. Their validity was tested twice: after their identification using the UCLA Phonotactic Learner, their ability to separate the harmonic and mixed sublexica was queried again in a quantitative way. This ability to separate the sublexica was statistically significant for all five patterns. It is to be noted, though, that none of these patterns seem to be very strong or categorical: their presence was only found with a combination of computational learning and manual examination.

The next task was to find out whether Finnish speakers are aware of sublexical patterns setting harmonic and mixed stems apart. This was investigated using a wug test, where speakers were tested on whether they use this knowledge to rule novel stems as more likely to be harmonic or mixed.

3.3 Wug study

The patterns identified in the previous section were tested to see if Finnish speakers are aware of these phonotactic generalizations and use them in determining the categoriazation of a novel stem as more likely to be harmonic or mixed. If speakers use these patterns in categorization, that is an argument for the hypothesis that speakers learn sublexical regularities as part of the grammar.

There are alternative hypotheses besides the sublexical one that could predict different outcomes for the experiment. One such hypothesis is that speakers are more sensitive to vowel frequencies in the language: either simply to the unigram frequency of the final vowel (overall frequency of word final -*a* and -*æ*), or to vowel-to-vowel dependencies. The latter would mean that speakers would be much more sensitive to the quality of the first stem vowel than to the intervening consonant(s), so the expected differences between the patterns set up for the experiment would be suppressed by the effects of the vowels themselves.

3.3.1 Methods

3.3.1.1 Stimuli

The creation of stimuli was automatized, as there were several requirements on the form of the wug stems that were to be selected. First, the UCLA Phonotactic Learner was trained on the *whole stemmed corpus* — the entire corpus was stemmed using Omorfi as described in Section 3.2 so that only existing Finnish stems are analyzed. All possible CIC(C)A forms, as well as (C)IA forms were generated, and their harmonicity score based on the entire grammar was calculated using the UCLA Phonotactic Learner. This was necessary, because the goal of the wug study was to isolate the effect of the patterns identified in the corpus on Finnish speakers' judgments, so only wug stems that had a harmony score of 0 were used in the study, as stems with a 0 harmony score are deemed consistent with the general phonotactics in Finnish.

After all phonotactically plausible wug stem forms were generated, stems that actually exist as a word in Finnish had to be filtered out using Omorfi. For a given stem like *tirk-*, both the non-existence of *tirka* and *tirkæ* had to be checked, as both forms were going to be present in the study. After the set of possible wugs was restricted to pairs of unknown words ending in *-a* or *-æ*, these words were annotated based on which pattern's requirements out of the 5 identified in the corpus study they satisfied. Finally the stimuli were sampled: 20 wugs for each pattern and 60 control wugs that had a CIC(C)A shape but did not fit any of the patterns.

After generating these nonce stems, it was noticed that a lot of them contained the voiced stops *b*, *d* or *g*. Since the corpus contained a lot of foreign or loanwords that the morphological parses accepted, there was no overall phonotactic constraint against these consonants. However, as the presence of a voiced stop (except for *d* in morphologically marked contexts) is a marker for unassimilated loanwords, it was not optimal to have stems with any of these segments in the set of stimuli. Therefore all wugs containing any voiced stops were eliminated and replaced with an acceptable nonce stem from a second run of the sampling.

The generation of wug stems ran into a problem with generating stems for Pattern 5, the pattern with a fricative + *j* cluster. Out of the four possible clusters *fj*, *vj*, *sj* and *hj*, not one wug word with a labial fricative was generated. The reason for this was that all possible wugs with *vj* or *fj* between the two vowels in the word had a harmony score higher than 0, as it seems that these clusters are rare or nonexistent in Finnish. The issue with this limitation is that, as discussed in Section 4.2.3, that the remaining two clusters behave quite differently: *sj* disprefers, while *hj* prefers harmonicity. Therefore the analysis of this pattern was split in two, and *sj* and *hj* were analyzed separately too. Analyzing the behavior of this pattern still grouped together still makes sense: speakers could be aware of the different behavior of the two possible clusters but they might only behave as if they would be part of one generalization. One stem with *fj* and with a low harmony score was handpicked to be included in the analysis, though, in order to see whether such a stem would indeed behave differently from others.

The final set of stimuli is presented in Table 3.23.

3.3.1.2 Procedure

The experiment had a forced choice design: participants had to decide for each stem, whether it sounds better as a Finnish word with a front low or a back low final vowel. Before the experiment started, participants filled out a consent form and a demographic form. This ensured that the subjects were over 18, and collected data about whether they lived, or had previously lived outside Finland for an extensive period. After the demographic form, the participants were shown 80 pairs of wugs out of the 160 pairs in total: out of the 60 control pairs, 30 were randomly selected for each subject, and out of the 20 pairs for each pattern, 10 were randomly sampled for presentation.

For each pair, the nonce words ending in -a and in -æ were presented in a random order. The task of the subjects was to press the button with the label of their preferred choice. The

Stimuli	Pattern tested	Focusing on
viitv-, vith-, vepr-, vemm-, tirk-, tern-, sikh-, siism-, siern-, serk-, seeht-, rirr-, remh-, rert-, rek-, reish-, reesm-, pirh-, pelk-, peesv-, nisv-, nish-, niirr-, netn-, nesp-, neisp-, neepr-, neem-, milv-, mils-, miir-, mehn-, meelt-, lirr-, liitl-, kimv-, kiisp-, kenn-, keim-, keeht-, jimp-, jiikn-, hinl-, hiisv-, hels-, heiht-, heik-, hehl-, herk-, pets-, lisn-, firb-, fiis-, fens-, feeh-, merv-, kikl-, jelm-, hiinr-, liihm-	Controls	to see what the default pattern is, if there is any
vij-, veij-, tiij-, tej-, sej-, siej-, seej-, riij-, reej-, mij-, miej-, meij-, meej-, kiej-, jeij-, jeej-, hiij-, fiej-, kij-, hej-	Pattern 1	to see if intervocalic *-j-* indeed prefers mixedness
ve-, ti-, te-, se-, re-, pi-, pe-, ni-, ri-, ne-, mi-, me-, li-, le-, ki-, ji-, je-, he-, hi-, fi-	Pattern 2	to see if hiatus indeed prefers mixedness.
tip-, lip-, siip-, siepp-, piep-, piepp-, niip-, niep-, mip-, kiep-, jipp-, jiip-, jiipp-, jiep-, hiepp-, vip-, viip-, riip-, riepp-, sip-	Pattern 3	to see if intervocalic *-p(p)-* after high vowels indeed prefers mixedness
veil-, siil-, siill-, sel-, rill-, riil-, peil-, peel-, nill-, miill-, miil-, mell-, lell-, leil-, leell-, kil-, rill-, fiel-, pill-, keell-	Pattern 4	to see if intervocalic *-l(l)-* indeed prefers harmonicity
tiesj-, siisj-, sihj-, riesj-, niehj-, nesj-, neehj-, liehj-, lesj-, keisj-, rehj-, fiesj-, fiehj-, feihj-, peisj-, kiihj-, meihj-, hehj-, nefj-	Pattern 5	to see if fricative + j clusters indeed prefer mixedness

Table 3.23: Stimuli for the wug study

prompt asked subjects to select the word that sounds better as a Finnish word to them.

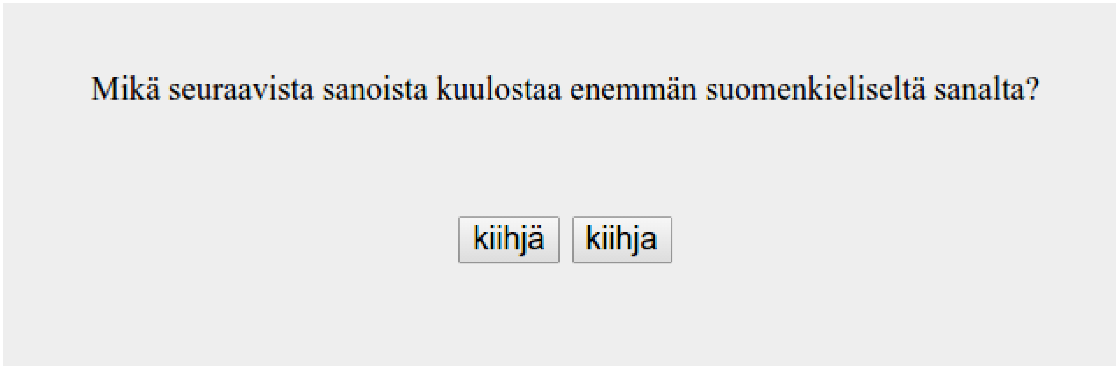


Figure 3.3: Screenshot of a forced choice test. The prompt translates to *Which of the following made up words sounds better as a Finnish word?*

The experiment was implemented in the Experigen online experiment framework (Becker and Levine 2010), and it was hosted on the S3 service of Amazon Web Services. The link to the experiment was disseminated on Facebook. Besides sharing the link and appealing to Finnish-speaking friends and friends of friends to do the experiment, it was also advertised to a focus group of Finnish speakers who are interested in the Finnish language.

3.3.1.3 Participants

The experiment was completed by 288 people in total. Since this number was much more than expected, and enough to acquire conclusive results, speakers who were likely to have a strong influence of a language other that Finnish were excluded. First, 14 people who reported not being (exclusively) a native Finnish speaker were excluded. After this, 18 speakers who filled the experiment while living outside of Finland were excluded. Finally, 90 subjects who reported to have lived 6 or more months abroad were also kept out of the experiment, yielding 172 subjects whose data was analyzed.

The mean age of the 172 subjects who filled out the experiment was 46.9 years (with

SD=13.41), ranging from 19 to 80 year olds. Out of these participants, 149 claimed to be able to speak English, followed by 120 who claimed some proficiency in Swedish (learning Swedish is obligatory in the Finnish education system). These two languages were followed by 69 speaking some German, 32 French, 20 Estonian, 15 Russian, 10 Spanish and 7 Italian. There were at most 3 speakers who reported proficiency in other languages, like Greek, Hungarian, Japanese, Karelian, or Norwegian.

3.3.1.4 Analysis

The results were downloaded from the server and the results of speakers who lived or had lived abroad were filtered out. The proportion of mixed responses was plotted for each pattern and for each stem vowel. The proportions of mixed responses by stem vowel were compared to the proportion of back vowels (resulting in mixed stems) after these stem vowels in the lexicon. Significance was determined using generalized linear mixed effect models with a binomial logit link function, using the `lme4` library in R (Bates et al. 2014). In the mixed effects models the response served as a dependent variable, and random intercepts and slopes were added by speaker and item.

3.3.1.5 Expected results

If the sublexical hypothesis turns out to be correct, it is expected that the proportion of choosing the harmonic or the mixed variant of the wug stem will depend significantly on the pattern the stem belongs to. The baseline for the proportion of mixed responses will be established by the control stimuli. The expectation is that patterns 1–4, which prefer mixedness based on the corpus study will show a higher proportion of mixed responses, while pattern 5, which prefers harmonicity based on the corpus study will show a lower proportion than the control stems. This is summarized in Table 3.24 below:

pattern	expected proportion of mixed responses
control	$p_{control}$
1–4	$> p_{control}$
5	$< p_{control}$

Table 3.24: Expected proportion of mixed responses based on the sublexical hypothesis

There might be some differences between the results and this expectation that would not question the validity of the sublexical hypothesis. If one or even two patterns deviate from the expected pattern but the other ones behave as expected, that might indicate that the given pattern presented some unexpected problems for the speakers but the overall results are as expected. Finally, the controls might show some unforeseen bias towards harmonicity or mixedness — so if patterns 1–4 show a mixed likelihood of $p_{control}$ while pattern 5 shows a much lower proportion, that would indicate that the controls themselves favor mixedness.

There are several other hypotheses that could explain a result not consistent with the expectations described above. First, if the plain unigram frequency of the final vowel is the most important information for speakers when deciding which form to prefer, then the proportion for mixed responses should be equal for all of the 5 patterns and the control, and it should match the overall frequency of -*a* final neutral stems in the language.

If speakers focus on the vowels instead of other sublexical characteristics, but they do not exclusively pay attention to the final vowel, they might rely mostly on vowel-to-vowel dependencies in their judgments. If this is the case, the proportion of mixed responses will depend on the quality of the first vowel: the best model for the participants' reaction will be the conditional probability of a final -*a* vowel, given the first vowel.

Table 3.25 below illustrates the two vowel-sensitive hypotheses. Unlike what was expected under the sublexical hypothesis, the proportion of mixed responses is expected to be the same across all patterns if vowels are all that Finnish speakers are considering when categorizing novel stems. If the most important variable is the unigram frequency of the last vowel, the expected probability is uniform across all words, while if the bigram frequency of the vowels is more important, then the quality of the stem vowel will be the only variable affecting the outcome.

pattern	expected proportion of mixed responses			
	unigram	bigram		
		after *e*	after *ii*	...
control	$p(a\#)$	$p(a\#\|e)$	$p(a\#\|ii)$	
1–4	$p(a\#)$	$p(a\#\|e)$	$p(a\#\|ii)$	...
5	$p(a\#)$	$p(a\#\|e)$	$p(a\#\|ii)$	

Table 3.25: Expected proportion of mixed responses based on a vowel-sensitive hypothesis

3.3.2 Results

Before analyzing the results of the experiment, the proportion of mixed stems in the lexicon ($p(a\#)$), and the proportion of mixed stems in the investigated subset of the lexicon, ie. in stems with a netural first vowel ($p(a\#|N)$) was determined to provide a simple Bayesian baseline mixed proportion for all of the more detailed results (Becker and Gouskova 2016 also use such conditional probabilities to build their sublexical model). It was established that mixedness prevails in the lexicon overall, as well as in the neutral-first subset of the lexicon:

(223) *mixed stems in the whole lexicon:* $p(a\#) = 75.29\%$

(224) *mixed stems amongst neutral-first disyllables:* $p(a\#|N) = 74.92\%$ (717 mixed out of 957 stems)

These proportions are very close to each other, and this small difference is not statistically significantly: mixedness is not less prevalent amongst CIC(C)A stems than what would be expected based on the whole lexicon (exact binomial two-sided $p = 0.79$).

3.3.2.1 Building the model

The first task of the analysis was to figure out which variables contributed significantly to speakers' categorization. To test this significance, generalized binomial mixed effect models were built with the (dis)harmonicity of the response as a dependent variable and the contributing factors: the `stem vowel` and sublexical `pattern` as independent variables. As the effects of patterns might be independent from each other, each pattern, including control stems as a "control" pattern were coded as independent binary variables. This leads to the intercept actually representing the grand mean, and the effect of each pattern on speakers' judgments can be independently assessed. Random slopes for pattern and stem vowel by speaker and random intercepts by speaker and by item were added to the model.

The stepwise selection process of finding the model that describes the data the best is shown on Figure 3.4 below. The first model tested had only an intercept as fixed effect, and neither the stem vowel, nor any patterns had been added (model A on Figure 3.4). The significance of the effects of `stem vowel` and `patterns` were tested by adding these factors to the intercept model one by one and performing model comparison: if the model with the added effect has a significantly higher likelihood, the factor is permanently added to the model and is considered significant.

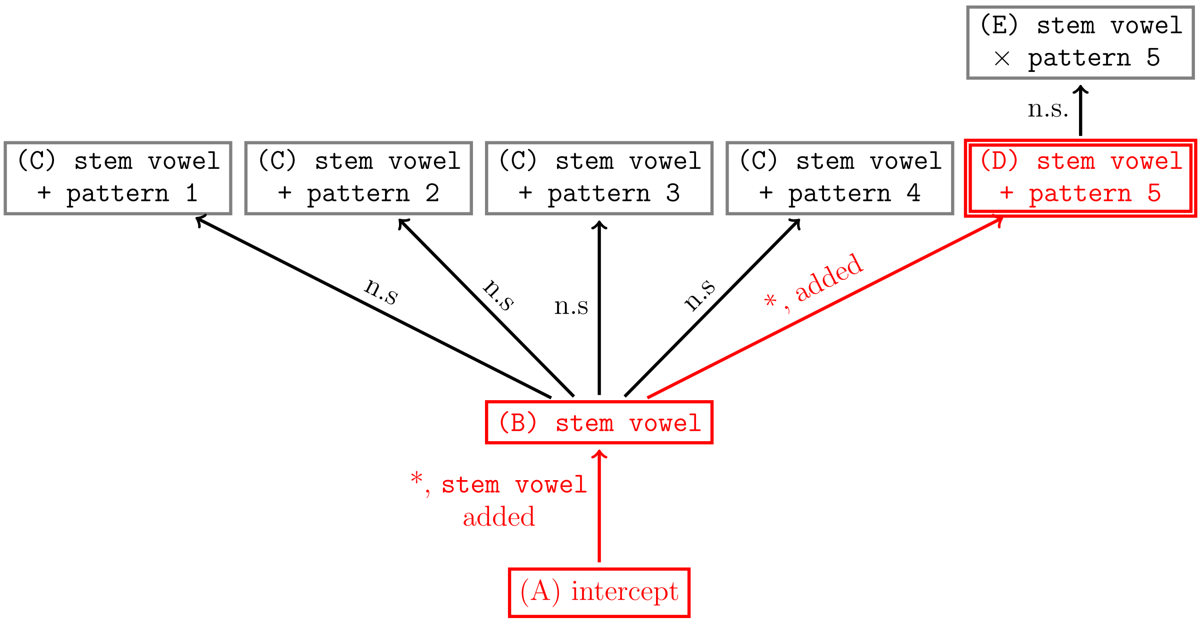


Figure 3.4: Selection of model for the mixed effects model for the stem vowel factor

First, `stem vowel` was added to the model (model B), as its effect was expected to be significant by both the sublexical hypothesis and the vowel only hypothesis. Its inclusion did indeed improve the model significantly ($\chi^2(5) = 24.51$, $p < 0.001$), so the `stem vowel` factor was included in subsequent models. The patterns were tested by adding them one by one to the `stem vowel` model and see whether doing this improves the model likelihood significantly. While this was not the case with patterns 1–4 (C models), including `pattern 5` (model D) as fixed effect did improve the model significantly, $\chi^2(1) = 3.99$, $p = 0.046$. Thus, the investigation of how the patterns affect the ratings is still relevant to the analysis of the results.

There was one more model to test: it had to be seen whether the interaction between the `stem vowel` and conforming to the phonotactic pattern of `pattern 5` was significant. This interaction term was added to the `stem vowel + pattern 5` model (yielding model E), and the likelihood ratio test was performed for the two models. The interaction turned out not to be significant ($\chi^2(5) = 0.6$, $p = 0.99$), therefore the interaction between the two effects does not have to be considered.

In the following subsections, the effects of the individual factors will be discussed. First, the results will be analyzed by stem vowel, then the analysis by sublexical pattern will be presented.

3.3.2.2 Results by stem vowel

First, the hypothesis that vowels in the stem matter in speakers' decisions about categorizing novel stems was tested. The overall proportion of mixed responses (62.54%) is lower than the proportion of mixed stems in the lexicon, but it is higher than 50%. This indicates that speakers are aware of the fact that mixed stems are preferred in the lexicon, but their performance undershoots this proportion.

The hypothesis that the stem vowel influences the decision of how to categorize novel stems more than the less natural sublexical phonotactic patterns seems to be confirmed by the analysis by stem vowels. Figure 3.5a presents the results analyzed by stem vowels, while Figure 3.5b shows the lexical statistics of Finnish CIC(C)A stems by stem vowel. Each bar represents the proportion mixed responses (black column) and harmonic responses (light blue column), with the total Ns (black in the bottom) and mixed and harmonic Ns (red in their respective columns) also shown. The horizontal lines represent the at chance level (red line), the proportion of mixed stems in the lexicon (green line) and the proportion of mixed stems in disyllables after a neutral vowel (blue line).

The correlation between the responses and the lexical statistics is the most visible with *i*-stems which prefer mixed stems both in the lexicon and in the nonce word preferences, while *ie*-stems prefer harmonicity the most in both the responses and in the lexicon. There seem to be a contradiction with stems that have *ee* in their first syllable, however, there are

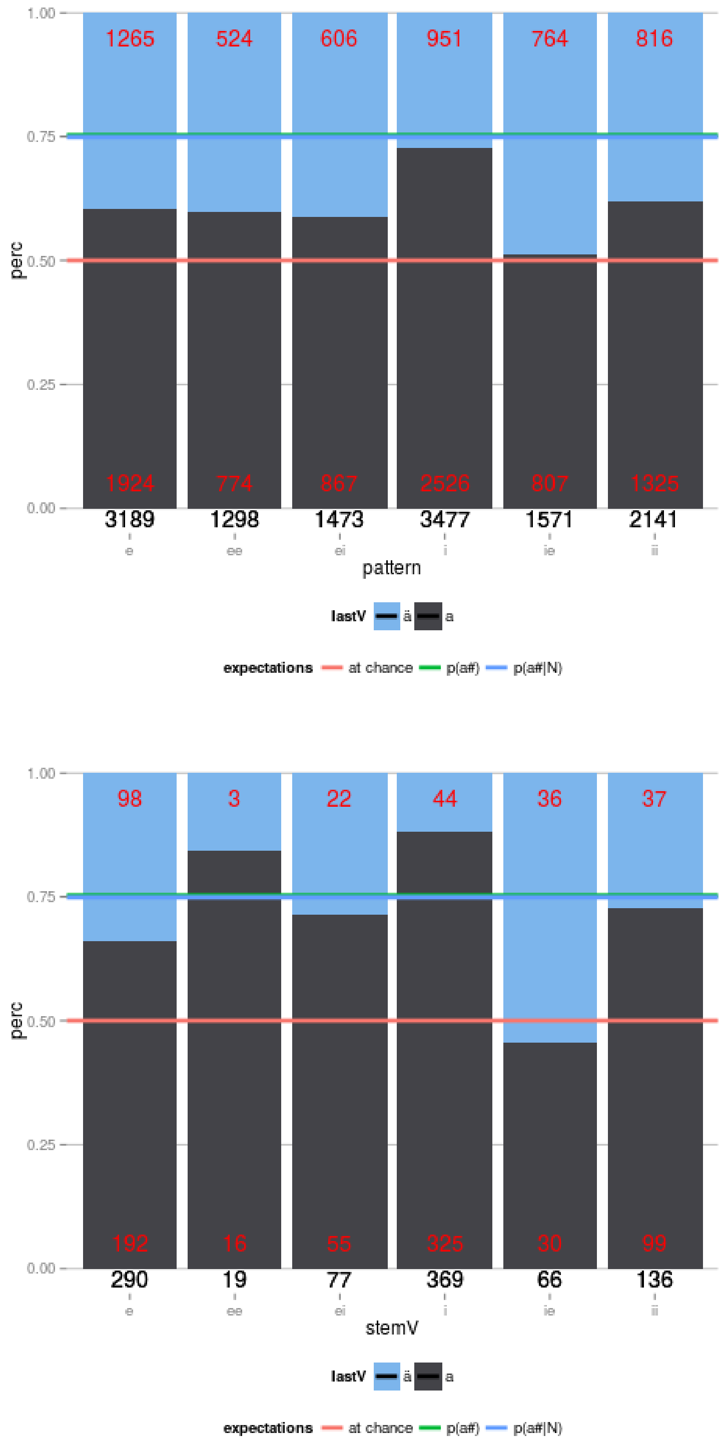


Figure 3.5: (top) Mixed responses by stem vowel (bottom) Proportion of mixed stems in the lexicon by stem vowel

only 19 such stems in the lexicon, so it seems that speakers do not have enough data to learn the distribution of mixed stems in these forms, so they aim towards the across the lexicon average. Such a small number might not lead speakers to identify a reliable generalization, as the reliabity of a rule to be extended to novel words is lower if the scope of the generalization is small (Albright 2002).

The significance of the effects of the stem vowel was tested by looking at the fixed effect coefficients in the final model (model D on Figure 3.4). As the stem vowel factor was coded with treatment coding, it had to have a baseline level, which was chosen to be [e]. The results are shown in Table 3.26. The value of the coefficient (β) and the probability of mixedness for the given pattern (the inverse logit function applied to the sum of the intercept and the coefficient; $\text{logit}^{-1}(\alpha + \beta)$) are shown, as well as p values provided by the `glmer` function.

stem vowel	β	$\text{logit}^{-1}(\alpha + \beta)$	p	sig.
ee	-0.088	0.64	0.76	n.s.
ei	0.068	0.675	0.82	n.s.
i	0.753	0.805	0.001	**
ie	-0.458	0.551	0.096	.
ii	0.064	0.674	0.799	n.s.

Table 3.26: Coefficients for stem vowels in the generalized mixed effects model built on stem vowels as independent variable.

These results confirm that a stem [i] vowel prefers mixed stems significantly more than [e]. The diphthong [ie] does not prefer mixedness significantly less with an $\alpha = 0.05$ significance level in this model, although in the exclusively `stem vowel` model without `pattern 5` being included (model B), the effect for [ie] is significant. The other vowels do not differ from the baseline [e] significantly under any circumstance.

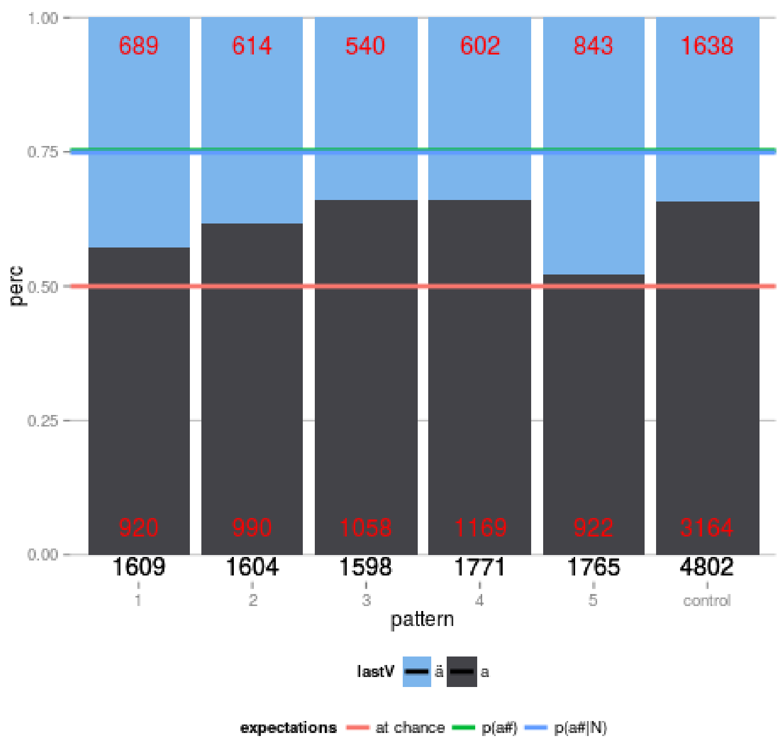


Figure 3.6: Mixed responses by pattern

3.3.2.3 Results by pattern

Figure 3.6 illustrates the results by pattern of the test word.

The results are consistent with the expectation based on the sublexical hypothesis. Pattern 5, which was expected to prefer harmonicity is indeed the only one where participants chose the mixed response at a lower level, around at chance (52.24%). The other patterns showed a pronounced preference for mixed stems. The control stems preferred harmonicity the least — the reason of which is not entirely clear. This might have been the cause of patterns 1–4 not being found contributing significantly to speakers' judgments.

Pattern 4, which was expected to prefer mixed stems although this difference was not significant actually has the second largest proportion of mixed responses (66%) amongst all patterns, trailing only pattern 3 (with 66.2%), and followed closely by control stems (65.89%), indicating that this pattern might be overlearned by the speakers. Therefore it seems that the lack of significance of the difference in the lexicon is only due to the lack of enough words with this pattern, yet speakers are still aware of this pattern. Patterns 1 and 2 also seem to prefer mixed stems too, but less so than the controls (57.18% and 61.72% respectively).

3.3.2.4 Further analysis of pattern 5

Pattern 5 shows a significant result, it is worthwile to take a closer look at how the stems belonging to this pattern behave. This is even more important after seeing in Section 3.2.5.7 that there are some subpatterns under pattern 5 in the lexicon: although fricative + *j* clusters preferred harmonicity overall, *sj* clusters overwhelmingly favored mixed stems in the lexicon.

When analyzing the different clusters separately for pattern 5, the hypothesis that speakers are aware of finer sublexical phonotactic regularities is further confirmed. The results show that Finnish speakers preferred harmonicity more when a stem had an *hj* cluster in it, than when the fricative was an *s* before the *j*. The results for pattern 5 separated by cluster, compared to the control results are seen in Figure 3.7. The results for the only *fj* cluster are also included, although these are based on the responses to only one stem.

The statistical significance of the contribution of different patterns was already discussed in Section 3.3.2.1: model comparison has shown that while patterns 1–4 do not seem contribute to speakers' judgments, including pattern 5 in the model describes these judgments significantly better. Indeed, in model D, which includes `stem vowel` and membership in `pattern 5` as well, the coefficient for the latter shows that stems that conform the shape

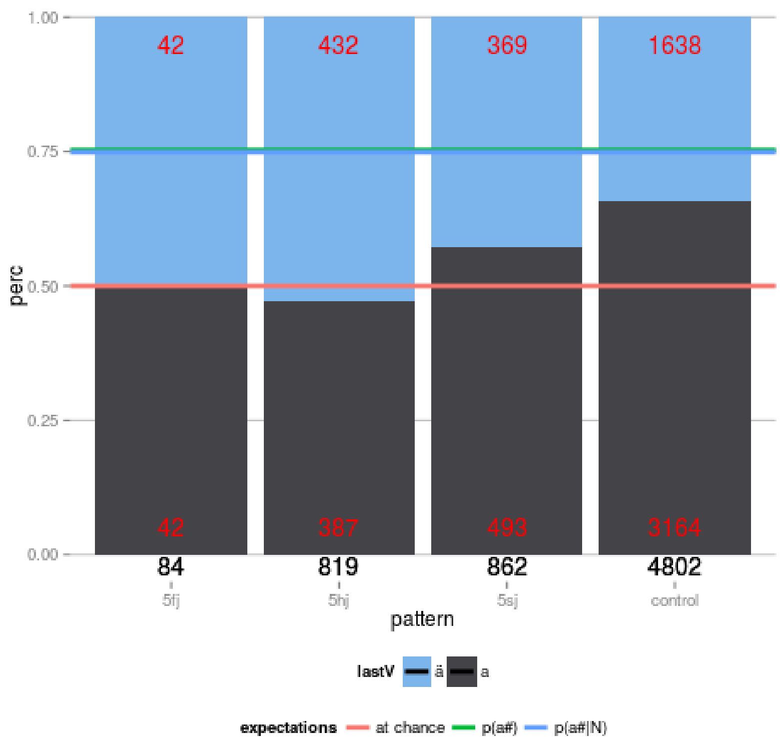


Figure 3.7: Mixed responses by exact consonant cluster in pattern 5

of pattern 5 are less likely to be categorized as mixed than other stems. This coefficient is $\beta = -0.519$, which means that compared to the overall probability of mixed judgments provided by the model intercept $\text{logit}^{-1}(\alpha) = 0.66$, the overall probability for pattern 5 stems is $\text{logit}^{-1}(\alpha + \beta(\text{pattern5})) = 0.54$, a difference that is significant in the model, $z = -2.012$, $p = 0.044$.

3.3.3 Discussion

The results of the experiment show that Finnish speakers are aware of sublexical phonotactic patterns underlying the distribution of mixed and harmonic stems in Finnish. Finnish speakers are aware of the influence a stem vowel has on the probability distribution of the backness feature of the following vowel. First of all, the probability of speakers choosing a mixed wug word like *kisja* over a harmonic one like *kisjæ* matches the overall proportion of mixed stems in the relevant sublexicon of neutral-first disyllables.

Speakers also pay attention to the quality of the stem vowel: since amongst CIC(C)A stems, a stem vowel [i] prefers mixedness the most in the lexicon, they prefer mixed novel stems with this vowel. Similarly, speakers do not prefer mixedness with [ie] as the stem vowel, despite the fact that they do prefer mixed stems with all other vowels, because the lexicon shows the exact same pattern.

The results about the vocalic patterns are in line with the bigram vowel-sensitive hypothesis sketched up in Table 3.25: the stem vowel matters to speakers when determining the sublexicon they are matching given a novel word.

The results of the wug study also support the hypothesis that certain non-vocalic phonotactic patterns can influence speakers' decisions when categorizing a novel stem as more likely to be harmonic or mixed. Although the effects of most consonantal patterns extracted from the lexicon were not seen to influence speakers' decisions significantly, speakers seemed to be aware of the effects of one pattern. This pattern states that fricative+/j/ clusters prefer harmonicity in general, contrary to the overall preference for mixed stems in the lexicon. Not only is this lexical preference matched by speakers in the wug experiment, speakers seemed to be also aware of the distribution of harmonic and mixed stems defined by subpatterns identified by the place of articulation of the fricative in the cluster as well.

The lack of sublexical learning of these consonantal patterns might be due to the low reliability of the predictions based on these patterns because of the low sample size available for the speakers in the lexicon (Albright 2002, Albright and Hayes 2003). Their effect might also be negligible because consonantal interactions with vowel harmony are not natural phonetically or widespread phonologically (Hayes et al. 2009), and therefore speakers might have a strong bias against learning such patterns (Becker et al. 2011) (though see Goldsmith and Riggle 2012 for informational theoretical arguments for the importance of consonantal interactions in Finnish vowel harmony). The effect of frequency matching was similarly low, though, for a stem-internal *ee*, because of low reliability reasons, and naturalness could not explain this underlearning result.

3.4 Conclusion

The results of the lexical and nonce word studies support the hypothesis argued for in this chapter: speakers are able to rely on lexical distributions of exceptional and non-exeptional stems when making judgments about the categorization of novel stems in vowel harmony patterns. More broadly, this supports the idea that these lexical distributions are the key to learning exceptional patterns in vowel harmony. In this sense, these results are in line with the analysis proposed in the Hungarian chapter, as sublexical learning seems to be the preferred way of describing exceptionality in vowel harmony again.

The importance of phonetic detail was shown to take a secondary role in exceptionality in the Hungarian chapter. Phonetic naturalness of the constraints characterizing sublexica seems to play a more important role, as speakers are more aware of more natural vowel-to-vowel patterns than consonantal patterns. The results are in line with Hayes et al. (2009): patterns that are phonetically not grounded (described by "unnatural" constraints, like pattern 5 above) are also learnable, but their effect is weaker than those that are more natural, like vocalic effects. The following subsections will discuss these points further.

3.4.1 Sublexical frequency matching

Following other studies, this paper also provides support for the idea that speakers are aware of the phonotactics and the frequency of certain subpatterns when categorizing stems. More specifically, speakers handle exceptional behavior by learning the phonotactic regularities about the exceptional sublexica.

Ernestus and Baayen (2003) have tested which past tense allomorph (-*te* or -*de*) Dutch speakers prefer, when presented with nonce verbs ending in obstruents, whose voicing is incompletely neutralized. Besides other factors, like the orthography of the presented wug stem, speakers were influenced by the suffix similar verbs take. The proportion their participants selected the voiced allomorph matched the proportion of verbs in the same group based on their form selecting that allomorph. Similarly, Albright and Hayes (2003) showed that when forming the past tense of an English nonce verb, speakers were influenced by verbs that had a similar phonological shape. These results were replicated for suffix choice for Hungarian transparent vowel stems (Hayes et al. 2009), and for non-predictable vowel-zero alternations in Russian stems (Gouskova and Becker 2013, Becker and Gouskova 2016). Linzen et al. (2013) have also shown that vowel-zero alternations in Russian prepositions depend non-categorically on the following word: several factors including the phonological form or the semantics of this word influences this alternation, and speakers do use these factors to decide whether to have a vowel or not in the form of the preposition.

These results mean that when encountering a novel item, speakers do not default to regularity necessarily, but they are willing to extend irregularity if the phonotactics of the new item indicate probable membership in an exceptional sublexicon. As Ernestus and Baayen (2003) show, and the other studies confirm, the proportion of irregular responses in a wug study matches the proportion of irregularity in the phonotactically similar sublexicon. This hypothesis of frequency matching is also confirmed in this study, as both the results for the vocalic patterns, and the subpatterns of the consonantal Pattern 5 behave in the predicted way.

3.4.2 Vocalic patterns are stronger than consonantal

Vowel-to-vowel effects were found to have a stronger effect than consonants in the stem on speakers' decisions. The reason for this is not entirely clear and several explanations are plausible at this stage of the research. Three hypotheses on this question are discussed below.

One such explanation might be that the consonantal patterns are simply not as strong in Finnish (and in Hungarian) as the vocalic patterns statistically. While there are only six relevant vowel phonemes (/e, ee, ei, i, ii, ie/) that occur in this sublexicon that Finnish speakers have to base on their analysis, there is no such restraint on the consonants occuring in this stem, while the number of possible intervocalic consonants, especially when taking into account consonant clusters is considerably more. This means that the sublexica divided by the stem vowel results in groups with still high enough membership (19–369 stems per stem vowel, 159.5 on average) to generalize, while the consonantal patterns selected in this chapter only describe 11 stems in the lexicon on average. Thus, it might be much harder for the speakers to establish a sublexicon that behaves in a certain way based on the consonants in the stem. Similar issues were encountered by Gouskova and Becker (2013) and Linzen et al. (2013) with monosyllabic yer words in their studies of Russian sublexica: there are just much fewer such stems in the lexicon for the speakers to generalize as widely on their lexical trends as with longer words, which provide a much better sample size.

Another explanation for the bigger effect of vocalic patterns is that speakers of a language with vowel harmony, like Hungarian or Finnish, are more biased to pay attention to the vowel quality when deciding about which allomorph of a suffix to use. While this attention generally focuses on the [αback/front] feature, or the [αround] feature in Hungarian, it is able to pick up the probabilistic effects of vowel height or vowel length as well. So when analyzing the phonotactic probabilities underlying the categorization of stems into harmonic and mixed, the first strategy they try is to see if looking at vocalic features (backness, roundness, height) helps them better predict the distributions. This strategy was useful for these speakers when learning about the basic mechanism for vowel harmony, so it will be naturally employed first. The effects of the consonants on the vowel, however, is not something that they used to base categorizing or allomorph selection choices on, so their role is considered either later in the process, or just on a lower level.

The third explanation for the preference to rely on vowel-to-vowel correspondences is based on underlying grammatical principles. As Hayes et al. (2009) propose, there might be a universal difference in the naturalness of effects on the vowel selection in a suffix. Patterns which are based on consonant-to-vowel interactions are in some way less natural than those, which are based on vowel-to-vowel interactions. This naturalness, which has a straightforward phonetic grounding, can be modeled in the grammar in several ways: one might argue for a different vocalic tier and a different consonantal tier in an autosegmental representation. In an Optimality Theory-based model, one could say that only few and select constraints exist, that penalize consonant-to-vowel cooccurences. In a more probabilistic constraint-based analysis, this argument would lead to positing a smaller weight to such constraint, so that their effects are much weaker than the constraints operating exclusively on the vocalic tier.

3.4.3 Strength of consonantal patterns in Finnish and Hungarian

The strength of the consonantal patterns in this study is still much lower than similar patterns found by Hayes et al. (2009) in Hungarian. Only one such pattern was found to have an effect on speakers based on the Finnish wug study, in Hayes et al. all four "unnatural" Hungarian consonantal patterns were found to be significant. There are several explanations for this difference.

First, while Hayes et al. looked for unnatural patterns based on the whole lexicon, the search space in this paper was limited to disyllabic stems of a CIC(C)A shape. Although the wug study also contained stems of this shape, speakers' judgments might have been based on a wider assessment of the lexicon — several stems of 3 or more syllables with multiple neutral vowels in them are left out of this study, as well as stems ending in a non-low vowel. If this is the case, a new lexical analysis on the full Finnish lexicon might find consonantal patterns that are reflected more by speakers' judgments, but this might not necessarily be the case. In order to find variation, the CIC(C)A shape cannot be extended too far, as more neutral vowels in the stem probably increases the likelihood of the word being categorized as harmonic. The reason for this is that in Hungarian, multiple neutral vowels in a row trigger harmonic affixation more than only one neutral vowel (this is the Count Effect of Hayes et al. 2009). A similar effect is also found by Ringen and Heinämäki (1999) in Finnish for the categorization of multisyllabic loanwords in Finnish — more neutral vowels at the end of the word like *arkkitehti* 'architect' were more likely to take front suffixes in the survey they did than stems like *analyysi* 'analysis'. Therefore, stems longer than the CIC(C)A shape in this study would be subject to a similar Count Effect, meaning that harmonicity would be much more favored by the addition of each vowel. Therefore, extending the structure to look for in the lexicon is not the most viable proposal.

Second, it might be the case that the Finnish patterns are weaker than the Hungarian ones based on the number of stems that belong to a given subpattern. While the patterns of Hayes et al. (2009) are represented by 41–168 stems in the lexicon, the Finnish patterns had to be extracted by phonotactic learning, as the exact membership of given patterns in the proper lexicon was lower, ranging 1–33 (11 on average, see Table 3.21). These patterns might have been too weak for the Finnish speakers to base their decisions on, while the more robust Hungarian patterns were learned and reproduced in the Hungarian study. While this problem might be a serious flaw of the design of this analysis, it might be worthwile to point out that the speakers seemed to follow the patterns discovered by the UCLA Phonotactic Learner instead of the raw numbers in the lexicon: while the learner implicated that Patterns 2 and 3 preferred mixed stems, and speakers did not prefer harmonicity for these patterns at all, the lexicon showed that there were no mixed stems belonging to these patterns in the lexicon. Similarly, while speakers did follow the Phonotactic Learner's analysis of Pattern 5, and its subpatterns, the lexicon provides minimal evidence (only 1 stem!) for this pattern. This might indicate that speakers do not only learn based on the direct evidence of stems with the same shape, but do some level of abstraction to understand the underlying phonotactics of the sublexica, like the Phonotactic Learner used in this study.

Finally, it might be the case that the structural difference between the Finnish and Hungarian exceptionality affected the results. Hungarian antiharmony is a clear case of morphological exceptionality: certain stems take different suffixes than what might regularly be expected. The Finnish case of exceptionality, as it has been discussed, is a static phonotactic irregularity: the stems by themselves do not behave exceptionally with regards to allomorph selection, their exceptionality is based on their non-harmonizing form. It might be the case, that this static irregularity is handled differently than the more morphologically consequential exceptionality in Hungarian.

3.4.4 Mixedness versus antiharmonicity as exceptional pattern

Connecting the results of this chapter with the Hungarian results might present some problems. Finnish mixed stems, unlike Hungarian antiharmony, is not a morphological exceptionality. This means, that there is no alternation or allomorph selection that is irregular for the exceptional stems, but their exceptionality is manifested only in a static phonotactic way. This raises the question whether these two kinds of exceptionality in vowel harmony can be handled together.

Certain models of morphophonology propose that allomorph selection and stable phonotactics might need different computations to learn. Hayes presents different computational models for these two tasks: the Minimal Generalization Learner (Albright and Hayes 2003) for learning alternations, the UCLA Phonotactic Learner (Hayes and Wilson 2008) for phonotactics. The Sublexical Learner model of Allen and Becker (2016) used in Becker and Gouskova (2016) also provides two grammars: one to select the sublexicon a stem is in ("gatekeeper" constraints) and one to evaluate the resulting surface form including base-derivative alignment ("grammar proper").

Treating the allomorph selection facet and the sublexical phonotactic generalization facet of vowel harmony entirely separate is not necessary though: under a classical Optimality Theory analysis, allomorphy selection happens in parallel to evaluation of phonotactics, using the same constraint set (Prince and Smolensky 1993/2004). It is clearly not impossible to handle morphological alternations and phonotactic regularities at least similarly. The Finnish case study and Hayes et al. (2009) show that there are commonalities between static and morphological exceptionality from vowel harmony: they both show preference of vocalic effects with the existence of weak consonantal effects in sublexical learning. The following chapter will present a computational model, which will discuss how these two facets of exceptionality in vowel harmony can be modeled in a parallel way.

Chapter 4

Learning the harmonicity of Finnish disyllables

This chapter addresses the learning problems introduced by the Finnish mixed stems and the Finnish disharmonicity pattern. These patterns, as seen in the previous chapter, are static generalizations on the grammaticality of stems in Finnish. There are three kinds of behaviors of vowels modeled by harmonicity classes: back vowels are Back, front rounded and front low vowels are Front and non-low, non-rounded front vowels count as Neutral. While Back and Neutral and Front and Neutral vowels are free to cooccur within the same stem, Back and Front vowels cannot be present within a stem at the same time.

The general problem discussed below is the learnability of the Finnish pattern. The learning of this pattern entails several tasks, like categorizing vowels and evaluating the grammaticality of their combinations. While approaching these tasks, different kinds of biases will be considered that the learner can use to help it discover the Finnish pattern. The usefulness of these biases can also shed some light on what prior knowledge helps actual Finnish speakers in acquiring vowel harmony.

A more specific problem discussed in this chapter is the relation between phonotactic learning and learning allomorphy. The Hungarian pattern discussed in Chapter 2 is about exceptionality in the selection of allomorph for a given stem: for example, the stem /hiːd/ 'bridge' selects back suffixes despite having a phonetically back vowel. However, the Finnish pattern in Chapter 3 deals with the learnability of static phonotactic generalizations on the acceptability of different stem forms — so the exceptionality of a stem like /lyyra/ 'lyre' is evident only by looking at the phonotactic patterns of the language and finding that front rounded vowels and back vowels are very infrequently found in the same stem. Most vowel harmony systems include both facets of exceptionality, having both alternating affixes and within-stem generalizations. The relationship between these two sides is not entirely clear: is learning one facet helpful to learn the other one? Or is the relationship only coincidental synchronically and the relationship between the two might be just the product of diachronic processes? This chapter attempts to answer these questions by exploring how learning one facet might be useful for learning the other one.

The problems sketched up above will be addressed using two existing learners: the UCLA Phonotactic Learner (Hayes and Wilson 2008) which is geared towards static phonotactic generalizations and the Sublexical Learner of Allen and Becker (2015), which focuses on allomorphy. The UCLA Phonotactic Learner is an inductive learner that is capable of finding generalizations in the form of constraints based on the training data, and assigns weights to these constraints so that the resulting maximum entropy grammar maximizes the probability of the observed data. The Sublexical Learner, on the other hand, works with constraints defined by hand, however, it is capable of defining the alternations between pairs of forms, and build two grammars using the pre-assigned constraint set. First, the "gatekeeper grammar" decides which sublexicon a form belongs to, which determines what kind of alternation that form undergoes, and then the "grammar proper" that evaluates the final form.

The training data to these learners will be focusing on a subset of the Finnish lexicon: the more frequent half of disyllablic stems in the webcorpus used in the previous chapter. This decision will be justified first with an exploratory analysis of the harmonicity patterns of the whole Finnish lexicon. To model the different learning problems that static phonotactic generalizations and allomorphy pose, the UCLA Learner will be run on two different sets on training data. The analysis and the frequent disyllabic **stems** can help to model the static generalizations that Finnish speakers are aware of. To model how this knowledge can be expanded to allomorphy, the learner will also be tested on full (possibly suffixed) **word forms**, whose stems are in the frequent disyllable set used before.

The results of the learning algorithms will show that providing a static phonotactic learner with all (suffixed and unsuffixed) word forms improves its performance. To perform better, some deeper understanding of the phonology of vowel harmony (the actual natural classes that make up the harmonic types vowel harmony cares about) is needed, which the static learner cannot discover by itself. The learner trained on alternations, however, is able to come up with these classes, provided with some low-level biases on vowel-to-vowel dependencies. The results suggest, therefore, that the two facets of vowel harmony need to be learned together, as their inputs do rely on each other.

First, the lexical statistics of harmonicity in Finnish, both for stems and word forms, will be discussed in Section 4.1. This discussion will lead to focusing in on the frequent disyllable set, that will be analyzed by the UCLA Phonotactic Learner in Section 4.2. This will be followed by the experiment of learning the generalizations about the vowel harmony alternations in Finnish using the Sublexical Learner in Section 4.3. Section 4.4 will summarize and discuss the results.

4.1 Lexical statistics of Finnish harmonicity

In order to assess how different learners are able to model Finnish speakers' acquisition of vowel harmony, it is important to get a picture of the relevant facts of the Finnish lexicon. In this section, the distribution of harmonic and non-harmonic vowel combinations in the Finnish lexicon will be summarized to see what kinds of generalizations can be expected to be discovered by the learners — both by Finnish speakers acquiring the language, and by the learning models used in this chapter.

The corpus used in this exploratory study is, the same as the one used in the last chapter (3.2.3). This dataset was created from the webcorpus of Zséder et al. (2012), cleaned and stemmed using the *Omorfi* tool (Pirinen 2014). The word forms were one step behind the corpus used from stem, as they were just collected before the stemming of the corpus.

There is a crucial difference between what is expected to be learned about affixed words and about plain stems. Stems in Finnish tend to be harmonic, but there is a fair amount of mixed stems, and quite a few which are disharmonic as they mix non-neutral front vowels (*æ, ø, y*) with back vowels. A summary of the type and token frequency of stems in Finnish as calculated from the webcorpus is seen below in Table 4.1, where all of the stems found in the corpus are analyzed.

harmonicity	type frequency		token frequency		grammatical?
back only	17 791	(16.1%)	202 908 225	(33.9%)	✓
neutral only	3 963	(3.6%)	58 645 925	(9.8%)	✓
front only	4 402	(4.0%)	42 827 920	(7.2%)	✓
back-neutral	60 787	(55.1%)	215 554 101	(36.0%)	✓
front-neutral	14 360	(13.0%)	73 883 891	(12.4%)	✓
disharmonic	9 041	(8.2%)	4 215 940	(0.7%)	×

Table 4.1: Summary of harmonicity of stems in the lexicon

Stems that have one type of harmonicity throughout (back, front or neutral) are considered harmonic, while ones that have one back and one front (non-neutral, ie. rounded) vowel in them are considered disharmonic. For clarity, back-neutral and front-neutral mixed stems are categorized separately. It is visible that disharmonicity is indeed quite rare, especially by looking at the token frequency: most disharmonic stems are infrequent loanwords. On the other hand the grammatical forms comprise 91.8% of types and 99.3% of tokens in the corpus.

The lexical statistics show a more complicated pattern, however, than just an overall dispreference of disharmonicity. A pure vowel harmony system with no transparency would require that each word contain vowels with the same value for the phonetic feature driving vowel harmony. This is, however, only characteristic of 36.7% of types and 63.3% of tokens in the corpus, as both disharmonic or back-neutral stems contain at least one [-back] and at least one [+back] vowel. Of these two types, back-neutral stems are grammatical according to linguistic descriptions, as a large number of native words have this pattern. The absolute majority of types and the relative majority of tokens contains both back and neutral (front non-low unrounded) vowels, which indicates that understanding the phonotactic distribution of transparent/neutral vowels is a very important facet of the Finnish pattern.

The statistics also show that the harmonically front vowels /æ/, /ø/ and /y/ are relatively infrequent in the lexicon. This is evidenced by the low frequency of front harmonic and front-neutral stems in the lexicon: only 17.0% of stem types contain one or more front vowels. Different strategies used by a learner make different predictions on how it would rate stems with Front vowels. If the learner generalizes on the phonetically based behavior of vowel harmony, it will rate such stems as equally grammatical as back only or neutral only stems. However, if the learner relies more on the relative frequencies of vowels and vowel bigrams, it might rate stems with Front vowels worse than other grammatical stems, as the scarcity of these vowels would influence its decisions.

Table 4.2 below shows the results for word forms. It is noticeable that disharmonicity is more widespread with word forms. This makes sense: loanwords, which are most likely to be disharmonic, are also suffixed regularly as native words, yielding many disharmonic tokens in the corpus (eg. *lyyra* 'lyra', *lyyrat* 'lyras', *lyyrassa* 'lyra.INESS', *lyyransa* 'his/her lyra', etc.). The source of the larger proportion of back-neutral forms is not entirely clear, but a quick look at the list back-neutral word forms hints that this might be due to the effect of a large number of derivational suffixes with neutral vowels.

harmonicity	type frequency		token frequency		grammatical?
back only	361 243	(5.3%)	147 350 437	(28.2%)	✓
neutral only	46 603	(0.7%)	37 501 395	(7.2%)	✓
front only	38 676	(0.6%)	26 742 170	(5.1%)	✓
back-neutral	4 036 885	(58.8%)	208 744 698	(39.9%)	✓
front-neutral	525 015	(7.6%)	78 447 993	(15.0%)	✓
disharmonic	1 858 763	(27.1%)	24 480 741	(4.7%)	×

Table 4.2: Summary of harmonicity of word forms in the lexicon

To see the distribution of the different types of harmonic behavior in Finnish suffixes, a short summary will be presented in Section 4.1.1. Two subanalyses have been run after this, in order to figure out whether the hypothesis that loanwords and other less frequent words like certain slang terms are responsible for the relatively high proportion of disharmonic and back-neutral stems. In the following subsections these will be presented: the lexical statistics on disyllables in Section 4.1.2 and the analysis of stems by frequency quantiles in Section 4.1.3.

4.1.1 Summary of the harmonic behavior of Finnish suffixes

Suffixes in Finnish that contain at least a vowel regularly alternate according to the harmonicity type of the stem, if the vowel in the suffix is not Neutral. Each harmonically Back vowel is paired up with a Front vowel: *a* with *æ*, *o* with *ø* and *u* with *y*. Neutral vowels do not participate, but they are transparent, so the nominal derivational suffix *-io* does alternate with *-iø*. There are some suffixes (as listed in Hakulinen et al. 2004), generally loan morphemes, that do not alternate according to vowel harmony, like *-ssa* 'feminine suffix', as no *-*ssæ* allomorph exists:

(225) *prinssi* 'prince' ∼ *prinsessa*, not **prinsessæ*

In the following analysis, suffixes will be categorized into alternating (containing a Back ∼ Front alternating vowel), regular non-alternating (containing only Neutral vowels) and irregular non-participating (like *-ssa* or *-fobia*). Of these, the first two behave grammatically for the purposes of vowel harmony. As the number of suffixes in Finnish is large, the nominal inflectional paradigm, the verbal inflectional paradigm, derivative suffixes and final particles are handled separately. The source for the list of suffixes is the Finnish grammar of Hakulinen et al. (2004), omitting the suffixes indicated to be part of a slang or a dialectal variety.

The data are summarized on Table 4.3 and visually represented on Figure 4.1. The data show that the majority (although a small majority only) of Finnish suffixes alternates according to vowel harmony (145 out of a total of 247, 58.70%). There is still a noticeable amount of non-alternating suffixes both in inflection and in derivation, which means that the more suffixes a stem takes on, the more likely it is for the word form to include a non-alternating suffix, therefore, more likely to be disharmonic, front-neutral or back-neutral.

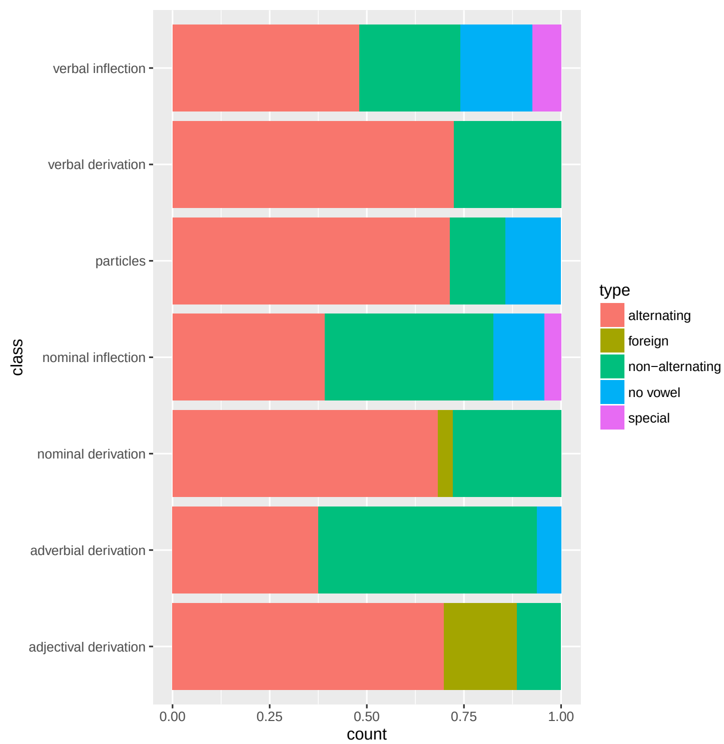


Figure 4.1: Summary of alternation types by suffix classes

morphology	alternating	non-alternating	foreign	no vowel/special
nominal inflection	9	10	0	4
verbal inflection	13	7	0	7
nominal derivation	54	22	3	0
adjectival derivation	37	6	10	0
adverbial derivation	6	9	0	1
verbal derivation	21	8	0	0
discourse particles	5	1	0	1

Table 4.3: Suffix harmonicity type by morphological class

Table 4.4 below shows the frequency of each vowel in Finnish suffixes. It can be seen again, that while many vowels (including *all* rounded vowels) in the suffixes in Finnish do alternate, the number of non-alternating vowels (the Neutral vowels /e/ and /i/) is quite high: the majority of the vowels found in suffixes (54.28%) do not alternate.

vowel	occurences	percentage
a ∼ æ	88	25.96%
o ∼ ø	34	10.03%
u ∼ y	33	9.73%
e	76	22.42%
i	101	29.79%
non-alternating a	7	2.06%
sum	339	100%

Table 4.4: Frequency of different vowels in Finnish suffixes

4.1.2 Analysis of disyllables

To make the analysis easier to handle, limiting it to disyllables makes sense. As disyllables were particularly of interest in the previous section, the behavior of these shorter words is worthy of being focused on. Another reason to look at words with two syllables is that most loanwords will be excluded from this set, and native Finnish words will most likely predominate. The analysis is also simpler with disyllables as there only two vowels have to be taken into account when characterizing the generalizations.

The table of the harmonicity of disyllabic stems below shows that disharmonicity is indeed even more infrequent in native words, but there is still a bit more words like *pake* (neutral vowel following a back one) than back only stems like *paka*:

harmonicity	type frequency		token frequency		grammatical?
back only	5 637	(35.2%)	103 368 691	(34.9%)	✓
neutral only	1 608	(10.0%)	27 051 107	(9.1%)	✓
front only	1 487	(9.3%)	24 786 973	(8.4%)	✓
back-neutral	5 789	(36.1%)	102 504 529	(34.6%)	✓
front-neutral	1 404	(8.8%)	37 861 268	(12.8%)	✓
disharmonic	94	(0.6%)	655 122	(0.2%)	×

Table 4.5: Summary of harmonicity of disyllabic stems in the lexicon

The more harmonic behavior of disyllables, as compared to all stems is also illustrated below in Figure 4.2:

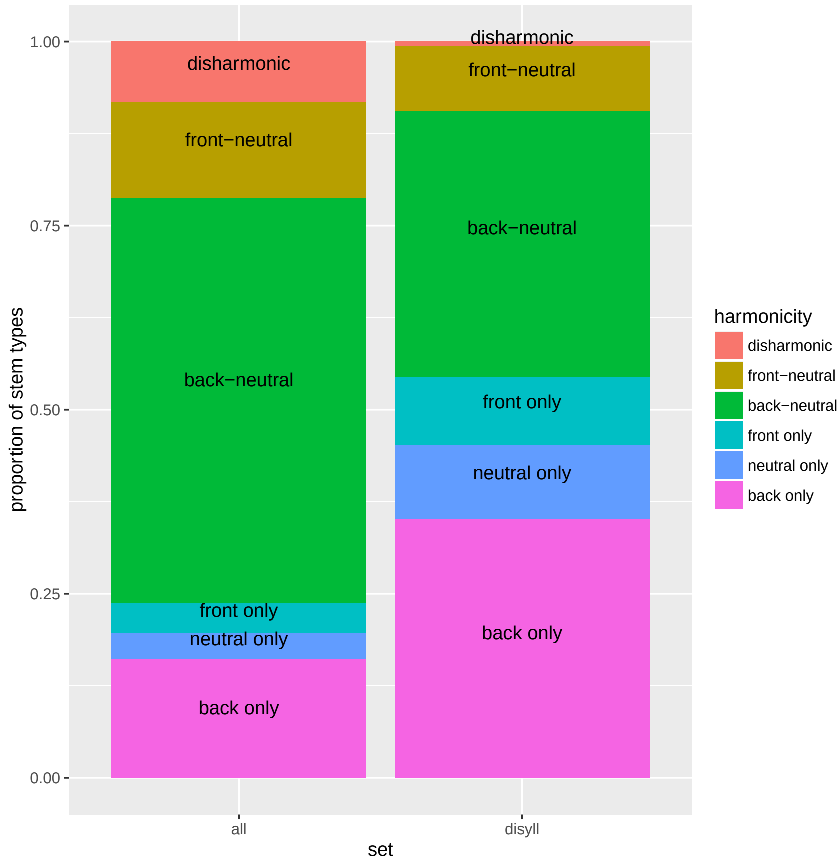


Figure 4.2: Comparison of harmonicity in all stems and in disyllabic stems only

4.1.3 Analysis by frequency of the stem

In order to see whether loanwords, ad hoc borrowings or other infrequent forms are indeed responsible for the high percentage of ungrammatical forms, the corpus was split to 10 bins based on the stem frequency. The least frequent stems are expected to be more likely to be loanwords and therefore more likely to be disharmonic, and the more frequent stems are expected to be more likely to be harmonic.

Figure 4.3 below illustrates the percentage of the different harmonic qualities of the stems in the 10 frequency bins. It is apparent that harmonicity is more and more frequent as the stems become more frequent. The proportion of disharmonic stems drops from 12.9% in the least frequent quantile to merely 1.9% in the most frequent one. The proportion of the grammatical but phonetically still not harmonic Back-Neutral stem type also falls as stems get more frequent, and conversely, the proportion of fully harmonic BB and NN stem types gets higher when the stems are more frequent.

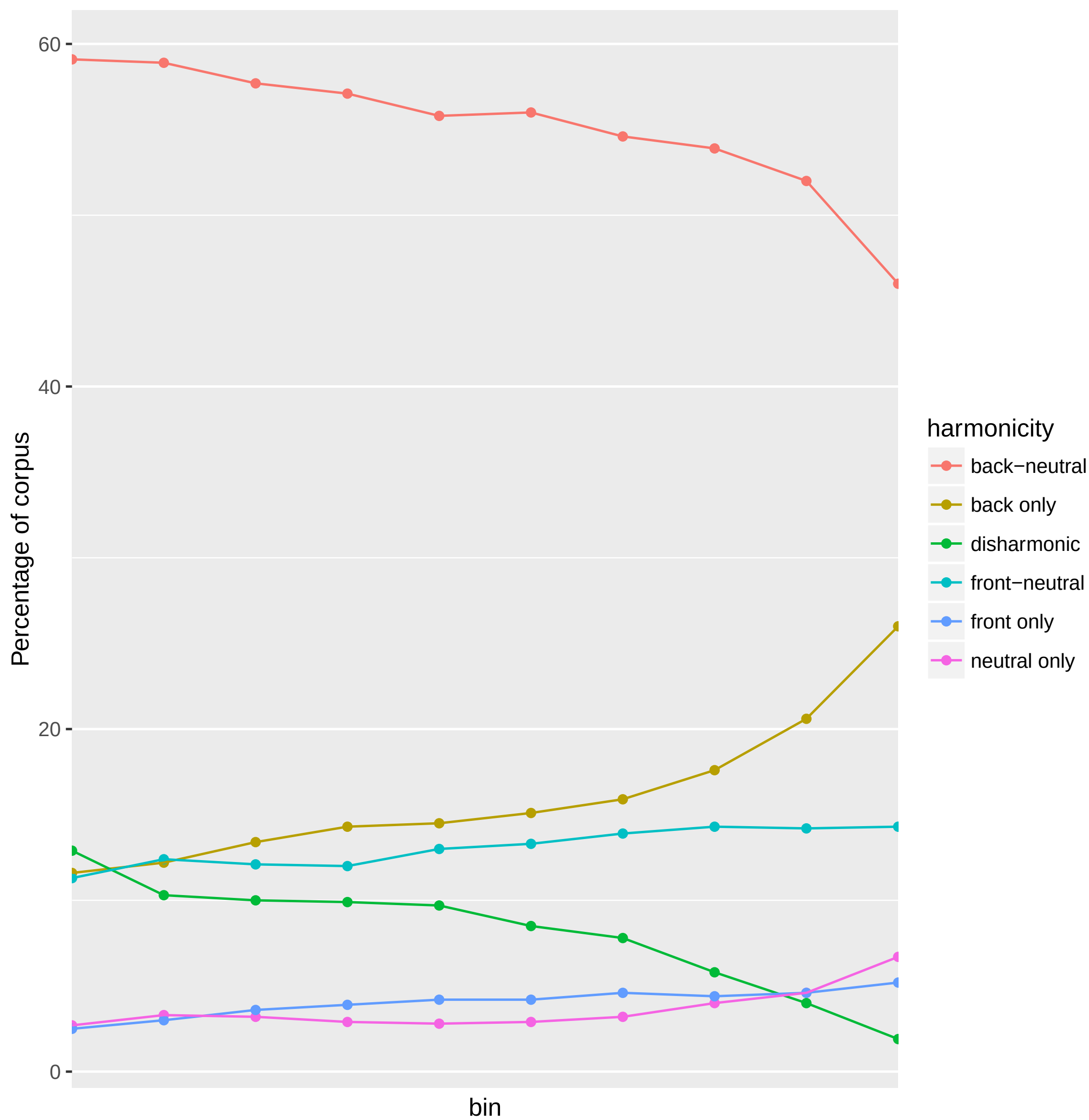


Figure 4.3: Harmonicity of stems by frequency in percentiles made up of 10%

4.1.4 The set of frequent disyllables to test learners on

The analysis in the following section was not run on the whole corpus, the learner was tested on frequent disyllables to see if it finds the patterns on a well defined subset of the lexicon that is most likely to be in everyday usage. As seen in the previous subsections (4.1.2–4.1.3), focusing on disyllables and on frequent stems helps the analysis close in on stems that are more likely be harmonic.

First, disyllabic stems were filtered out from the full corpus. The resulting number of stems depended on the phonemic inventory used in the analysis. The number of stems with all diphthongs described as such by Hakulinen et al. (2004) counted as being the nucleus of one syllable, there was a total of 16157. To get an even more manageable set of stems with more everyday usage, this set was split in half by frequency, yielding 8081 stems with a minimum token frequency of 352. The summary of harmonicities for the more frequent half of the disyllabic stems is summarized below in Table 4.6.

harmonicity	type frequency		token frequency		grammatical?
back only	3 003	(37.2%)	104 235 307	(34.7%)	✓
neutral only	792	(9.8%)	28 266 514	(9.4%)	✓
front only	686	(8.5%)	25 016 706	(8.3%)	✓
back-neutral	2 853	(35.3%)	103 674 010	(34.5%)	✓
front-neutral	706	(8.7%)	38 796 261	(12.9%)	✓
disharmonic	41	(0.5%)	662 911	(0.2%)	×

Table 4.6: Summary of harmonicity of the frequent disyllables

This table shows that this subset of the corpus is indeed more regular and more harmonic than the full lexicon. Most importantly, the frequency of ungrammatical stems in this subset of the corpus is much lower (0.5%) than in the full corpus (8.2%). The percentage of back-neutral stems is still high at 35.3%, but this reflects the high frequency of such stems in the lexicon. The majority (55.5%) of the stems in this more focused subset is completely harmonic, with both vowels having the same harmonicity (back only, front only and neutral only stems). These facts suggest that using the frequent disyllable stems in the analysis exposes the learner to stems more typical of the Finnish stems a human learner would encounter in acquisition.

4.1.5 Expected frequency of disharmonicity

Before looking at how the UCLA Phonotactic Learner deals with the Finnish data, one more analysis was attempted to account for the results above. The actual proportion of nonharmonic stems as shown in Tables 4.1—4.6 might also seem exaggerated due to the larger number of mixed stems if multiple vowels are randomly drawn independently from each other to form stems.

The individual probability of Front, Back, and Neutral vowels in the lexicon, is summarized below:

(226) $p(F) = 0.121$

(227) $p(B) = 0.494$

(228) $p(N) = 0.385$

It is already visible that Front (that is, front rounded) vowels are much less frequent in the language than the other two categories. To produce the expected proportions of different harmonicities for disyllabic stems, refined calculations are used: the probability of the different categories are calculated separately by the first and the second syllable. As illustrated below, Front vowels are even less frequent in the second syllable:

	first syllable	second syllable
Front	0.143	0.102
Back	0.546	0.489
Neutral	0.312	0.409

Table 4.7: Probabilities of different harmonic categories in the first two syllables of stems

Based on these numbers, it is possible to predict what the proportion of different harmonicities is if the vowels were independent from each other. This is summarized below:

	Expected	Observed	Expected %	Observed %	O/E
back only	4280	5637	26.7	35.2	1.32
neutral only	2040	1608	12.7	10.0	0.79
front only	232	1487	1.4	9.3	6.41
back-neutral	6018	5789	37.6	36.1	0.96
front-neutral	1440	1404	9.0	8.8	0.97
disharmonic	2006	94	12.5	0.6	0.05

Table 4.8: Predicted and observed proportion of harmonicities in disyllables

This table shows that the observed frequencies of most harmonic patterns can be quite accurately predicted without positing dependence between the two stem vowels in a disyllable. Back only stems are a little bit more frequent and neutral only stems are a little bit less frequent than what would be expected if each harmonicity were to be sampled independently from each other. The proportion of front-neutral stems and, crucially, the high frequency of back-neutral stems are very close to what is expected based on the individual category-level frequencies.

The remaining two patterns, however, are able to show that there is a need for dependence between the harmonicities of two vowels in the disyllable — that is, there is a need for modeling vowel harmony to explain the patterns. Harmonic stems with front rounded vowels only are much more (more than 6 times more) frequent than what would be expected if vowels were independent from each other. On the other side of the coin, disharmonic stems (stems with back and front rounded vowels) are very scarce: one would expect 20 times more such disyllables without vowel harmony. This also means that the disharmonic category is indeed the only ungrammatical in the Finnish lexicon, therefore it is the only category that the computational models should strive to exclude from the grammar.

4.2 Learning disyllables with the UCLA Phonotactic Learner

4.2.1 Setup

This section will describe the input and the parameter settings of the UCLA Phonotactic Learner's learning algorithm. While setting up the learner, there were 3 points of decision where more than one option was taken into consideration. In these cases, opting for a different choice could reveal different facets of the pattern, and the comparison of the results of the two learning processes could also be helpful in understanding what assumptions are necessary in learning vowel harmony.

These three points of decision occured when selecting the training set, the feature system used in the analysis and when choosing whether there were prior constraints used in training. Table 4.9 summarizes the options described by these conditions and the subsections below discuss the setup in detail.

condition	option 1	option 2
training data condition	stems	word forms
feature system condition	phonetic features only	harmonic features added
prior constraints condition	discovery learning	assisted learning

Table 4.9: Summary of the 3 conditions used in the training process

In choosing the training set, comparing the analysis of stems and the analysis of word forms helped to illustrate the difference between learning static phonotactic regularities and allomorphy, as discussed below in Section 4.2.1.1.

Choosing two different feature systems (described in Section 4.2.1.3), one with standard phonetic features and one with the harmonic features Back, Neutral and Front added could reveal how important it is to learn prior assumptions about the phonological system of Finnish vowels when learning the vowel harmony pattern. A system relying on phonetic features might struggle with learning that the low front vowel patterns with the front rounded vowels; on the other hand, the system using the added harmonic features might overfit the data and learn an overall dispreference for the infrequent front vowels.

Finally, the learner was run with two set of preset constraints (Section 4.2.1.4). First, it was trained with a set of preselected constraints that evaluate different harmonic patterns (assisted learning). This grammar did not attempt to find more constraints based on the phonotactic generalizations about the data, as its task was solely to assign weights to the constraints describing vowel harmony selected by hand. The result of this grammar provided a baseline for the second training, where the learner was instructed to find constraints from scratch (discovery learning). The results of this learning could then be compared with the assisted learning results to see how well the Phonotactic Learner fares in discovering and isolating the patterns needed to describe the generalizations on the Finnish data.

4.2.1.1 Training data

There were two different analyses using two training corpora for two different learning tasks. To model how Finnish speakers learn static phonotactic generalizations, the UCLA Phonotactic Learner was trained on the list of frequent disyllablic **stems** set up as described in Section 4.1.4.

To see how Finnish speakers are able to generalize their static phonotactic way of understanding vowel harmony to using harmonicity to determine allomorphy of suffixes all **word forms** (both suffixed and unsuffixed) were also tested. The training data for this analysis consisted of all word forms in the corpus whose stems were in the frequent disyllable list used in the stems analysis. This control over the input ensured that the results of the word form and the stem analyses could be compared with each other without any intervening additional effects to account for.

4.2.1.2 Projected tiers

The learner was set up to work not only on the basic segmental tier, but to learn generalizations from a tier on which exclusively vowels were projected. Since the last chapter showed that consonantal patterns are much weaker than vocalic patterns for Finnish speakers, analyzing vowel-to-vowel correspondences were given priority here, justifying the decision to include the vowel tier.

In their information theoretical study of Finnish vowel harmony, Goldsmith and Riggle (2012) have found that projecting a vocalic tier actually gives worse results than a plain bigram model. A reason they mention is that consonantal effects might be more important than what a tiered representation might imply. Their study, however, unlike this chapter, does not focus on the learning of the phenomenon, but it evaluates models based on information theoretical tools, which might give a very different outlook to how a vowel harmony pattern should be analyzed.

Another reason for separating the vocalic tier was that with (at most) trigram constraints generalizations with more than 2 vowels were impossible to learn. With no separate vocalic tier, learning vowel to vowel correspondences turned out to be hard for the learner as more local patterns dominated the resulting grammars.

4.2.1.3 Feature system

All analyses were run using two slightly different feature systems. Tables 4.10 and 4.11 below illustrate the two systems employed in the analyses.

	[-back]		[+back]	
	[+rounded]	[-rounded]	[+rounded]	[-rounded]
[+high, -low]	y	i	u	
[-high, -low]	ø	e	o	
[-high, +low]		æ		a

Table 4.10: The phonetic features only analysis of the Finnish vowel system

The **phonetic features only** system (Table 4.10) uses fairly standard binary vocalic features: this system employs all of the features that were used in the previous chapter. This means that vocalic features are defined phonetically: for example, [±back] determines if the vowel is back or front phonetically (phonetically front vowels including both harmonically Front and Neutral vowels), [±low] determines if it is low, etc. This analysis models the knowledge of a speaker with exclusively phonetic knowledge about the vowel system. For a learner using this feature system, learning vowel harmony involves the discovery that non-low front vowels behave differently both in their distributions in stems and in allomorphic alternations than the other front vowels: they can cooccur with back vowels and are transparent, while the other [-back] vowels cannot be grammatically mixed with back vowels within the stem, alternate with back allomorphs in suffixes and are generally not transparent. This involves the problem of working out generalizations over a non-natural class subset of the phonemes that can only be described with disjunctive specifications ([-back, +low] *or* [-back, +round]).

There is a formal problem that using such a feature system for describing the Finnish pattern introduces. There is no way to refer to the set of Front vowels (the low front vowel and the rounded front vowels) in a single conjunctive feature bundle description. One solution is that constraints against different vocalic patterns refer to a disjunctive set of features, as in (229) below:

(229) $*[+\textsc{back}]\begin{bmatrix} \textsc{-back} \\ \textsc{+rounded} \text{ or } \textsc{+low} \end{bmatrix}$, ruling out *a…ø* and *a…æ*

The problem with this solution is that such feature bundles are not permissible in most theories of grammar, as they do not describe natural classes. As an alternative, each intuitive generalization containing reference to Front vowels has to be described by two different constraints, such as in (230–231), a solution that misses the generalization that these vowels behave in the same way.

(230) $*[+\textsc{back}]\begin{bmatrix} \textsc{-back} \\ \textsc{+rounded} \end{bmatrix}$, ruling out *a…ø*

(231) $*[+\textsc{back}]\begin{bmatrix}\textsc{-back}\\ +\textsc{low}\end{bmatrix}$, ruling out *a*. . .*æ*

	[+Front]	[+Neutral]	[+Back]
[+high, -low]	y	i	u
[-high, -low]	ø	e	o
[-high, +low]	æ		a

Table 4.11: The harmonic features added analysis of the Finnish vowel system

For the **harmonic features added** system (Table 4.11), the Back, Front and Neutral harmonic categories are added as binary features, in place of the [±back] and the [±rounded] phonetic features. Other vocalic features are not altered, so the [±low] binary feature still encode that particular aspect of the vowel's height. This feature system models a learner that has a prior knowledge about how Finnish vowels are categorized phonologically in vowel harmony, or one that uses phonological information to set up its feature system. Using these harmonicity classes might turn out to be necessary to describe the pattern, but how a learner can develop such a feature system without the distributional information provided by the vowel harmony pattern is not entirely clear. It might be the case that learning the feature system simultaneously with the phonotactics is the key to the best performance for a learner.

The reason for using both of these systems is that they both represent a different hypothesis on what kind of representation the learner needs to learn vowel harmony. The phonetic features only system models a learner who has only the phonetic facts to rely on when learning the vowel harmony pattern. This learner has the task to figure out that non-low front vowels behave very differently than other front vowels both in the phonotactics of the stem and in the allophonic alternation of suffixes. On the other hand, the harmonic features added system models a learner that has some prior knowledge on the phonology of Finnish thats help it discover the underlying mechanics of vowel harmony. Formalizing the constraints that describe the intuitive generalizations underlying vowel harmony is more straightforward in this system, as the constraint in (232) suffices in describing disharmonic disyllabic stems with back vowels in their first syllable.

(232) *[+BACK][+FRONT], ruling out *a...ø* and *a...æ*, but not *a...e*

4.2.1.4 Prior constraints

The learning process was done under two conditions based on whether the learner was given any prior constraints to work with or not. The different outputs based on these two conditions can be compared in order to understand how well the learner is able to discover constraints needed to model vowel harmony.

In the **assisted learning** condition, the learner was trained on the corpus by instructing it to weigh preset constraints, which refer to the harmonicities of two neighboring vowels (eg. *[+BACK][+FRONT]. These constraints represent an approximation to how a learner with prior knowledge of what kinds of restrictions it has to look for in a vowel harmony system would find the generalizations of the Finnish vowel harmony pattern. This learner has the advantage in knowing that it the vowel-vowel interactions matter the most in such a system, and also that it can ignore certain distinctions (eg. the height features for [+back] vowels).

This setting provides a baseline for the next condition: as a target for the discovery learner, the results of learning can be compared across these two conditions. This can show how close the learner is able to get to discovering the regularities of Finnish vowel harmony.

In the **discovery learning** condition, the learner was trained on the training corpus and it was instructed to discover constraints, which were analyzed by giving wug words to the grammar and summarizing their scores by their harmonicity. This setting models the unbiased learner that has no prior information in how it should approach understanding the phonotactic regularities of the language.

4.2.1.5 Parameters and testing

The learner was instructed to find 55 constraints in both the word form and the stem analysis conditions. This seemingly random number was used because trying to learn more constraints caused the program to crash in certain conditions — and because 55 constraints seemed to be enough for the learner to produce an acceptable result. To see how the learned constraints fared in discovering grammaticaly patterns, the resulting grammars were tested on wugs as well.

To understand how all combinations of back, neutral and front vowels for 2 syllables are evaluated, all vowel combinations for a word structured as $\mathrm{pV_1kV_2}$ were included in the first test word list. The vocalic inventory tested was based on Hakulinen et al. (2004): it contained all 8+8 short and long monophthongs of Finnish (*a, o, u, æ, e, i, ø, y* and their long counterpairs), and 18 diphthongs. The diphthongs included 7 with an *i* offglide (*ai, oi, ui, æi, ei, øi, yi*), 4 with an *u* offglide (*au, ou, eu, iu*), 4 with an *y* offglide (*æy, øy, ey, iy*) and the 3 lowering diphthongs *uo, ie* and *yø*. All other vowel-vowel sequences were analyzed as hiatus, with two vowel phonemes next to each other.

Examples for these wug words are *paka, pakaa, ..., paki, pakii, pakei, paky, pakyø*, etc., *pika, peeka, pøøka, ... piku, piky, ...*. Trisyllabic wugs were also tested when dealing with longer, non-disyllabic forms, in order to have a closer understanding about the possibly different patterns shown by stems and suffixed forms. The structure of these wugs was $\mathrm{pV_1kV_2lV_3}$, with all permutations of the possible Finnish vowels in all syllables again.

4.2.2 Expectations

A summary of the hypotheses tested with the UCLA Phonotactic Learner is presented in Table 4.12. Before running the learner, the expectations were that discovery learning will be able to approach the baseline set by the assisted learner. It is not expected that the discovery learner will fare better than the assisted learner: this would mean that the statistical analysis provides a much better understanding of the simple vowel harmony pattern than the standard phonological analysis of vowel harmony.

result	sanity check	prior phonology	better with word forms
assisted > discovery	✓		
assisted < discovery	×		
harmonic added > phonetic only		✓	
harmonic added ≤ phonetic only		×	
word forms > stems			✓
word forms < stems			×

Table 4.12: Summary of expectations and hypotheses for the UCLA Phonetic Learner

The further results test different hypotheses on how learners can acquire vowel harmony. If the harmonically added feature set fares better than the phonetic only feature set, that would mean that the learner has to have an understanding of the phonological patterns of Finnish before being able to acquire the static phonotactic generalizations. This prior knowledge might come from learning the alternations of suffixes in Finnish, which could help the learner restructure its focus on certain natural classes that seem to matter more in Finnish phonology.

Similarly, if full, possibly suffixed word forms are easier to learn than stems, that would indicate that learing is more optimal if the input is not a lexicon of stems, but a full corpus of all forms encountered. If the result turns out to show the contrary scenario, that would mean that learning is more likely to operate on a more abstract lexical level.

4.2.3 Results

This section summarizes the results — first, it will discuss the results on the analysis of stems, followed by the analysis of word forms. The results will show that it is quite hard for the discovery learner to perform as well as the assisted learner in modeling the Finnish pattern on the corpus. However, learning can be helped by using a harmonic added feature set and by providing the learner with a corpus with full, possibly suffixed word forms.

4.2.3.1 Learning stems

4.2.3.1.1 Assisted learning The first step was to see how the learners fare if they are supervised with a preset grammar with only the bigram constraints on the harmonicity of the word and no other constraints. These results could then serve as a target to the discovery grammar. The constraint weights learned by the phonetic features only feature set is summarized in Table 4.13 and the weights learned by the harmonic features added set is shown in Table 4.14:

constraint	penalizes	phonetic only
	harmonic	
*[+BACK][+BACK]	back only	0.00
*[-LOW,-ROUNDED][-LOW,-ROUNDED]	neutral only	0.41
*[-BACK,+ROUNDED][-BACK,+ROUNDED]	front only	0.88
*[-BACK,+LOW][-BACK,+LOW]	front only	0.00
	transparent	
*[+BACK][-LOW,-ROUNDED]	back-neutral	0.00
*[-LOW,-ROUNDED][+BACK]	back-neutral	0.13
*[-LOW,-ROUNDED][-BACK,+LOW]	front-neutral	0.67
*[-BACK,+LOW][-LOW,-ROUNDED]	front-neutral	0.64
*[-LOW,-ROUNDED][-BACK,+ROUNDED]	front-neutral	2.68
*[-BACK,+ROUNDED][-LOW,-ROUNDED]	front-neutral	1.28
	disharmonic	
*[-BACK,+ROUNDED][+BACK]	disharmonic	3.73
*[+BACK][-BACK,+ROUNDED]	disharmonic	3.63
*[-BACK,+LOW][+BACK]	disharmonic	4.57
*[+BACK][-BACK,+LOW]	disharmonic	4.55

Table 4.13: Weighing all harmonic constraints — phonetic features only

constraint	penalizes	harmonic added
harmonic		
*[+Back][+Back]	back only	0.00
*[+Neutral][+Neutral]	neutral only	0.43
*[+Front][+Front]	front only	0.50
transparent		
*[+Back][+Neutral]	back-neutral	0.00
*[+Neutral][+Back]	back-neutral	0.15
*[+Neutral][+Front]	front-neutral	1.47
*[+Front][+Neutral]	front-neutral	0.98
disharmonic		
*[+Front][+Back]	disharmonic	4.13
*[+Back][+Front]	disharmonic	4.02

Table 4.14: Weighing all harmonic constraints — harmonic features added

The results of both feature systems show that the learner succeeded in separating ungrammatical patterns from grammatical ones: all constraints penalizing disharmonic stems are weighted higher than all of the other constraints. There is also some variation among the weights assigned to constraints ruling out grammatical forms. This variation seems to show frequency effects: constraints referring to harmonically Front vowels, which are the least frequent of the three vowel harmonicity classes in the language, received a higher weight than the other constraints penalizing grammatical stems.

For a clearer summary, the wug words used in the discovery learning study were used here too: each wug was assigned its score and these harmony scores were averaged across harmonicity category. The results now can be shown side by side as in Table 4.15 below. This table also shows that these grammars did fairly well in separating the ungrammatical forms from the grammatical ones. Importantly, they did not deem the Front-Front harmonic stems ungrammatical despite the overall scarcity of Front vowels in the data, but they still ran into problems with the Front-Neutral stems:

vowels	**phonetic only**	**harmonic added**	**grammatical?**	**frequency**
		harmonic forms		
BB	0.00	0.00	✓	3003
FF	0.39	0.50	✓	686
NN	0.41	0.43	✓	792
		front-neutral forms		
FN	1.07	0.98	✓	423
NF	2.01	1.47	✓	283
		back-neutral forms		
BN	0.00	0.00	✓	1533
NB	0.13	0.15	✓	1320
		disharmonic forms		
BF	3.94	4.02	×	22
FB	4.01	4.13	×	19

Table 4.15: Wug results of the assisted grammars

4.2.3.1.2 Discovery learning In discovery learning, the learner was instructed to find 55 constraints that describe the phonotactics of the tested frequent disyllabic stems well. The 6 highest ranked constraints on the vowel tier for each grammar are listed on Table 4.16.

constraint	description	harmonic?	weight
	phonetic only		
*VVV	no trisyllables	×	6.49
*#V#	no monosyllables	×	6.31
*[+BACK][-BACK,+LOW]	eg. **a.æ*, **o.æ*, ...	✓	4.94
*[-BACK,+LOW][+BACK]	eg. **æ.a*, **æ.o*, ...	✓	4.55
*#[+ROUNDED,+RAISING]	no first syllable raising dipthong	×	4.27
*##	no words with no vowels	×	3.91
*[+BACK][+RDING,+FTING]	no back vowel followed by y-diphthong, eg. **a.ey*, **o.øy*, ...	✓	3.7
	harmonic added		
*VVV	no trisyllables	×	6.83
*#V#	no monosyllables	×	6.17
*[U-DIPH][+FRONT]	no u-diphthongs followed by Front vowels, eg. **au.ø*, **ou.æ*, ...	✓	4.03
*#[+NEUTRAL,+Y-DIPH]	no y-diphthongs after Neutral vowels. Motivates against front-neutral patterns! Eg. **e.øy*, **i.æy*, ...	?	3.87
*##	no words with no vowels	×	3.83
*[+U-DIPH][+I-DIPH]	eg. **au.ai*, **eu.oi*, ...	?	3.74

Table 4.16: Highest ranked constraints on the vowel tier in discovery learning

These top constraints include several constraints that penalize disharmonic patterns, as well as some constraints where vowel-vowel interactions are restricted but the direction of this restriction is not clearly against disharmony. There are also some highly ranked constraints that conspire to restrict the structure of the stem to disyllables.

The results by harmonicity are summarized below in table 4.17. One grammar was trained on either of the two feature system conditions, and the score of the disyllabic wug words (the /pVkV/ set) assigned to them by these grammars were calculated. These wug word scores were summarized by grouping the tested wugs by their harmonic patterns and calculating the average score assigned by the grammar. In the table below, each column shows the aggregate mean score for a given harmonicity type for a given feature set used.

vowels	phonetic only	harmonic added	grammatical?	frequency
		harmonic forms		
BB	2.52	1.57	✓	3003
FF	5.86	3.57	✓	686
NN	3.45	3.81	✓	792
		front-neutral forms		
FN	4.37	4.43	✓	423
NF	5.69	5.57	✓	283
		back-neutral forms		
BN	2.45	3.40	✓	1533
NB	4.54	2.95	✓	1320
		disharmonic forms		
BF	9.11	8.52	×	22
FB	9.20	7.80	×	19

Table 4.17: Comparison of feature systems

Evaluating how well grammars fare is done by looking at how well it separates the scores of well-formed representations from ungrammatical ones (Hayes and Wilson 2008). It is expected that the ungrammatical forms will be assigned higher scores than grammatical forms, and this difference will be larger than the variation between grammatical scores.

The results above show that the learners were able to find that disharmonicity is discouraged in the grammar in general, as the widest separation between scores was mostly a separation between the best ungrammatical forms and the worst grammatical forms. This separation was even wider, 3.25 between BF (ungrammatical) and FF (grammatical) stems for the phonetic only feature set than for the harmonic added feature set with a separation of 2.23 between the FB (ungrammatical) and NF (grammatical) harmonicities. This might indicate that the phonetic only feature set captures the generalization behind vowel harmony better, but the range between the lowest and highest scored fully harmonic harmonicity type is wider for this feature set (3.34 between BB and FF) than for the harmonic added feature set (2.24 between BB and NN), which points to the latter having discovered the basic underlying pattern a little bit better. An alternative explanation for the wider separation between grammatical and ungrammatical forms is that the phonetic only feature set has a higher number of binary features than the harmonic added condition, which means it had to work with a higher number of natural classes, which can lead to a better fit on the frequency distributions among the attested forms.

The grammars assigned quite high scores to Front-Neutral and Front-Front harmonic stems as well, although these forms are perfectly grammatical in Finnish. The average score was 5.03 for the phonetic only and 5.00 for the harmonic added feature system. The reason for these high scores can be explained by the overall low frequency of Front vowels in the corpus.

While discovery learning did provide the expected scores, the results seem to be quite noisy. The markedness of Front-Neutral stems and Front-Front stems would be surprising if one were to follow the traditional description of Finnish vowel harmony: stems with these shapes are fully grammatical in Finnish and would not be treated as exceptional in any sense. The results for discovery learning above show, however, that the learner rated Front-Front stems (with a harmony score of 5.86 for phonetic only and 3.57 for harmonic added) and most Front-Neutral stems worse than other grammatical patterns. The reason for this is probably underexposure, as the frequency of these stems is lower, but as discussed above and seen on Table 4.8, Front-Front stems are much more frequent than what would be expected without vowel-to-vowel dependencies, so a learner should be able to categorize these forms with other grammatical patterns.

4.2.3.1.3 Evaluation of discovery learning The results of the two learning conditions were compared to see how well did the learner discover the vowel harmony pattern under the discovery learning condition, with the assisted learning results serving as a baseline. A direct comparison of constraint weights between discovery and assisted learning is not possible, as the set of constraints learned in the discovery grammars are not a subset of the constraints in the assisted learning grammars. The comparison of the results of discovery learning with the results of assisted learning is still possible by comparing the range of scores assigned to wug forms with different harmonic combinations, by measuring the separation between grammatical and ungrammatical forms and by calculating the ratios of the scores assigned by the discovery and the assisted learning grammars.

As discussed above, the minimal **separation** between the grammatical and the ungrammatical forms was 3.25 for the phonetic only and 2.23 for the harmonic added feature set in discovery learning. The mere existence of these separation shows that the learner was at least partially successful in figuring out what is grammatical and what is not in the data set. The size of the separation was of a similar magnitude for the assisted learning in the harmonic added condition (2.66), but slightly lower for the phonetic only condition (1.93), which is also indicative of the success of discovery learning. The main difference between the results is that the target grammars have generally low scores for grammatical forms, with scores well under 1 for all completely harmonic (FF, NN and BB) stems; while there is more noise in form of higher harmony scores in the discovery grammar results.

This main difference is apparent from the **range** of scores between the best and the worst rated grammatical forms. As Table 4.18 below shows, this range is consistently lower for the assisted learning condition, which indicates less noise, and less overfitting to the underlying frequency distributions. This difference is even more apparent when the front-neutral stems are set aside, as the range of scores gets minuscule for the assisted learning condition, and the harmonic added feature set condition fares better in discovery learning when these stems are disregarded. The reason for the high scores assigned to front-neutral stems is the relatively low frequency of these stems, which the learner notices and formulates generalitions against this pattern.

feature set	**front-neutral included**		**without front-neutral**	
	assisted	**discovery**	**assisted**	**discovery**
phonetic only	2.01	3.42	0.41	3.42
harmonic added	1.47	4.00	0.43	2.24

Table 4.18: Range between the lowest and highest average score for a grammatical form

It seems a little problematic that the range between the highest and lowest average score for grammatical forms is higher than the separation between grammatical and ungrammatical harmonicities. This indicates that the discovery learner has not succeeded completely in modeling vowel harmony: there is no clean boundary between forms that the grammar rejects as ungrammatical and forms that it accepts as grammatical. This boundary might be drawn so that front only and/or front-neutral stems could be incorrectly considered ungrammatical. This issue indicates that the low frequency of Front vowels influences the UCLA Phonotactic Learner too much, and the low observed frequency of stems with Front vowels leads it to consider such stems as marginally ungrammatical. As speakers do not consider front only and front-neutral stems as ungrammatical, further steps should be taken to help the learner in modeling the vowel harmony pattern better.

4.2.3.2 Learning possibly suffixed word forms

Learning vowel harmony based exclusively on stem forms might not accurately model the real world learning problem, and as discussed above, comparing the learned grammar's efficiency on stems and on words could give more insight into how the grammar works. Adding suffixed forms to the training corpus might also raise the observed frequency of Front vowels, so that the grammar could separate ungrammatical forms from grammatical forms better than it did with stem forms.

To be able to make the comparison with the stem results easier, the learner was trained on those words in the corpus, whose stems were amongst the 8081 frequent disyllables analyzed in the previous section. This yielded 278086 word form types for the UCLA Phonotactic Learner to train on. All of the other parameters were the same as for the stem learning condition.

4.2.3.2.1 Assisted learning Table 4.19 below illustrates how the grammar performs if the set of constraints is predetermined. There is a very clear separation between grammatical and ungrammatical forms in this condition, indicating that the assisted learner fares really well if it has access to all suffixed word forms. The worst scores for grammatical structures are given to front-neutral forms, which are, again, probably due to the scarcity of Front vowels in the language. However, front only forms are especially infrequent, yet the learner makes a good job in classifying these stems as grammatical.

vowels	phonetic only	harmonic added	grammatical?	frequency
		harmonic forms		
back only	0.00	0.00	✓	41502
front only	0.88	0.48	✓	9387
neutral only	0.00	0.00	✓	10777
		transparent forms		
front-neutral	1.68	0.68	✓	60739
back-neutral	0.00	0.00	✓	154662
		disharmonic forms		
disharmonic	5.88	5.56	×	1019

Table 4.19: Results of the ranking of the pre-learned constraints.

The scores of front only and front-neutral stems are, however, quite well separated from the disharmonic forms, indicating that the learner has succeeded in modeling the main facets of Finnish vowel harmony, and it can serve as the baseline for the discovery learning results.

4.2.3.2.2 Discovery learning The learner, as in the stem learning condition, was instructed again to find 55 constraints on either the full string or on the projected vocalic tier. The 6 highest ranked constraints on the vowel tier are listed for illustrative purposes on Table 4.20. Most of these constraints penalize disharmonic structures, but overwhelmingly, these constraints are highly specialized for one segment, like the ones penalizing back vowels before /æ/. The grammar using added harmonic features, though, managed to weigh the general disharmonicity penalizing constraint *[+Back][+Front] high, as it is the 5th highest ranked constraint on the vowel tier.

constraint	description	harmonic?	weight
	phonetic only		
*[+BACK][-BACK,+LOW]	no back vowel before /æ/, eg. **a.æ*, **o,æ*, ...	✓	6.83
*[-BACK,-HIGH,+ROUNDED][+BACK]	no back vowel after /ø/, eg. **ø.a*, **ø.u*, ...	✓	6.76
*[+ROUNDING,-BACKING][+BACK]	no back vowel after an y-diphthong, eg. **øy.a*, **æy.o*, ...	✓	6.06
*[+BACKING][-BACK,+LOW]	no /æ/ after an u-dipthong, eg. **au.æ*, **ou.æ*, ...	✓	5.95
*##	no words with no vowels	×	5.55
*[-BACK,+LOW][+BACK]	no back vowel after /æ/, eg. **æ.a*, *æ.o*, ...	✓	5.32
	harmonic added		
*[+U-DIPHTHONG][+FRONT]	no Front vowel after an u-dipthong, eg. **au.æ*, **ou.y*, ...	✓	7.57
*#[+NEUTRAL,+Y-DIPHTHONG]	no first syllable *ey, iy*	×	5.79
*[+Y-DIPHTHONG][+BACK]	eg. **ey.a*, **øy.u*, ...	✓	5.61
*[+Y-DIPHTHONG]#	no final syllable y-diphthong	×	5.47
*[+BACK][+FRONT]	no BF disharmonic sequences, like **u.y*, **a.ø*, ...	✓	5.32
*##	no words with no vowels	×	5.25

Table 4.20: Highest ranked constraints on the vowel tier in discovery learning for word forms

The results for the discovery learning for the possibly suffixed word forms is summarized in Table 4.21 below, where these results are compared with the results on the stems only condition, as seen in the previous section. The results are, again, illustrated by showing the average score assigned by the grammar to a set of wug words used for testing purposes. As the input for training was now not limited to disyllables, the wugs the grammar was tested on now consisted of di- and trisyllabic word forms shaped pV_1kV_2 and $pV_1kV_2lV_3$ respectively. If there was both at least one Back and at least one Front vowel in the wug form, it was classified as disharmonic. Otherwise, if the form contained Neutral vowels mixed with Back or Front vowels, it was classified as back-neutral or front-neutral respectively, and words were classified as back only, neutral only and front only when the vowels in the word had the same harmonic category.

vowels	phon.	(stems)	harm.	(stems)	grammatical?	frequency
			harmonic forms			
back only	4.21	2.52	2.69	1.57	✓	41502
front only	5.40	5.86	5.00	3.57	✓	9387
neutral only	10.74	3.45	6.89	3.81	✓	10777
			transparent forms			
front-neutral	10.00	5.03	7.53	5.00	✓	60739
back-neutral	8.54	3.50	6.61	3.18	✓	154662
			disharmonic forms			
disharmonic	14.66	9.16	13.19	8.16	×	1019

Table 4.21: Results of discovery learning for words whose stem is disyllabic, compared with disyllabic stems only

It can be seen that the grammar gives much worse scores to certain harmonic patterns, and seems to be doing a worse job than the grammars trained on the stems only. This is understandable: the suffixed forms show more complex patterns as they are longer and they might involve several affixes, some of which might not be harmonic, therefore the data contains more noise. However, it is not entirely clear why back-neutral forms are so dispreferred in these grammars, and why neutral-neutral harmonic forms are judged to be worse than the front only forms which were dispreferred by the grammars trained on stems.

4.2.3.2.3 Evaluation of discovery learning The separation between the grammatical harmonicity type with the highest rating and the ungrammatical harmonicity type (disharmonic) was 3.91 for the phonetic only and 5.66 for the harmonic added feature set. As the separation was wider for the harmonic added feature set, it seems that it did a better job in finding the grammatical patterns in the language. The separation between grammatical and ungrammatical forms was of a similar magnitude in raw scores in the target grammar with assisted learning (4.20 for phonetic only and 4.89 for harmonic added), but the main difference between discovery and assisted learning can be found when measuring the range of scores with grammatical forms, as seen on Table 4.22.

feature set	**assisted**	**discovery**
phonetic only	1.68	6.53
harmonic added	0.68	4.84

Table 4.22: Range between the lowest and highest average score for a grammatical form

This range of scores within grammatical harmonicity types raises a red flag for the discovery learner again. The learner using phonetic features only does not have a clear boundary between grammatical and ungrammatical forms at all, as the range of scores for different types of grammatical words is higher than their separation from disharmonic stems. The learner using the harmonic features added feature set does a slightly better job, but the separation between grammatical and ungrammatical forms is almost of the same magnitude than the range of grammatical scores. Both grammars failed to approach the quality of the results for the hand assisted learning: the latter ones clearly separate disharmonicity from all other types of stem harmonicities by having a narrow range between grammatical scores and a much higher separation between grammatical and ungrammatical forms.

4.2.3.3 Discussion

The UCLA Phonotactic Learner has done a good job in rating disharmonic stems and word forms as the least acceptable of the wug stems and word forms it was tested on. In all eight models, including both assisted and discovery learning grammars, disharmonic stems were assigned the highest average scores (were judged to be the least acceptable).

The learning process of these models was not completely successful, though, as the separation between disharmonic and non-disharmonic forms was not always wide enough to convincingly split the lexicon into grammatical and ungrammatical. To assess the quality of how well a given grammar succeeded in this splitting the wug stems into grammatical and ungrammatical subsets, the range of grammatical forms and the separation between grammatical and ungrammatical forms were compared.

input	feature	constraints	grammatical range	separation	s/r
stem	phonetic only	assisted	2.01	1.93	0.96
stem	phonetic only	discovery	3.42	3.25	0.95
stem	harmonic added	assisted	1.47	2.54	1.73
stem	harmonic added	discovery	4.00	2.23	0.56
word form	phonetic only	assisted	1.68	4.20	2.50
word form	phonetic only	discovery	6.53	3.91	0.60
word form	harmonic added	assisted	0.68	4.89	7.22
word form	harmonic added	discovery	4.84	5.66	1.17

Table 4.23: Measuring the success of grammars by comparing the range of grammatical scores and the grammatical-ungrammatical separation

Table 4.23 above summarizes all grammars that were the results of the learning processes in the previous section, with the range between grammatical forms, the separation between the highest rated grammatical and the lowest rated ungrammatical form, and the ratio s/r between these two numbers. If this ratio is below 1.0, like in most discovery grammars and in the phonetic only grammars on stems, that means disharmonicity is not separated very well from the grammatical forms: there is no clean boundary between the two subsets.

Based on this table, the discovery grammar trained on suffixed word forms and using the harmonic added feature system fared unambiguously the best, as the separation is the widest for this grammar both in absolute value and in relative to the range of grammatical forms. However, the hand assisted grammar with the same settings reaches a much better result with an s/r ratio of 7.22.

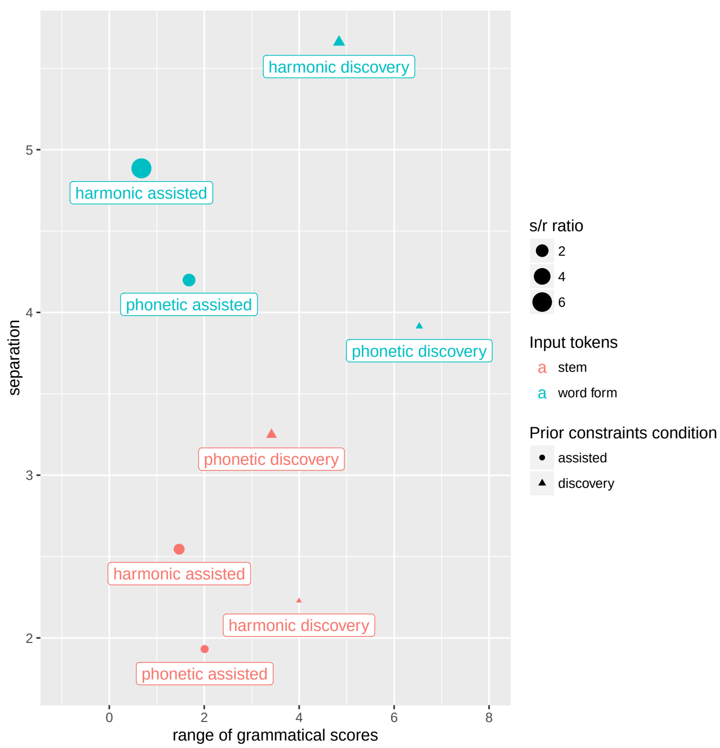


Figure 4.4: Range of grammatical forms and separation for different grammars

The results are visualized on Figure 4.4 above too. The higher and more to the left a grammar is on this plot, the better it fared in establishing a separation. It is visible on this figure as well, that with possibly suffixed word forms included, the grammars fared better, although with the price of a somewhat extended range of grammatical patterns for the discovery grammars. This indicates, as already discussed above, that the problem lies in these grammars overfitting the frequency distributions of different harmonicity types, which is especially problematic with front only and front-neutral stems, as harmonically Front vowels are infrequent in the input.

The fact that learning word forms did help the grammar, but the learning process was not entirely successful, a possible strategy to move forward is to help out the learner with the knowledge that it has to sample from a more limited set of constraints, which would help it reach scores closer to the assisted grammars. In order to know what the limited set of constraints, or the limited set of features to build these constraints from is, the learner needs to have access to the alternations that Finnish grammar shows. Since discovering these alternations necessarily implies discovering the relevant features for the static generalizations, it could only help in the grammars achieving a better result.

4.3 Learning alternations with the Sublexical Learner

The previous section showed that the UCLA Phonotactic Learner does not do a perfect job in discovering the relevant constraints that restrict disharmonicity in the Finnish vowel harmony pattern. The problem seemed to be an overfitting to frequency patterns in the language that are irrelevant for vowel harmony, like the low frequency of front rounded vowels. Another issue was that the learner performed poorly when it was not aware of the language-specific distinction between vowels that participate in harmony and neutral vowels. The idea proposed in this section is that learning allomorphy patterns would help out the learner in what features to pay more attention to, and in what kinds of structures it should constrain.

Learning allomorphy was modeled using the Sublexical Learner of Allen and Becker

(2015). The input to the learner was the nominal stems from the list of frequent disyllables used in the previous section, so that the results could be comparable. As the task of the learner was to learn allomorphy, a suffixed form for the stems was provided for the learner to learn from.

The goal of training the Sublexical Learner on the Finnish suffixation data was to see if it is able to figure out the relevant natural classes for modeling Finnish vowel harmony. If it is able to do so, that might indicate that either a serial or a parallel learning of alternations (as with the Sublexical Learner) and of the static phonotactic generalizations (as with the UCLA Phonotactic Learner in the previous section) is the best model of acquiring the Finnish vowel harmony pattern.

4.3.1 Setup

This section will discuss the setup for the Sublexical Learner. Again, as in the setup of the UCLA Phonotactic Learner, there were several options to choose from when deciding on the input of the training and on the parameters given to the learner.

4.3.1.1 Training data

The training data for the Sublexical Learner consists of pairs of stems and their suffixed forms. The learner is capable of learning the nature of alternations and how different alternations are distributed in the lexicon by breaking it down to sublexica, and building a gatekeeper grammar for each allomorph participating in the alternation.

For example, consider a hypothetical example of a perfectly regular vowel harmony with all stems with back vowels being suffixed with *-na* and all stems with front vowels are suffixed with *-næ*. The learner would first break down the lexicon into two groups: the *-na*-

suffixing and the *-næ*-suffixing sublexica. Then, by looking at the phonotactic regularities within these sublexica, it would build two gatekeeper grammars: one for the *-na*-suffixing sublexicon where the constraint *[-BACK] would have a very large weight and one for the *-næ*-suffixing sublexicon with a highly weighted *[+BACK] constraints. Thus it would be able to predict that a new stem with back vowels would with an almost 100% probability would be classified in the *-na*-suffixing sublexicon.

Learning the exact nature of alternations in Finnish is outside the scope of this chapter. To allow the learner to focus on the relevant question of separating the front-suffixing and back-suffixing sublexica of Finnish stems, a simplified artificial suffixed form was created instead of using a real one. This suffixed form was generated using the Omorfi morphological parser (Pirinen 2014), which was used in the previous chapter as well. Because suffixation frequently introduces consonant alternations in the stem (eg. *kæsi* "hand" ∼ *kætenæ* "hand.ESS"), the actual suffixed stems were not used, but the vowel quality of the suffix was determined as Front or Back, and the suffixed form was actually just a concatenation of the nominative form and an F or B suffix determining the harmonicity of the suffix (eg.*kæsi* ∼ *kæsiF*).

4.3.1.2 Constraints

The Sublexical Learner has to be provided with a set of constraints that it uses both to categorize stems and to evaluate final word forms. The given constraints can refer to just one segment (like *[+BACK], which assigns violation marks to back vowels), or it can specify longer spans of segments. These constraints are first used in a gatekeeper grammar to classify stems into sublexica, and then the same constraints are used in the "proper" grammar to evaluate the final surface form.

There were several types of constraint systems tested with the learner. There were two systems with unigram constraints — constraints that are violated by the presence of a given feature set anywhere in the word. The two unigram constraint sets differed in how many features could be present maximally in a feature bundle describing a constraint: the unigram full model did not have any such restrictions, while the unigram 2-max model limited the number of features to be 2 at most within a feature bundle. There was also a bigram constraint set tested, which included all possible pairs of the feature bundles defined in the 2-max unigram model, allowing any number of consonants to intervene between these two vowels.

When selecting the constraints to use in this analysis, there must be some limitations employed as using all possible phonotactic combinations would cause the model to overfit the training data, if the learner would be even able to successfully run with such a vast number of possible constraints. This limitation does introduce some underlying bias to the constraint system used, though. The chosen assumption to introduce was that only vowels are relevant in conditioning the vocalic alternation in the suffix. The results of the wug study in the previous chapter, as well as the findings of Hayes et al. (2009) justify this assumption, as it is clear that speakers pay much more attention to vowel-vowel correspondences than consonantal effects in the case of an alternation driven by vowel harmony.

The goal of the **unigram full** model is to identify the features or the feature combinations that are the best in separating back-suffixing stems from front-suffixing ones. In order to achieve this, the input constraints to the grammar included all markedness constraints that can be legally expressed using the features [±back], [±rounded], [±high] and [±low]. Illegal feature combinations are [+high,+low] as no such segments exist, and [+high,-low] and [-high,+low] as well as they are redundant, and can be simply replaced with [+high] and [+low] respectively. There are $3^4 = 81$ feature combinations up to 4 features (each feature can be +, - or not present), the number of possible combination for each illegal high-low combination is 3^2 (for the two other feature values). There is also 1 empty combination, so there are $3^4 - 3 * 3^2 - 1 = 53$ legal constraints used in this system. An example for a constraint in this system would be *[-BACK,+ROUNDED,+HIGH], which would be violated by any /y/ /yy/ vowels in the word. One would expect such a constraint to have a high weight in the back gatekeeper grammar, meaning that stems with these vowels would be unlikely to be classified into the back-suffixing sublexicon.

The **max-2 unigram** model strove to achieve the same goal as the unigram full model, but the number of constraints was reduced to distract the learner less to certain particular combinations. The number of features in a constraint was limited to 2 at most, so the *[-BACK,+ROUNDED,+HIGH] constraint could not be included in the model. Feature combinations which selected the same subset of vowels that another, possibly simpler, feature bundle were also excluded. This limited the number of constraints to 24.

The **max-2 bigram** model was extended to base its assesments not on the individual vowels in the stem, but on vowel-vowel sequences, as the order of vowels clearly matters in the Finnish grammar. The number of constraints therefore was $24^2 = 576$. An example for such a bigram constraint is *[+BACK]C_0[+ROUNDED,+HIGH], which would penalize back vowels followed by a high rounded vowel in the stem.

There was no bigram full model for two reasons. It will be seen that the max-2 unigram model actually fared better than the unigram full model, so it made sense to limit the search space for the learner. An explanation for why limiting the number of constraints was helpful for the learner is that it introduced a bias towards finding generalizations for bigger natural classes, instead of overfitting to individual segments, and this bias turned out to improve the analysis significantly. The number of constraints for a bigram full model would also have been prohibitively high ($53^2 = 2809$), which also would have presented challenges in computational capacity.

A different way of exploring the search space is to use the constraints discovered by the **phonotactic learner** when learning over word forms. This condition would model a learning situation where phonotactic learning and the acquisition of alternations happens more or less simultaneously. If the results indicate that using the constraint system of the UCLA Phonotactic Learner improves the performance of the Sublexical Learner, that might indicate that such an overlap between learning these two facets of vowel harmony is advantageous for the learner.

4.3.1.3 Feature system

Only the phonetic only feature system was used with the Sublexical Learner, as part of the goals that are set in front of this learner is whether it is able to discover those feature combinations that are the most useful in understanding Finnish vowel harmony. In order to do so, the plain, phonetically based [±back] and [±round] features are used instead of the privative [Front], [Neutral] and [Back] harmonically defined features.

4.3.1.4 Testing

The grammars were tested on the same wug disyllables that were used with the UCLA Phonotactic Learner — words with a pV_1kV_2 structure. The task of the Sublexical learner was to figure out which sublexicon it assigns different kinds of wug stems to. It is expected of the learner that it assign front suffixes to front only, neutral only and front-neutral stems, and back suffixes to back only and back-neutral stems. There is very little data in the lexicon that can help the learner deal with disharmonic forms, so the results for such stems might not be as accurate as the results for other types of stems.

4.3.2 Results

4.3.2.1 Full unigram constraints

The results of the Sublexical Learner trained with all possible unigram constraints is summarized below in Table 4.24. Each row, again, represents the average of the scores or percentages for wug stems within a given harmonicity type. There are two columns illustrating the performance of the grammar: the difference between the score assigned by the back-suffixing gatekeeping grammar and the score assigned by the front-suffixing grammar is shown (positive numbers indicating preference for back-suffixation), as this difference is responsible in deciding which sublexicon the grammar assigns stems into. The Sublexical Learner also assigns probabilities to each wug's classification, and these probabilities are averaged as well. Finally, the lexical statistics are also shown, in the percentage of back-suffixing stems in the given stem type.

vowels	score difference	% back	regular suffix	lexicon % back
harmonic forms				
BB	34.04	92.09	back	100.00%
FF	9.36	65.13	front	0.00%
NN	22.99	99.98	front	1.21%
front-neutral forms				
FN	16.17	77.48	front	0.00%
NF	16.17	77.48	front	0.00%
back-neutral forms				
BN	28.32	89.21	back	100.00%
NB	28.32	89.21	back	100.00%
disharmonic forms				
BF	21.50	79.38	front	20.00%
FB	21.50	79.38	back	100.00%

Table 4.24: Sublexical results with full unigram constraints

These results indicate that the full unigram learner failed to split the grammar to back- and front-suffixing sublexica, as it assigns all stems to be more likely back-suffixing. It is capable of differentiating between harmonically Front and Neutral vowels, though, as Front vowels show a lower back-front grammar difference, starting to lean towards front-suffixing.

It is worthwile to investigate the constraints that differentiate between the two sublexica the most, and how do constraints that circumscribe the harmonic Back, Front and Neutral vowel categories fare. The top 5 constraints preferring back suffixes are shown in Table 4.25 below. This table shows that the constraint *[+BACK], that describes the phonetic feature value needed for the harmonically Back vowels was found to be an important constraint in setting the sublexica apart, as it is the constraint that prefers back suffixes the second most. The feature [+rounded] appears in several of the highest ranking constraints, which can be explained by the fact that the rounded back vowels /o/ and /u/ are much more frequent than the rounded front vowels /ø/ and /y/, so roundness and backness do correlate, so rounded vowels are indeed more likely to trigger as back-suffixing overall.

constraint	**back grammar**	**front grammar**	**preference for back**
*[+ROUNDED,+HIGH]	0.00	-1.49	1.49
*[+BACK]	-0.18	-1.50	1.32
*[+ROUNDED]	0.00	-1.12	1.12
*[+ROUNDED,-LOW]	0.00	-1.12	1.12
*[-BACK,+ROUNDED,+HIGH]	-0.43	-1.49	1.06

Table 4.25: Top 5 constraints categorizing stems as back-suffixing in the full unigram model

Table 4.26 below similarly summarizes the top 5 constraints that prefer front suffixes. The two constraints circumscribing the phoneme /æ/ are ranked first and second. The fourth and fifth highest ranked constraints refer to [-rounded] segments, which, as seen with the back vowels, does correlate with frontness, but shows that the learning algorithm didn't exactly succeed in discovering the relevant feature bundles for harmonic behavior. The vowel /a/ violates the highly ranked *[+LOW] and *[-ROUNDED,+LOW], both of which prefer front-suffixing, so the problems with the learner are even more pronounced looking at the constraints preferring front-suffixing.

constraint	**back grammar**	**front grammar**	**preference for back**
*[-BACK,+LOW]	-3.36	-0.06	-3.29
*[-BACK,-ROUNDED,+LOW]	-3.36	-0.06	-3.29
*[+LOW]	-1.01	-0.14	-0.87
*[-ROUNDED,+LOW]	-1.01	-0.14	-0.87
*[-ROUNDED,-HIGH]	-0.80	0.00	-0.80

Table 4.26: Top 5 constraints categorizing stems as front-suffixing in the full unigram model

The failure of the learning algorithm to identify the feature bundles most important for setting the two sublexica apart is also apparent when looking at the constraints that select harmonically Front vowels. While the constraints describing /æ/ are rated to be front-preferring, as seen above, the constraints describing front rounded vowels are not categorized correctly: *[-BACK,+ROUNDED] has a -0.93 score in the back and -1.12 score in the front grammar, preferring back-suffixing by 0.19. This mistake is not helped out by the constraints picking out /ø/ and /y/ individually, as the former only prioritizes front suffixing by a score of 0.36, the latter actually prefers back suffixing by 1.06.

To summarize the results above, the full unigram model failed to learn the underlying pattern of Finnish vowel harmony. There are several reasons for this failure: the high number of constraints caused overfitting to the model and failing to learn the generalizations, and the unigram nature of the model meant that it could not differentiate between stems based on the order of vowels. The two following models attempted to address these issues.

4.3.2.2 Max-2 unigram constraints

It is possible that the failure of the full unigram model was due to the high number of constraints leading to overlearning for individual vowels, or distracting the learner in some other way in understanding the underlying pattern. The model should perform better in modeling Neutral, and especially Front vowels as preferring front suffixes. The max-2 model was run to see how a simpler model might make better generalizations.

vowels	**score difference**	**% back**	**regular suffix**	**lexicon % back**
		harmonic forms		
BB	29.27	99.99	back	100.00%
FF	-11.33	16.04	front	0.00%
NN	12.90	92.32	front	1.21%
		front-neutral forms		
FN	0.78	54.95	front	0.00%
NF	0.78	54.95	front	0.00%
		back-neutral forms		
BN	21.00	99.23	back	100.00%
NB	21.00	99.23	back	100.00%
		disharmonic forms		
BF	8.89	79.16	front	20.00%
FB	8.89	79.16	back	100.00%

Table 4.27: Sublexical results with the max-2 unigram constraints

The results of the learner are shown in Table 4.27. These results are somewhat more encouraging indeed than the results of the full unigram model: at least the front only stems are now firmly classified into the front-suffixing sublexicon. There are still two kinds of problems surfacing with this model: the overall inclination of Neutral vowels to influence the stem they are in to be classified as back-suffixing (whereas they should prefer front suffixing in neutral-only and front-neutral stems); and the built-in inability to distinguish between stems based on the order of the vowels in them. These problems might be solved with introducing bigram constraints.

constraint	**back grammar**	**front grammar**	**preference for back**
*[+BACK]	-0.30	-2.58	2.28
*[+BACK,+HIGH]	0.00	-1.58	1.58
*[+BACK,+ROUNDED]	0.00	-1.58	1.58
*[+HIGH]	0.00	-0.93	0.93
*[-BACK,+HIGH]	0.00	-0.93	0.93

Table 4.28: Top 5 constraints categorizing stems as back-suffixing in the max-2 unigram model

Looking at the top 5 constraints preferring back (Table 4.28) and front (Table 4.29) suffixes, the max-2 unigram learner fared much better than the full unigram model. The back-suffixing sublexicon is preferred by the general *[+BACK] constraint the most, and the next 2 following top back-preferring constraints reference backness directly. However, the fourth and fifth constraints preferring back-suffixes the most refer to high vowels, the fifth one specifically selecting *front* high vowels, which is problematic. This result is probably due to the /i/ vowel occuring in back-neutral stems very frequently, however, it does not exclude the harmonically always Front /y/ vowels from its effect.

constraint	back grammar	front grammar	preference for back
*[-BACK,+LOW]	-4.86	-0.22	-4.64
*[-BACK,+ROUNDED]	-2.58	-0.11	-2.46
*[-ROUNDED,-HIGH]	-0.59	0.00	-0.59
*[-HIGH]	-0.55	0.00	-0.55
*[-ROUNDED,+LOW]	-0.99	-0.54	-0.46

Table 4.29: Top 5 constraints categorizing stems as front-suffixing in the max-2 unigram model

The two most front-preferring constraints select the two natural classes that define the harmonically Front class: the top constraint describing /æ/ and the one below pointing out the front rounded vowels. These two constraints are also well separated from the other constraints that favor front-suffixing, as their effect is much stronger (4.64 and 2.46 preference) than the third one (0.59).

The max-2 unigram model fared much better than the full unigram constraint set in identifying the relevant features that separate back-suffixing and front-suffixing stems in Finnish. Its performance is still far from being perfect: the order of the stem vowels still could not matter in classification, and (probably partially as a consequence) neutral only stems are still miscategorized as back-suffixing. The analysis with bigram constraints was conducted with the hope that expanding the constraint space this way will improve the performance of the learner.

4.3.2.3 Bigram constraints

The results of the bigram model are illustrated in Table 4.30. The grammar equipped with bigram constraints managed to produce the Finnish pattern with a fairly acceptable accuracy, except for neutral-only stems. Front only stems are now exclusively front-suffixing, and front-neutral stems are also classified into the right sublexicon. The percentages still indicate a stronger bias towards back-suffixing than what the lexicon would warrant, which is manifested very strongly in the case of neutral only stems. These stems, which are almost exclusively front-suffixing (98.79% of neutral only stems) in the lexicon, yet are categorized into the back-suffixing sublexicon with a 64.99% probability.

vowels	score difference	% back	regular suffix	lexicon % back
		harmonic forms		
BB	31.73	100.00	back	100.00%
FF	-19.24	0.00	front	0.00%
NN	6.69	64.99	front	1.21%
		front-neutral forms		
FN	-11.57	5.65	front	0.00%
NF	-10.97	10.96	front	0.00%
		back-neutral forms		
BN	23.83	100.00	back	100.00%
NB	32.72	100.00	back	100.00%
		disharmonic forms		
BF	-0.68	36.48	front	20.00%
FB	22.60	100.00	back	100.00%

Table 4.30: Sublexical results with bigram constraints

Investigating the top ranked constraints is more impressionistic in this model than with the unigram models with much fewer constraints, but it still provides insight in how the model understood the training data. Table 4.31 below shows the 5 constraints that prefer back-suffixing the most. All of these constraints make reference to a [+back] vowel segment, showing that the importance of this feature is very high for distinguishing between the two sublexica.

constraint	back grammar	front grammar	preference for back
*[+BACK] C_0 [-ROUNDED]	-0.20	-1.53	1.33
*[+BACK] C_0 [-BACK,-ROUNDED]	-0.18	-1.28	1.09
*[+BACK] C_0 [-ROUNDED,-LOW]	-0.18	-1.28	1.09
*[-LOW] C_0 [+BACK]	0.00	-1.09	1.09
*[+BACK] C_0 [+BACK]	0.00	-0.98	0.98

Table 4.31: Top 5 constraints categorizing stems as back-suffixing in the max-2 bigram model

Table 4.32 presents the constraints with the highest preference for front-suffixing. Again, all constraints in the top 5 refer to at least one [-back] vowel segment, identifying the most important feature that drives vowel harmony. All of these constraints also refer to either one of the two natural classes that make up the harmonically Front set of vowels: 4 refer to rounded front vowels and one to the low front vowel /æ/.

constraint	back	front	back preference
*[-BACK,+ROUNDED] C_0 [-BACK]	-1.52	0.00	-1.52
*[-BACK,+ROUNDED] C_0 [-BACK,-ROUNDED]	-1.18	0.00	-1.18
*[-BACK,+ROUNDED] C_0 [-BACK,-LOW]	-1.17	-0.04	-1.14
*[-BACK] C_0 [-BACK,+LOW]	-0.97	0.00	-0.97
*[-HIGH] C_0 [-BACK,+ROUNDED]	-0.93	0.00	-0.93

Table 4.32: Top 5 constraints categorizing stems as front-suffixing in the max-2 bigram model

The results of the bigram model look much closer to the expectations based on the lexicon than the unigram models: front-neutral stems are now front-suffixing, and the main problem with the ordering of vowel is solved. As back+back and front-back disharmonic stems are indistinguishable for unigram models, their output showed that the learner was confused that they demand different suffixation patterns. The bigram model managed to separate these two types: back+back stems are leaning towards front-suffixing and front-back stems are consistently analyzed as back-suffixing, both results being consistent with in the lexical trends.

To summarize, the bigram model succeeded in fixing the problems presented by the results of the unigram models. Harmonically Front vowels now firmly prefer front-suffixing, and Neutral vowels are influencing front-neutral stems towards back-suffixing much less. The problems with the order of vowels in the stem have been solved: Back-Front disharmonic stems are now correctly identified as mostly front suffixing and Front-Back disharmonic stems are back-suffixing with a 100% probability.

4.3.2.4 Constraints based on the Phonotactic Learner

Table 4.33 below illustrates the results of learning with constraints provided by the UCLA Phonotactic Learner trained on word forms, using a phonetic only feature system and instructed to discover constraints characteristic to the corpus. The 55 constraints provided by the phonotactic learner were reformatted to the formalism used by the Sublexical Learner, and used as the input constraints to build the gatekeeper and the proper grammar on.

vowels	score difference	% back	regular suffix	lexicon % back
harmonic forms				
BB	0.39	57.26	back	100.00%
FF	-5.08	22.86	front	0.00%
NN	-0.20	48.62	front	1.21%
front-neutral forms				
FN	-3.46	23.05	front	0.00%
NF	-2.43	35.37	front	0.00%
back-neutral forms				
BN	-0.04	51.55	back	100.00%
NB	0.75	61.67	back	100.00%
disharmonic forms				
BF	-2.10	38.93	front	20.00%
FB	-0.34	43.78	back	100.00%

Table 4.33: Sublexical results with constraints discovered by the Phonotactic Learner

The results show that the learner does improve on the performance of the full unigram model, but fares poorer than the max-2 unigram and, most importantly, the max-2 bigram models. It manages to learn that neutral only and front only stems prefer front suffixes, unlike the full unigram model. However, the main comparison for this model has to be the bigram model, as they both employ constraints that refer to vowel-to-vowel correspondences, and its performance is much worse than the bigram model. Its performance is worsened by a lack of certainty in the categorization of stems: it fails to make decisions close to 100% or 0%. The reason for this might be the low number of constraints (55 versus the 576 of the bigram model), and that these constraints are not as generic as they can be when set up by hand and there is more variety for the grammar to work on. The fact that the result is still promising and *better* than the full unigram model shows that overlap between learning phonotactics and alternations might be beneficial for both learners.

4.3.3 Discussion

The results of the Sublexical Learner show that when trained on the relevant alternation, a phonotactic learner is able to isolate the features and feature bundles that are the most relevant for the Finnish vowel harmony pattern. When presented with a limited search space of maximum 2 features specified for a segment in a constraint, the learner finds the four phonetic feature bundles [+back]; [-back,+rounded], [-back,+low]; and [-back,-rounded,-low] that define harmonically Back, Front and Neutral vowels respectively.

This successful identification shows that a learner when presented with dynamic phonotactic generalizations is clearly able to find the relevant feature bundles that differentiate between the two sublexica. These feature bundles can, therefore, be used when learning static phonotactic generalizations. This ability, as seen in Section 4.2 significantly improves the performance of the phonotactic learner, as it is able to generalize based on the classes of phonemes that matter in both facets of vowel harmony.

4.4 Summary

This chapter discussed the learnability of the Finnish vowel harmony pattern using maximum entropy learners, one that works on static phonotactic generalizations and one that classifies stems in sublexica based on their morphological behavior. Both the UCLA Phonotactic Learner and the Sublexical Learner showed that they are able to figure out the underlying pattern if the corpus, the feature set used and the constraint set used introduces certain biases that gear them towards the correct assumptions.

The UCLA Phonotactic Learner succeeded in evaluating the pattern when assisted by constraints that directly address front-back vowel harmony. If it is instructed to discover constraints on its own, it does not succeed that well to separate grammatical and ungrammatical Finnish forms. The performance of both the assisted and the discovery learner improves if the training corpus includes all word forms, not just stems and if the feature set reflects a prior phonological knowledge of what the important features for understanding Finnish vowel harmony are. This implies that in order to succeed, learning phonotactics of vowel harmony has to be preceded or has to be simultaneous by other learning processes that work on a similar input and discover the relevant features to deal with vowel harmony.

This process was modeled with the Sublexical Learner, which classifies word pairs that represent alternations into different sublexica that determine what kind of alternation a given base undergoes. This learner showed that it is capable of identifying the crucial constraints which are the underlying forces of Finnish vowel harmony. There were some biases that helped the learner succeed with the alternations as well. One such bias was the limitation on the maximal number of features in a feature bundle in a constraint to 2, which forced the learner to focus on broader groups of vowels instead of overfitting to each possible sound and natural class. Another bias encoded the Sublexical Learner's performance was the focus on vowels only: this bias seems to be present in actual speakers as seen in the previous chapter. This bias can be extended by hypothesizing that learners are aware that vowel-vowel interactions are the most relevant for vowel harmony, and can use bigram constraints on a vocalic tier when classifying stems. Using these bigram constraints also improved the performance of the learner greatly.

The conclusion of this chapter is that there has to be a part of phonotactic learning that occurs, or is amended after learning morphophonemic alternations. For the acquisition of the vowel-to-vowel correspondence restrictions to succeed, learning alternations has to happen either before or simultaneously. It might be the case that learning generalizations on more complex vowel-to-vowel dependencies happen much later than the acquisition of basic phonotactics, or the results or early learning is later revised and refined based on the knowledge gained during the learning of higher level grammar. Regardless, the results indicate that learning the alternations does help shrink the search space for constraints in phonotactic learning. The range of possible constraints and the natural classes that are restricted in these constraints is too large for the phonotactic learner to successfully achieve the judgments that are available by using linguistic insight. This range can successfully be shrunk by using the alternations to classify stems and to discover the most relevant classes, using which the phonotactic learner is able to separate grammatical and ungrammatical forms effectively.

Chapter 5

Conclusions

This dissertation analyzed several types of exceptionality patterns in vowel harmony and set out to find out what kinds of biases help speakers in learning and maintaining such patterns. Nonce word studies were employed to explore this question: whether speakers use certain biases in an experimental design carefully controlled for a certain bias when determining whether a novel word should be exceptional or regular can shed some light on what kinds of information speakers use in acquisition.

Chapter 2 investigated the Hungarian antiharmonicity pattern, and sought out whether a phonetic level analysis is possible for this phenomenon. The findings of Benus and Gafos (2007) were not clearly replicated: there was no consistent phonetic difference found between antiharmonic and regular stems when the word is unsuffixed, and only a tiny difference was found with suffixed word forms. The differences even within suffixed words were below the perceptibility threshold of what Hungarian listeners were able to distinguish.

Finding a lack of evidence for the phonetic distinction on the surface was not enough to dismiss the phonetic level hypothesis. The nonce word rating experiment showed that even if a phonetic difference is artificially introduced to what Hungarian speakers hear, this phonetic difference was not used by Hungarian listeners to categorize a novel stem as regular or antiharmonic. These results together led to the rejection of the phonetics first analysis of antiharmonicity. An alternative, lexical level analysis was presented using positively ranked spreading constraints inspired by Kimper (2011).

Chapter 3 tested the validity of the lexical level hypothesis of exceptionality in vowel harmony. More precisely, it tested whether speakers are aware of phonotactic regularities within sublexica defined based on the vowel harmony features of the stem's vowel. It was shown that Finnish speakers match of the probability of what kind of harmonicity they expect in novel stems to the frequencies found in the lexicon, confirming the lexical level hypothesis that exceptionality, even when fully grammatical as in mixed stems, is marked in the lexicon. However, not all patterns mattered equally to Finnish participants: consonantal effects on vowel harmony have been shown to be much weaker.

Finally, Chapter 4 discussed how the acquisition of the grammaticality of certain exceptional patterns like mixed stems and the low acceptability of other exceptional patterns like disharmony can be modelled computationally. A problem it encountered was the difference between static generalizations and dynamic behavior of stems (allomorph selection). The tools used to learn the former needed information that can only be inferred based on knowledge on the latter to perform well. This indicates that phonotactic learning must be extended so that the learning of higher level phenomena can have influence on it.

The sections below will discuss the evaluation of the main results of this dissertation in more depth. Section 5.1 will summarize the arguments against a phonetic level and for a lexical level analysis of exceptional patterns in vowel harmony. The implications of the computational modelling results will finally be summarized in Section 5.2.

5.1 Biases influencing exceptionality in vowel harmony

The main goal of this dissertation was to find out what kinds of biases matter and what kinds of biases do not really influence how exceptionality patterns in vowel harmony can be learned. In Chapter 1, two main sources of such bias were discussed: phonetic detail and sublexical phonotactic regularities. The case studies in Chapters 2 and 3 tested the details on how much these factors might influence speakers.

Table 5.1 below summarizes previous studies on exceptional vowel harmony patterns together with the results found in this dissertation. Based on these studies, the existence of a phonetic bias is controversial and it has been shown in Chapter 2 that while a minor segmental level difference might be present in some studies, speakers do not use such a bias when deciding on how to categorize a novel stem. Conversely, the presence of sublexical bias in vowel harmony exceptionality patterns has been clearly shown to influence speakers when categorizing nonce words, and there have not been any studies that found a lack of such an effect.

	arguments for	**arguments against**
phonetic bias	Benus and Gafos (2007): articulatory	Blaho and Szeredi (2013): pilot acoustic
	Benkő and van de Vijver (2016): acoustic on existing stems	Benkő and van de Vijver (2016): acoustic on novel stems
		Chapter 2 acoustic
		Chapter 2 perception
		Chapter 2 rating
sublexical bias	Hayes et al. (2009) on Hungarian	none
	Chapter 3 on Finnish	

Table 5.1: Results of experiments on biases used by speakers in exceptionality in vowel harmony patterns

To summarize the above, this dissertation argues against a phonetic level analysis of exceptionality in vowel harmony systems. It has shown that speakers did not use phonetic information in the input to categorize novel stems as regular or irregular, while it has shown, confirming the findings of Hayes et al. (2009), that information about sublexical phonotactic regularities are used by the speakers when they face such a nonce word categorization test.

Section 5.1.1 below will further discuss the conclusions of the results on phonetic bias and Section 5.1.2 below will present the conclusions on sublexical bias.

5.1.1 Phonetics first

As reviewed in the introductory chapter under Section 1.4.1, there has been an ongoing discussion in the literature on the effect of phonetic bias on phonological behavior, and particularly on vowel harmony. The goal of Chapter 2 was to test the plausibility of a phonetic level analysis of the Hungarian antiharmonic pattern. The first experiment sought out to verify the findings of Benus and Gafos (2007), that antiharmonic stems have a somewhat different vowel realization than regular harmonic stems. The results showed that there is indeed a very small, inconsistent, statistically not significant difference between the two kinds of vowels. The subsequent question to answer was whether Hungarian speakers are able to use such differences when categorizing novel stems as harmonic or antiharmonic.

First, the threshold of a perceivable difference between a retracted [i̠] and a full [i] had to be tested to set up the nonce word experiment, as well as to test whether Hungarian speakers have an enhanced perception for such differences. As the results suggested that a difference of 250 Hz in F2 was needed between the two vowels, no such enhanced perception can be argued for.

The nonce experiment explored whether a definitely perceivable 250 Hz difference is used by Hungarian speakers when determining how a novel stem selects affixes. The results, again, were negative: Hungarians did not change their categorization based on segmental level differences in the stem.

The results of these experiments in Chapter 2 point to a similar direction: the presence of a phonetic bias cannot be proven, and a phonetic first analysis of vowel harmony seems to be untenable. While one negative result might be questioned, and particular results might be ascribed to some kinds of task effects, there is a clear trend of negative results working against the phonetic first analysis.

5.1.2 Lexicon first

The presence of an alternative bias was also discussed in this dissertation: the effect of lexical phonotactic regularities affecting speakers' judgments on exceptional behavior in vowel harmony. These effects can be based on different interpretations of how sublexica are defined within in different ways based on the type of exceptionality. For the allomorph selection case, like the Hungarian antiharmonicity pattern, the sublexica affecting speakers' categorization are based on what allomorphs of subsequent affixes do given stems select: ones that are regular (harmonic) select front suffixes and antiharmonic stems select back suffixes.

Hayes et al. (2009) have already shown that phonotactic regularities in sublexica defined as such are used by Hungarian speakers when determining the allomorph selection after a novel (either mixed, or potentially antiharmonic) item. Chapter 3 tested whether the sublexical generalization extends to a classification of stems based on their static phonotactic probabilities instead of their allomorph selection. The results showed that speakers are aware phonotactic regularities present in the Finnish lexicon, and they base their judgments on whether to adapt a novel stem into the lexicon as mixed or harmonic by frequency matching using these phonotactic regularities.

It has also been shown by the studies presented in Chapter 3, that not all sublexical constraints are regarded equally by speakers when categorizing novel forms. Similarly to Hayes et al. (2009) and Becker et al. (2011), phonotactic interactions between segments that would be phonetically and phonologically unnatural were underlearned. For vowel harmony patterns, this means that the effects of sublexical patterns based on consonants in the stem matter much less for the speakers when making the decision to categorize than regularities based on the vowels in the stem: speakers are aware that vowel-to-vowel correspondences are more natural in vowel harmony than consonantal interference.

5.2 Learnability of static exceptionality

Having seen that sublexical biases are much more likely to affect how speakers learn exceptional patterns in vowel harmony systems, Chapter 4 provided a model on how speakers might learn the regularities of exceptional sublexica in a static pattern like Finnish mixed stems and disharmonicity.

The learning model employed maximum entropy grammars, so that the gradient nature of grammaticality judgment that had also been found in the wug study could be modeled. Static generalizations were modeled using the UCLA Phonotactic Learner, testing several conditions on how the training set was set up and what feature system was used. The task of learning the Finnish disharmonicity and the harmonicity pattern of frequent disyllables turned out to be quite difficult for the learner. The key to improve its performance was to include more phonological knowledge based on the Finnish allomorph selection pattern in the feature system: front vowels pattern into Front and Neutral vowels the same way in the dynamic and in static phonotactic facet of vowel harmony.

This knowledge was shown to be able to be bootstrapped given the affix vowel alternation pattern, using the Sublexical Learner. If the computational model is predictive of the learners' behavior, this means that learning how allomorphy works in Finnish must partially happen before, or simultaneously with the acquisition of static phonotactic generalizations.

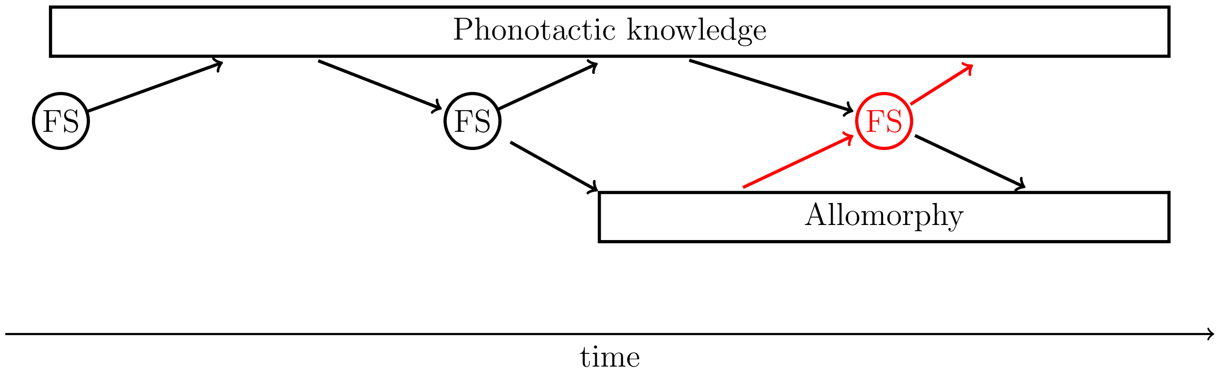


Figure 5.1: Modeling the influence of the acquisition allomorphy on learning phonotactics via a share feedback pattern (highlighted) through the feature system (FS), the abstract knowledge about the behavior phonemes in the language.

It is clear, though — based on acquisition studies — that acquiring phonotactics must start much earlier than the acquisition of higher level phenomena like allomorph selection. However, acquiring the more complicated vowel harmony patterns like the distribution and grammaticality of mixed and disharmonic stems might actually happen during, or after the learner has access to higher level information if the phonotactic learning does not stop after the acquisition of basic phonotactic regularities, but happens for a longer period, so that new knowledge can influence it later on. Figure 5.1 above illustrates this model, based on the computational results of Chapter 4: learning allomorphy can influence the abstract knowledge on the phonological inventory modeled by the feature system, which can feed back to facilitate further phonotactic learning.

To summarize the conclusions of the learning model in this dissertation, maxent grammars are able to model the acquisition of exceptional phonotactic patterns in vowel harmony. It is necessary, though, for the results of phonotactic learning and allomorphy learning to inform each other: without the phonological classification of vowels learned in allomorphy acquisition, the phonotactic learner performs on a much worse level.

Bibliography

Aaltonen, Olli, Osmo Eerola, Åke Hellström, Esa Uusipaikka and A. Heikki Lang (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *Journal of the Acoustical Society of America* **101**, pp. 1090–1105.

Alanne, Vieno Severi (1968). *Finnish-English general dictionary.* Werner Söderström Osakeyhitiö.

Albright, Adam (2002). Islands of reliability for regular morphology: Evidence from Italian. *Language* , pp. 684–709.

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

Allen, Blake and Michael Becker (2015). Learning alternations from surface forms with sublexical phonology. Manuscript submitted for publication. `http://ling.auf.net/lingbuzz/002503`.

Allen, Blake and Michael Becker (2016). The sublexical morphophonological learner. Web application, version of April 13, 2016. `http://sublexical.phonologist.org/`.

Anderson, Lloyd B. (1980). Using assymetrical and gradient data in the study of vowel harmony. In: Robert M. Vago (ed.), *Issues in Vowel Harmony: Proceedings of the CUNY Linguistics Conference on Vowel Harmony*, John Benjamins, pp. 271–330.

Arik, Engin (2015). An Experimental Study of Turkish Vowel Harmony. *Poznań Studies in Contemporary Linguistics* **53**, pp. 359–374.

Baer-Henney, Dinah, Frank Kügler and Ruben Vijver (2015). The Interaction of Language-Specific and Universal Factors During the Acquisition of Morphophonemic Alternations With Exceptions. *Cognitive Science* **39**, pp. 1537–1569.

Baković, Eric (2000). Harmony, dominance and control. Ph.D. thesis, Rutgers University. `http://roa.rutgers.edu/files/360-1199/roa-360-bakovic-2.pdf`.

Baković, Eric and Colin Wilson (2000). Transparency, strict locality and targeted constraints. In: *West Coast Conference on Formal Linguistics (WCCFL)*, vol. 19, pp. 43–56.

Baski, Imre (2009). Otuz yıl öncesi Deliorman ağızlarının şimdiki zaman ekleri üzerine. In: *Türkiye Türkçesi Ağız Araştırmaları Çalıştayı Bildirileri*, Türk Dil Kurumu Yayınları, Ankara, pp. 73–92.

Bates, Douglas, Martin Maechler, Ben Bolker and Steven Walker (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7, `http://CRAN.R-project.org/package=lme4`.

Becker, Michael and Maria Gouskova (2016). Source-oriented generalizations as grammar inference in Russian vowel deletion. *Linguistic Inquiry* **47**, pp. 391–425. `http://phonologist.org/projects/yers/`.

Becker, Michael and Jonathan Levine (2010). Experigen: an online experiment platform. `https://github.com/tlozoot/experigen`.

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

Becker, Michael, Lauren Eby Clemens and Andrew Nevins (forth.). Generalization of French and Portuguese Plural Alternations and Initial Syllable Protection. *Natural Language & Linguistic Theory* .

Beddor, Patrice Speeter, James D. Harnsberger and Stephanie Lindemann (2001). Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics* **30**, pp. 591–627.

Benkő, Ágnes and Reuben van de Vijver (2016). *A phonetic reflex of suffix in unsuffixed monosyllabic words with transparent vowels in Hungarian*. Poster presented at the 24th Manchester Phonology Meeting, Manchester, United Kingdom.

Benus, Stefan (2005). Dynamics and transparency in vowel harmony. Ph.D. thesis, New York University.

Benus, Stefan and Adamantios I. Gafos (2007). Articulatory characteristics of Hungarian 'transparent' vowels. *Journal of Phonetics* **35**, pp. 271–300.

Berko, Jean (1958). The child's learning of English morphology. *Word* **14**, pp. 150–177.

Binnick, Robert I. (1991). Vowel harmony loss in Uralic and Altaic. In: William G. Boltz and Michael C. Shapiro (eds.), *Studies in the Historical Phonology of Asian Languages*, John Benjamins, Amsterdam, vol. 77 of *Current Issues in Linguistic Theory*, pp. 35–52.

Birtalan, Ágnes (2003). Oirat. In: Juha Janhunen (ed.), *The Mongolic Languages*, Routledge Language Family Series, Routledge, New York, pp. 210–228.

Blaho, Sylvia and Daniel Szeredi (2013). Hungarian neutral vowels: a microcomparison. *Nordlyd* **40**, pp. 20–40.

Boersma, Paul and David Weenink (2012). *Praat: doing phonetics by computer. Version 5.3.16.* `http://www.praat.org`.

Boley, Jon and Michael Lester (2009). Statistical Analysis of ABX Results Using Signal Detection Theory. In: *Audio Engineering Society Convention 127*, Audio Engineering Society.

Boyce, Suzanne E. (1990). Coarticulatory organization for lip rounding in Turkish and English. *The Journal of the Acoustical Society of America* **88**, pp. 2584–2595.

Braver, Aaron and William G. Bennett (2016). Phonotactics of noun class disambiguation in Xhosa. In: Gunnar Ólafur Hansson, Ashley Farris-Trimble, Kevin McMullin and Douglas Pulleyblank (eds.), *Proceedings of the 2015 Annual Meeting on Phonology*, vol. 3.

Bybee, Joan L. and Carol Lynn Moder (1983). Morphological classes as natural categories. *Language* , pp. 251–270.

Callow, John C. (1965). Kasem nominals: a study in analyses. *Journal of West African Languages* **2**, pp. 29–36.

Campbell, Lyle (1980). The psychological and sociological reality of Finnish vowel harmony. In: Robert M. Vago (ed.), *Issues in Vowel Harmony*, Benjamins Amsterdam, pp. 245–269.

Casali, Roderic F. (2008). ATR harmony in African languages. *Language and Linguistics Compass* **2**, pp. 496–549.

Clements, George N. and Engin Sezer (1982). Vowel and consonant disharmony in Turkish **2**, pp. 213–255.

Demirci, Kerim and Zhanat Kuanysbaeva (2016). Kazakçanın ağızları ve özbekçenin çimkent ağzına etkisi üzerine. In: Mustafa Sarıca and Serap Sarıca (eds.), *Dil Bilimleri Kültür ve Edebiyat*, Padam Yayınları, Ankara, pp. 330–343. http://www.academia.edu/download/45061357/pdm-kazakcaagizlar.pdf.

Duncan, Liisa (2008). Vowel Quality in Finnish Loanwords of Swedish Origin: An Acoustic Study . Generals Paper. University of Toronto.

E. Abaffy, Erzsébet (2004a). Hangtörténet. (Az ómagyar kor) [Phonology of the Old Hungarian period]. In: Kiss and Pusztai (2004), pp. 301–351.

E. Abaffy, Erzsébet (2004b). Hangtörténet. (Az ősmagyar kor) [Phonology of the Proto-Hungarian period]. In: Kiss and Pusztai (2004), pp. 106–128.

Ernestus, Mirjam and Harald Baayen (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* **79**, pp. 5–39.

Ernestus, Mirjam and Harald Baayen (2006). The functionality of incomplete neutralization in Dutch. the case of past-tense formation. In: Louis M. Goldstein, Douglas H. Whalen and Catherine T. Best (eds.), *Papers in Laboratory Phonology 8*, Mouton de Gruyter, Berlin/New York, pp. 27–49.

Finley, Sara (2008). Formal and cognitive restrictions on vowel harmony. Ph.D. thesis, Johns Hopkins University.

Finley, Sara (2012). Typological asymmetries in round vowel harmony: Support from artificial grammar learning. *Language and Cognitive Processes* **27**, pp. 1550–1562.

Finley, Sara (2015). Learning nonadjacent dependencies in phonology: Transparent vowels in vowel harmony. *Language* **91**, pp. 48–72.

Finley, Sara and William Badecker (2012). Learning biases for vowel height harmony. *Journal of Cognitive Science* **13**, pp. 287–327.

Fowler, Carol A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General* **112**, p. 386.

Gafos, Adamantios I. (1999). *The Articulatory Basis of Locality in Phonology.* Outstanding Dissertations in Linguistics, Garland.

Goldsmith, John (1985). Vowel Harmony in Khalkha Mongolian, Yaka, Finnish and Hungarian. *Phonology Yearbook* , pp. 253–275.

Goldsmith, John and Jason Riggle (2012). Information theoretic approaches to phonological structure: the case of Finnish vowel harmony. *Natural Language & Linguistic Theory* **30**, pp. 859–896.

Gouskova, Maria (2012). Unexceptional segments. *Natural Language & Linguistic Theory* **30**, pp. 79–133.

Gouskova, Maria and Suzy Ahn (2016). Sublexical phonotactics and English comparatives. Ms. New York University.

Gouskova, Maria and Michael Becker (2013). Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language & Linguistic Theory* **31**, pp. 735–765.

Gouskova, Maria and Tal Linzen (2015). Morphological conditioning of phonological regularization. *The Linguistic Review* **32**, pp. 427–473.

Gouskova, Maria, Luzia Newlin-Łukowicz and Sofya Kasyanenko (2015). Selectional restrictions as phonotactics over sublexicons. *Lingua* **167**, pp. 41–81.

Hakulinen, Auli, Riitta Korhonen, Maria Vilkuna and Vesa Koivisto (2004). *Iso suomen kielioppi. [Large Finnish grammar].* Suomalaisen kirjallisuuden seura.

Hansson, Gunnar Ólafur (2008). Diachronic explanations of sound patterns. *Language and Linguistics Compass* **2**, pp. 859–893.

Hansson, Gunnar Ólafur (2010). *Consonant Harmony: Long-Distance Interaction in Phonology.* UC Publications in Linguistics, University of California Press, Berkeley, CA.

Harris, John (1994). Monovalency and opacity: Chichewa height harmony. In: *UCL Working Papers in Linguistics*, vol. 6, pp. 509–547.

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

Hayes, Bruce, Robert Kirchner and Donca Steriade (eds.) (2004). *Phonetically Based Phonology*. Cambridge University Press, Cambridge.

Hayes, Bruce, Kie Zuraw, Péter Siptár and Zsuzsa Londe (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language* **85**, pp. 822–863.

Jäger, Gerhard (2007). Maximum Entropy models and Stochastic Optimality Theory. In: Annie Zaenen, Jane Simpson, Tracy Holloway King, Joan Maling, Jane Grimshaw and Chris Manning (eds.), *Architectures, rules, and preferences: Variations on themes by Joan W. Bresnan*, CSLI, Stanford, CA, pp. 467–479.

Jun, Jongho and Jeehyun Lee (2007). Multiple stem-final variants in Korean native nouns and loanwords. *Journal of the Linguistic Society of Korea* **47**, pp. 159–187.

Kardestuncer, Aino (1983). Vowel harmony and gerundive compounds in Turkish. *Acta Linguistica Hafniensia* **18**, pp. 55–64.

Kastner, Itamar (2016). Form and Meaning in the Hebrew Verb. Ph.D. thesis, New York University.

Kaun, Abigail (1995). An Optimality-Theoretic typology of rounding harmony. Ph.D. thesis, UCLA.

Kaun, Abigail R. (2004). The typology of rounding harmony. In: Hayes et al. (2004), pp. 87–116.

Kendall, Tyler and Erik R. Thomas (2012). *vowels: Vowel Manipulation, Normalization, and Plotting*. R package, version 1.2, `http://CRAN.R-project.org/package=vowels`.

Kewley-Port, Diane (2001). Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context, and training. *Journal of the Acoustical Society of America* **110**, pp. 2141–2155.

Kimper, Wender (2011). Competing triggers: transparency and opacity in vowel harmony. Ph.D. thesis, University of Massachusetts Amherst.

Kiparsky, Paul and Karl Pajusalu (2003). Towards a typology of disharmony. *The Linguistic Review* **20**, pp. 217–242.

Kis, Tamás (2005). A veláris i̯ a magyarban. [The velar i̯ in Hungarian]. In: Tamás Hoffman, Tamás Kis and István Nyirkos (eds.), *Magyar Nyelvjárások [Hungarian Dialects]*, Debreceni Egyetemi Kiadó, Debrecen, vol. XLIII, pp. 5–26.

Kiss, Jenő (ed.) (2001). *Magyar dialektológia. [Hungarian dialectology]*. Osiris, Budapest.

Kiss, Jenő and Ferenc Pusztai (eds.) (2004). *Magyar nyelvtörténet. [The history of Hungarian]*. Osiris, Budapest.

Korn, David (1969). Types of Labial Vowel Harmony in the Turkic Languages. *Anthropological Linguistics* **11**, pp. 98–106.

Kornai, András, Péter Halácsy, Viktor Nagy, Csaba Oravecz, Viktor Trón and Dániel Varga (2006). Web-based frequency dictionaries for medium density languages. In: Adam Kilgarriff and Marco Baroni (eds.), *Proceedings of the 2nd International Workshop on Web as Corpus*, pp. 1–9.

Kuhn, Jeremy (2012). Harmony via Positive Agreement: Evidence from trigger-based count effects. In: Hsin-Lun Huang, Ethan Poole and Amanda Rysling (eds.), *NELS 43: Proceedings of the 43rd Meeting of the North East Linguistic Society*, vol. 1, pp. 253–264.

Labov, William (1971). Methodology. In: William Orr Dingwall (ed.), *A survey of linguistic science*, University of Maryland Linguistics Program, College Park, MD, pp. 412–497.

Labov, William, Malcah Yaeger and Richard Steiner (1972). *A quantitative study of sound change in progress*, vol. 1. US Regional Survey.

Labov, William, Mark Caren and Corey Miller (1991). Near-mergers and the suspension of phonemic contrast. *Language Variation and Change* **3**, pp. 33–74.

Levi, Susannah V. (2002). Glides, laterals, and Turkish vowel harmony. In: Mary Andronis, Christopher Ball, Heidi Elston and Sylvain Neuval (eds.), *Proceedings from the 37th meeting of the Chicago Linguistics Society*, Chicago Linguistics Society, Chicago, pp. 379–394.

Lewis, Geoffrey L. (1967). *Turkish language*. Clarendon Press, Oxford.

Lightner, Theodore (1972). *Problems in the theory of phonology*. Linguistic Research Inc., Edmonton.

Lindén, Krister, Miikka Silfverberg and Tommi Pirinen (2009). HFST tools for morphology – an efficient open-source package for construction of morphological analyzers. In: *State of the Art in Computational Morphology*, Springer, Berlin, pp. 28–47. `http://hfst.sourceforge.net`.

Linebaugh, Gary Dean (2007). Phonetic Grounding and Phonology: Vowel Backness Harmony and Vowel Height Harmony. Ph.D. thesis, University of Illinois at Urbana-Champaign.

Linzen, Tal, Sofya Kasyanenko and Maria Gouskova (2013). Lexical and phonological variation in Russian prepositions. *Phonology* **30**, pp. 453–515.

Macmillan, Neil A. and C. Douglas Creelman (1991). *Detection theory: A user's guide.* Cambridge University Press, Cambridge.

Mailhot, Frédéric and Charles Reiss (2007). Computing long-distance dependencies in vowel harmony. *Biolinguistics* **1**, pp. 28–48.

McCarthy, John (2009). Harmony in Harmonic Serialism. Ms., University of Massachusetts, Amherst. Available as ROA-1009 at www.roa.rutgers.edu.

Mermelstein, Paul (1978). Difference limens for formant frequencies of steady-state and consonant-bound vowels. *Journal of the Acoustical Society of America* **63**, pp. 572–580.

Miller, Jeff (1996). The sampling distribution of d'. *Perception & Psychophysics* **58**, pp. 65–72.

Milroy, James and John Harris (1980). When is a merger not a merger?: the MEAT/MATE problem in a present-day English vernacular. *English World-Wide* **1**, pp. 199–210.

Moreton, Elliott (2008). Analytic bias and phonological typology. *Phonology* **25**, p. 83.

Mtenje, Al D. (1985). Arguments for an autosegmental analysis of Chichewa vowel harmony. *Lingua* **66**, pp. 21–52.

Nádasdy, Ádám and Péter Siptár (1994). A magánhangzók [The vowels]. In: Ferenc Kiefer (ed.), *Strukturális Magyar Nyelvtan 2. Fonológia* [Structural Hungarian Grammar 2. Phonology], Akadémiai Kiadó, Budapest, pp. 42–182.

Nevins, Andrew (2004). Conditions on (Dis)Harmony. Ph.D. thesis, MIT.

Öhman, Sven (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America* **39**, pp. 151–168.

O'Keefe, Michael (2004). Akan vowel harmony. Bachelor's thesis, Swarthmore College.

Pater, Joe (2010). Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In: Steve Parker (ed.), *Phonological argumentation: Essays on evidence and motivation*, Equinox, London, pp. 123–154.

Pater, Joe and Elliott Moreton (2012). Structurally biased phonology: complexity in learning and typology. *Journal of the English and Foreign Languages University, Hyderabad* **3**, pp. 1–41.

Pirinen, Tommi (2014). *Open Source Finnish Morphology (OMorFi)*. Department of Modern Languages, University of Helsinki. `http://code.google.com/p/omorfi/`.

Polgárdi, Krisztina (1998). Vowel Harmony. An account in terms of government and optimality. Ph.D. thesis, Universiteit Leiden, Den Haag.

Port, Robert F. and Penny Crawford (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics* **17**, pp. 257–282.

Port, Robert F. and Michael L. O'Dell (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics* **13**, pp. 455–471.

Prasada, Sandeep and Steven Pinker (1993). Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes* **8**, pp. 1–56.

Prince, Alan and Paul Smolensky (1993/2004). *Optimality Theory: Constraint interaction in generative grammar*. Blackwell.

Puthuval, Sarala (2013). Loanwords, prominence and the basis for Mongolian vowel harmony. Master's thesis, University of Washington.

Rebrus, Péter and Miklós Törkenczy (2005). Uniformity and contrast in the Hungarian verbal paradigm. In: Laura J. Downing, T. Alan Hall and Renate Raffelsiefen (eds.), *Paradigms in phonological theory*, Oxford University Press, Oxford, pp. 263–295.

Rhodes, Russell (2010). A Survey of Vowel Harmony with Special Focus on Neutral Segments, Glides, and Diphthongs. Dissertation Prospectus. University of California, Berkeley.

Ringen, Catherine O. and Orvokki Heinämäki (1999). Variation in Finnish vowel harmony: An OT account. *Natural Language & Linguistic Theory* **17**, pp. 303–337.

Ringen, Catherine O. and Robert M. Vago (1998). Hungarian vowel harmony in Optimality Theory. *Phonology* **15**, pp. 393–416.

Sammallahti, Pekka (1988). Historical Phonology of the Uralic Languages, with Special Reference so Samoyed, Ugric, and Permic. In: Denis Sinor (ed.), *The Uralic Languages. Description, History and Foreign Influences*, pp. 478–554.

Seongyeon, Ko (2011). Vowel contrast and vowel harmony shift in the Mongolic languages. *Language Research* **47**, pp. 23–43.

Siptár, Péter and Miklós Törkenczy (2000). *The phonology of Hungarian.* Oxford University Press, Oxford.

Stanislaw, Harold and Natasha Todorov (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers* **31**, pp. 137–149.

Suomi, Kari (1983). Palatal vowel harmony: A perceptually motivated phenomenon? *Nordic Journal of Linguistics* **6**, pp. 1–35.

Suomi, Kari, James M. McQueen and Anne Cutler (1997). Vowel Harmony and Speech Segmentation in Finnish. *Journal of Memory and Language* **36**, pp. 422–444.

Suomi, Kari, Juhani Toivanen and Riikka Ylitalo (2008). *Finnish sound structure: phonetics, phonology, phonotactics and prosody.* Studia Humaniora Ouluensia 9, University of Oulu, Oulu, Finland.

Svantesson, Jan-Olof, Anna Tsendina, Anastasia Karlsson and Vivan Franzén (2005). *The Phonology of Mongolian*. Oxford University Press.

Szeredi, Daniel (2012). Possible theoretical relevance of subphonemic vowel reduction in Hungarian. In: Camelia Constantinescu, Bert Le Bruyn and Kathrin Linke (eds.), *Proceedings of ConSOLE XVII*, pp. 299–319.

Trudgill, Peter (1974). *The social differentation of English in Norwich*. Cambridge University Press, Cambridge.

Vago, Robert M. (1980). *The sound pattern of Hungarian*. Georgetown University Press.

Vroomen, Jean, Jyrki Tuomainen and Beatrice de Gelder (1998). The Roles of Word Stress and Vowel Harmony in Speech Segmentation. *Journal of Memory and Language* **38**, pp. 133–149.

Warner, Natasha, Allard Jongman, Joan Sereno and Rachèl Kemps (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *Journal of Phonetics* **32**, pp. 251–276.

Yu, Alan C. L. (2007). Understanding near mergers: The case of morphological tone in cantonese. *Phonology* **24**, pp. 187–214.

Zimmer, Karl E (1967). A note on vowel harmony. *International Journal of American Linguistics* **33**, pp. 166–171.

Zséder, Attila, Gábor Recski, Dániel Varga and András Kornai (2012). Rapid creation of large-scale corpora and frequency dictionaries. In: Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk and Stelios Piperidis (eds.), *Proceedings of the Eight*

International Conference on Language Resources and Evaluation (LREC12), European Language Resources Association (ELRA), Istanbul, Turkey, pp. 1462–1465.