Title
Distance-based decay in long-distance phonological processes: a probabilistic model for Malagasy, Latin, English, and Hungarian

Author
Zymet, Jesse

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Distance-based decay in long-distance phonological processes:

a probabilistic model for Malagasy, Latin, English, and Hungarian

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Arts

in Linguistics

by

Jesse Adam Zymet

2014

ABSTRACT OF THE THESIS

Distance-based decay in long-distance phonological processes:

a probabilistic model for Malagasy, Latin, English, and Hungarian

by

Jesse Adam Zymet

Master of Arts in Linguistics

University of California, Los Angeles, 2014

Professor Bruce Hayes (Co-Chair)

Professor Kie Zuraw (Co-Chair)

Many—or perhaps all—long-distance assimilatory and dissimilatory phonological processes produce lexical or free variation exhibiting a broad generalization: the likelihood of process application decreases as the transparent distance between the trigger and the target increases (a phenomenon that I call *distance-based decay*). This thesis provides a unified analysis of distance-based decay, drawing from thousands of data reflecting three long-distance phonological processes across four languages. I account for the data within the framework of Maximum Entropy Harmonic Grammar (Smolensky 1986, Goldwater and Johnson 2003, Hayes and Wilson 2008), which allows for an adequate treatment of variation. I argue that distance-based decay can be captured across the four surveyed languages by an invariant decay

function—a simple negative power function—that interacts with two language-specific parameters: the weight of the AGREE (for assimilatory cases) or DISAGREE (for dissimilatory cases) constraint and the weight of IDENT. My account is therefore an extension of Kimper 2011, who posits a scaling factor that scales the weight of markedness to account for the decay effect present in vowel harmony in Hungarian. While it is the case that decay rates differ across the languages I survey, my analysis accounts for such differences purely with the weights of markedness and faithfulness; that is, I show that differences in decay rate can be modeled without having to posit language-specific decay functions.

The thesis of Jesse Adam Zymet is approved.

Robert Daland

Ed Stabler

Bruce Hayes, Committee Co-Chair

Kie Zuraw, Committee Co-Chair

University of California, Los Angeles

2014

This thesis is dedicated in loving memory to my father,

Mark Jay Zymet.

TABLE OF CONTENTS

LIST OF FIGURES

List of Tables

ACKNOWLEDGMENTS

I am deeply indebted to the thesis co-chairs, Bruce Hayes and Kie Zuraw, for many, many hours of discussion, draft suggestions, abstract submission suggestions, and presentation suggestions. I could not ask for better advisors for this project, and could only hope to one day become even half the phonolo-giants that they are. Many thanks to my other committee members, Robert Daland and Ed Stabler, for their keen insights and commentary. Many more thanks to audience members at the Phonology Seminar at UCLA, Phlunch at UC Santa Cruz, and WCCFL 32 at the University of Southern California.

Even more thanks to my linguistics friends for plentiful spiritual and emotional support, and for great times out on the town. All work and no play makes Jesse a dull boy—you guys make my experience at UCLA an awesome one. I dare say the raucous nights in Koreatown with Margit Bowler and Phil Côté-Boucher actually led to the betterment of this thesis.

This project has been in the works for a while now. As it can sometimes be challenging for me to keep my mouth shut, I am further indebted to my family and non-linguistic friends for cheering me on throughout this long process while putting up with blabber about "distance-based decay", "*k*-basins", "decay functions", "transparent syllables", "opaque segments", and "Maximum Entropy Harmonic Grammar". They must be bursting with excitement over what kinds of discussion I have in store for them for when I begin work on my dissertation.

Finally, I am grateful to have been supported by a UCLA Graduate Summer Research Fellowship as well as a UCLA Graduate Research Fellowship while working on this project.

Distance-based decay in long-distance phonological processes: a probabilistic model for Malagasy, Latin, English, and Hungarian

Jesse Zymet
University of California, Los Angeles

1 Introduction

Many assimilatory and dissimilatory phonological processes are classified as long-distance, taking the following form:

(1) /AXB/ → [AXB′] where A is the trigger, X is transparent material, and B and B′ are respective underlying and surface forms of the target

Several—or perhaps all—of these processes produce either lexical or free variation or both. At the statistical level, the variation appears to exhibit a broad generalization: the likelihood of process application decreases as the length of X increases (a phenomenon that I call *distance-based decay*). That is, the grammar regulates the trigger and the target more reliably when they are closer together than when they are farther apart.

This thesis provides a unified analysis of distance-based decay, drawing from thousands of data reflecting three long-distance phonological processes across four languages. I account for the data within the framework of Maximum Entropy Harmonic Grammar (Smolensky 1986, Goldwater and Johnson 2003, Hayes and Wilson 2008), which allows for an adequate treatment of variation. I argue that distance-based decay can be captured across the four surveyed languages by an invariant decay function—a simple negative power function—that interacts with two language-specific parameters: the weight of the AGREE (for assimilatory cases) or DISAGREE (for dissimilatory cases) constraint and the weight of IDENT. My account is therefore an extension of Kimper 2011, which posits a scaling factor that scales the weight of markedness to account for the decay effect present in vowel harmony in Hungarian. While it is the case that decay rates differ across the languages I survey, my analysis captures such differences purely with the weights of markedness and faithfulness; that is, I show that differences in decay rate can be modeled without having to posit language-specific decay functions. The decay function takes as its argument a measure of transparent distance and returns a value that is then multiplied by weight of the AGREE or DISAGREE constraint.

The thesis is organized as follows. Section 1.1 gives a brief review of the current literature on distance-based decay. Section 2 gives an empirical background for the decay effect. Section 3 incorporates the decay function into Maximum Entropy Harmonic Grammar and demonstrates how the function models the decay effect in long-distance phonological processes. Section 4 determines a satisfactory unit of transparent distance to serve as input to the model and finds parameter values that maximize model accuracy. Section 5 concludes.

1.1 Distance-based decay in the literature

Distance-based decay has been observed in a growing body of literature. Walker and Mpiranya 2006 observe a binary distinction in coronal harmony in Kinyarwanda and in retroflex harmony in Sanskrit, showing that both processes apply obligatorily in local environments, while optionally in nonlocal environments. Martin 2005 shows that syllabic distance—but not segmental distance—between the trigger and the target is significant in producing a decay effect in sibilant harmony reflected in Navajo compounds; in addition, Martin observes a distinction between local and nonlocal application, showing that sibilant harmony applies obligatorily when the trigger and the target are in adjacent syllables but optionally when they are separated by one or more transparent syllables. Finally, Hayes and Londe 2006 observe a three-way distinction based on transparent distance, showing that vowel harmony in Hungarian applies most reliably when the harmonizing vowels are in adjacent syllables, less reliably when they are separated by one neutral vowel, and even less reliably when they are separated by two neutral vowels. Taken together, these findings suggest that distance-based decay is a general phenomenon emerging within data that reflect a variety of different processes and not just one in particular.

One way in which phonologists account for distance-based decay is through positing distance-based markedness constraints (Hansson 2001, Martin 2005, Hayes and Londe 2006), which penalize trigger-target pairs at different distances. Martin 2005, for example, accounts for sibilant harmony in Navajo compounds within the framework of Stochastic Optimality Theory (Boersma 1997) by positing a constraint that regulates sibilant harmony when the sibilants are in adjacent syllables as well as a constraint that regulates sibilant harmony when they are in nonadjacent syllables. The local markedness constraint has a ranking value higher than that of the nonlocal markedness constraint, thus enforcing the decay effect.

Kimper 2011 proposes an alternative to the constraint-family account within the framework of Harmonic Grammar (Smolensky and Legendre 2006): posit a single markedness constraint that regulates the cooccurrence restriction both locally and nonlocally, and scale down the weight of the constraint with increasing distance. In particular, Kimper accounts for the decay effect arising in Hungarian vowel harmony through scaling the weight of a SPREAD constraint by multiplying it x times over with a constant, $k \in (0,1)$ when there are x transparent moras. Effectively, the weight of SPREAD in such case becomes $w(\text{SPREAD}) * k^x$, which is then compared against the weight of IDENT to yield faithful forms at longer distances. As we will see, I opt to account for the decay effect using Kimper's method of scaling the weight of markedness.

We now turn to an exposition of the data surrounding distance-based decay in three long-distance phonological processes across four languages.

2 Empirical background: long-distance phonological processes with distance-based decay

Distance-based decay is evident in long-distance vowel assimilation and dissimilation and long-distance consonant assimilation and dissimilation. This section covers distance-based decay as it occurs in three such processes across four languages: vowel dissimilation in Malagasy, liquid dissimilation in Latin and English, and vowel harmony in Hungarian.

Malagasy, a Western Austronesian language spoken primarily in Madagascar, has a vowel inventory composed of four vowels:

	[-back]	[+back]
[+high]	/i/	/u/
[-high -low]	/e/	
[+low]	/a/[1]	

Table 1: *Malagasy vowel inventory*

Distance-based dissimilation of rounding on the high vowels (hereafter referred to as labial dissimilation) can be observed in data from Beaujardière 2004:

	UR	Passive imperative verb form	Gloss
Faithful items			
(2a)	/bata+u/	[bata-u]	'lift'
(2b)	/fana+u/	[fana-u]	'heat'
Items undergoing local and nonlocal vowel dissimilation			
(3a)	/bab**u**+**u**/	[bab**u**-**i**], *[bab**u**-**u**]	'plunder'
(3b)	/t**u**v+**u**/	[t**u**v-**i**], *[t**u**v-**u**]	'fulfill'
(3c)	/t**u**da+**u**/	[t**u**da-**i**], *[t**u**da-**u**]	'prevent'
(3d)	/g**u**rabah+**u**/	[g**u**rabah-**i**], *[g**u**rabah-**u**]	'spluttering'
Items with opaque front vowels			
(4a)	/**turi**+**u**/	[**turi**-**u**], *[**turi**-**i**]	'preach'
(4b)	/**ure**+**u**/	[**ure**-**u**], *[**ure**-**i**]	'massage'

[1] /a/ is in fact taken to be central (Nordhoff et al. 2013).

(2a) and (2b) reveal the underlying form of the passive imperative suffix to be -/u/. (3a) shows that if the passive imperative suffix attaches to a stem that ends in /u/, then it dissimilates, surfacing instead as -[i].[2] This can be seen as an effect arising from the O(bligatory) C(ontour) P(rinciple) (Goldsmith 1976, McCarthy 1986), which states that adjacent segments bearing identical features are prohibited. (3b) through (3d) show that labial dissimilation can apply across transparent material, suggesting that OCP can regulate segments even at a distance. /a/ is therefore transparent to labial dissimilation, while (4a) and (4b) show that /e/ and /i/ are opaque to the process. Labial dissimilation can thus be represented as follows:

(5) Labial Dissimilation:

$u \rightarrow i \; / u(C_0 a)_0 C_0 + \_\_$
suffix /u/ *becomes* -[i] *after stem* /u/ *if nothing but consonant clusters (possibly having zero length) or* [a] *come in between* /u/*-segments.*

The rule in (5) is inadequate in that it does not predict the existence of a decay effect produced by greater amounts of transparent material. Consider cases where stem-internal /u/ is separated from suffix -/u/ by growing numbers of transparent syllables:

	Transparent Syllables	UR	Passive imperative verb form	Gloss
(6a)	0 syllables	/ba.b**u**.+**u**/	[ba.b**u**.-**i**]	'plunder'
(6b)	0 syllables	/t**u**.v+**u**/	[t**u**.v-**i**]	'fulfill'
(6c)	0 syllables	/d**u**.r+**u**/	[d**u**.r-**i**]	'burn'
(6d)	1 syllable	/r**u**.va.+**u**/	[r**u**.va.-**u**]	'palisade'
(6e)	1 syllable	/**u**n.da.n+**u**/	[**u**n.da.n-**i**]	'bolster'
(6f)	1 syllable	/t**u**.da.+**u**/	[t**u**.da.-**i**]	'prevent'
(6g)	2 syllables	/b**u**.ra.ra.h+**u**/	[b**u**.ra.ra.h-**u**]	'scattered'
(6h)	2 syllables	/k**u**.ta.ba.+**u**/	[k**u**.ta.ba.-**u**]	'disorder'
(6i)	2 syllables	/g**u**.ra.ba.h+**u**/	[g**u**.ra.ba.h-**i**]	'spluttering'

[2] As far as I am aware, the passive imperative suffix -/u/ is the only suffix to which labial dissimilation applies.

As the number of transparent syllables increases, the likelihood of labial dissimilation decreases. This generalization is exhibited by a set of forms extracted from Beaujardière 2004, an online Malagasy dictionary. The figures are provided below:

Vowel dissimilation in Malagasy: /**u**σ^n+**u**/ → [**u**σ^n-**i**]			
Transparent syllables (σ)	Faithful forms	Dissimilated forms	Proportion of dissim'd forms
n = 0	4	989	0.99
n = 1	196	201	0.51
n = 2	28	4	0.13
n = 3	4	0	0.00

Table 2: *figures for distance-based decay in labial dissimilation in Malagasy*

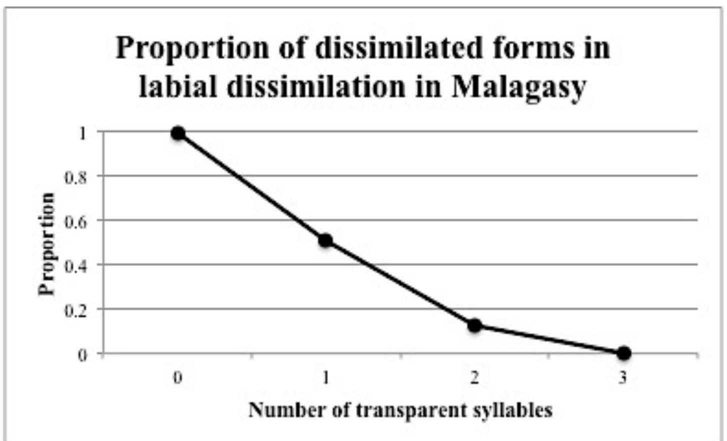


Figure 1: *graph of distance-based decay in labial dissimilation in Malagasy*

In forms with zero transparent syllables, the trigger and target are either adjacent segmentally or separated only by consonants. Local labial dissimilation is nearly categorical with 99.6% of the forms undergoing the process. In nonlocal settings where the trigger and target are separated by a positive number of transparent syllables (i.e., syllables that have /a/ in the nucleus), labial dissimilation displays distance-based decay: application is variable, with rates decreasing as the number of transparent syllables increases. Note that these figures do not include forms in which the opaque vowel /i/ comes in between the trigger and target.

Data extracted from the Perseus Digital Library (http://www.perseus.tufts.edu/hopper/) reveal that Latin as well contains a long-distance phonological process—liquid dissimilation—that regulates the distribution of [l] and [r] in particular contexts:

	UR	Adjective form	Gloss
Faithful items			
(7a)	/kib+a:lis/	[kib-a:lis]	'pertaining to food'
(7b)	/fanit+a:lis/	[fanit-a:lis]	'pertaining to a temple'
Items undergoing local and nonlocal liquid dissimilation			
(8a)	/so**l**+a:**l**is/	[so**l**-a:**r**is]	'pertaining to the sun'
(8b)	/wu**l**g+a:**l**is/	[wu**l**g-a:**r**is]	'pertaining to wheat'
(8c)	/**l**a:n+a:**l**is/	[**l**a:n-a:**r**is]	'pertaining to wool'
(8d)	/**l**apid+a:**l**is/	[**l**apid-a:**r**is]	'pertaining to rocks'
Items with opaque /r/			
(9a)	/**l**ito**r**+a:**l**is/	[**l**ito**r**-a:**l**is]	'pertaining to the seashore'
(9b)	/sepu**l**k**r**+a:**l**is /	[sepu**l**k**r**-a:**l**is]	'pertaining to a tomb'

Latin has an adjectival suffix, -/a:lis/, whose underlying form is apparent from (7a) and (7b). If the suffix attaches to a stem ending in /l/, then it dissimilates, surfacing instead as -[a:ris], as in (8a).[3] (8b) through (8d) show that stem-internal /l/ need not be stem-final in order to trigger liquid dissimilation on -/a:lis/. (9a) and (9b) show that /r/ blocks liquid dissimilation.[4]

The following data illustrate how the process is sensitive to increasing distance:

	Transparent Syllables	UR	Adjective form	Gloss
(10a)	0 syllables	/so.**l**+a:.**l**is/	[so.**l**-a:.**r**is]	'pertaining to the sun'
(10b)	0 syllables	/mu.**l**+a:.**l**is/	[mu.**l**-a:.**r**is]	'pertaining to mules'
(10c)	0 syllables	/du.p**l**+a:.**l**is/	[du.p**l**-a:.**r**is]	'pertaining to two'

[3] Bennett 2012 notes that only -/a:lis/ undergoes dissimilation, as opposed to other /l/-bearing suffixes such as -/ilis/ or -/ulus/ (e.g., /kalk+ulus/ → [kalk-ulus], 'small stone').

[4] My corpus has two exceptions to opaque /r/: [palpebr-a:lis] and [lucern-a:lis] can be optionally pronounced [palpebr-a:ris] and [lucern-a:ris] respectively.

(10d) 1 syllable /pa.**le**.+aː.**lis**/ [pa.**le**.-aː.**lis**] ‘pertaining to chaff’
(10e) 1 syllable /**la**ː.n+aː.**lis**/ [**la**ː.n-aː.**ris**] ‘pertaining to wool’
(10f) 1 syllable /a.**le**.+aː.**lis**/ [a.**le**.-aː.**ris**] ‘pertaining to chance’

(10g) 2 syllables /**lek**.tu.+aː.**lis**/ [**lek**.tu.-aː.**lis**] ‘pertaining to beds’
(10h) 2 syllables /di.**lu**.wi.+aː.**lis**/ [di.**lu**.wi.-aː.**lis**] ‘pertaining to floods’
(10i) 2 syllables /**la**.pi.d+aː.**lis**/ [**la**.pi.d-aː.**ris**] ‘pertaining to stone’

As is the case with labial dissimilation in Malagasy, separating candidate forms for liquid dissimilation based on the number of transparent syllables reveals that the process is subject to distance-based decay. The figures are shown below:

Liquid dissimilation in Latin: /**l**σ^n+aː.**l**is / → [**l**σ^n-aː.**r**is]			
Transparent syllables (σ)	Faithful forms	Dissimilated forms	Proportion of dissim’d forms
$n = 0$	0	131	1.00
$n = 1$	20	49	0.71
$n = 2$	29	13	0.31
$n = 3$	4	0	0.00

Table 3: *figures for distance-based decay in liquid dissimilation in Latin*

Application rate of liquid dissimilation decreases as number of transparent syllables increases:

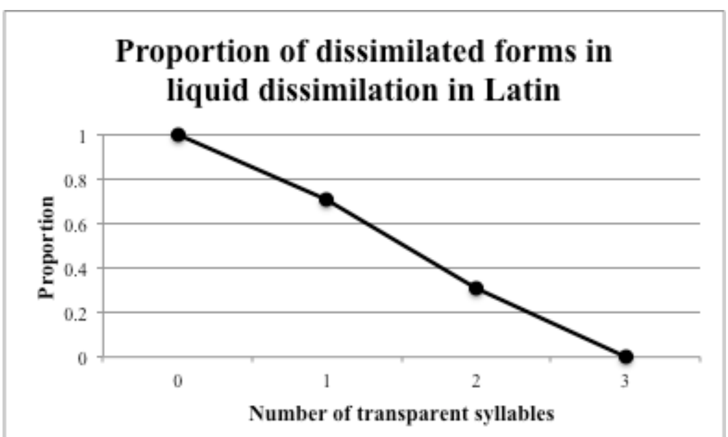


Figure 2: *graph of distance-based decay in liquid dissimilation in Latin*

Data extracted from the O(xford) E(nglish) D(ictionary) (www.oed.com) show that English displays long-distance consonant dissimilation in words borrowed from Latin:

	UR	Adjective form	Gloss
Faithful items			
(11a)	/dɪst+əl/	[dɪst-əl]	'distal'
(11b)	/ejpɪk+əl/	[ejpɪk-əl]	'apical'
Items displaying local liquid dissimilation			
(12a)	/soʊ**l**+ə**l**/	[soʊ**l**-ə**ɹ**]	'solar'
(12b)	/vi**l**+ə**l**/	[vi**l**-ə**ɹ**]	'velar'
(12c)	/kɑndə**l**+ə**l**/	[kɑndə**l**-ə**ɹ**]	'condylar'
Items showing nonlocal liquid dissimilation			
(12d)	/**l**un+ə**l**/	[**l**un-ə**ɹ**]	'lunar'
(12e)	/**l**əkjun+ə**l**/	[**l**əkjun-ə**ɹ**]	'lacunar'
Items with opaque /ɹ/			
(13a)	/**flɔɹ**+ə**l**/	[**flɔɹ**-ə**l**]	'floral'
(13b)	/æ**l**pɛst**ɹ**+ə**l**/	[æ**l**pɛst**ɹ**-ə**l**]	'alpestral'

The underlying form English adjectival suffix, -/əl/, is revealed in (11a) and (11b). As is the case in Latin, if the adjectival suffix attaches to a stem ending in /l/, then it undergoes liquid dissimilation, surfacing as -[əɹ], as shown in (12a) through (12c). (12d) and (12e) show that liquid dissimilation can apply even when the trigger and target are distant from one another. (13a) and (13b) show that /ɹ/ blocks the process.

Liquid dissimilation in English is not simply the relic of borrowing words from Latin. The OED contained 527 candidate forms, and of these, 125 of them had corresponding Latin forms in the Perseus Online Dictionary. Of these 125 forms, twelve of them underwent dissimilation in one language but not the other. Six underwent dissimilation in Latin but not in English, and six underwent dissimilation in English but not in Latin.

The following data show that liquid dissimilation in English is as well subject to distance-based decay:

	Trigger-target distance	UR	Adjective form	Gloss
(14a)	Same syllable	/soʊ**l**+ə**l**/	[soʊ**l**-**əɹ**]	‘solar’
(14b)	Same syllable	/kɑndə**l**+ə**l**/	[kɑndə**l**-**əɹ**]	‘condylar’
(14c)	Adjacent syl.s	/**l**i.g+ə**l**/	[**l**i.g-ə**l**]	‘legal’
(14d)	Adjacent syl.s	/p**l**ej.n+ə**l**/	[p**l**ej.n-**əɹ**]	‘planar’
(14e)	Adjacent syl.s	/**l**u.n+ə**l**/	[**l**u.n-**əɹ**]	‘lunar’
(14f)	1 transparent syl.	/**l**ej.bi.+ə**l**/	[**l**ej.bi.-ə**l**]	‘labial’
(14g)	1 transparent syl.	/p**l**u.vi.+ə**l**/	[p**l**u.vi.-ə**l**]	‘pluvial’
(14h)	1 transparent syl.	/**l**ə.kju.n+ə**l**/	[**l**ə,kju.n-**əɹ**]	‘lacunar’

One can observe that the number of transparent syllables is not exactly what divides the data in (14a) through (14h) into distinct classes; for example, zero transparent syllables come in between the trigger and target in both (14a) and (14c), and yet the two forms are organized into different classes. We will nonetheless put this consideration aside, returning to characterizing the unit of distance for liquid dissimilation in English later on. The figures surrounding distance-based decay of the process are shown below:

Liquid dissimilation in English: /**l**σ^n+ə**l**/ → [**l**σ^n-**əɹ**]			
Trigger-target distance	Faithful forms	Dissimilated forms	Proportion of dissim’d forms
Same syllable[5]	1	303	0.99
Adjacent syl.s	60	39	0.65
1 transp. syl.	85	10	0.10
2 transp. syl.s	24	1	0.04
3 transp. syl.s	4	0	0.00

Table 4: *figures for distance-based decay in liquid dissimilation in English*

5 The single faithful form that OED lists in which the trigger and the target are in the same syllable is [pri.**l**-ə**l**], ‘prelal’.

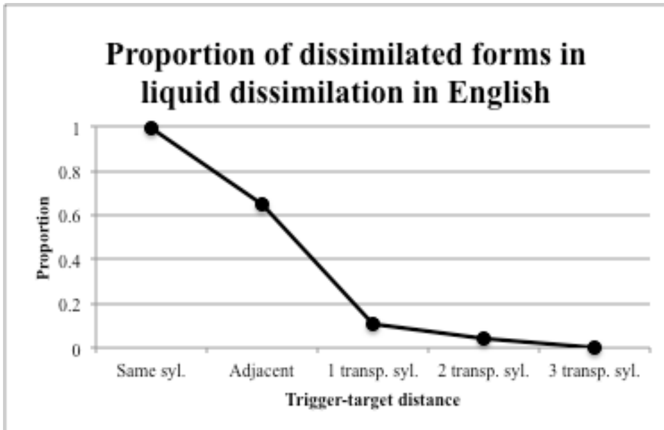


Figure 3: *graph of distance-based decay in liquid dissimilation in English*

The table in (14) and the graph in Figure 3 show that the number of transparent syllables has an erosive effect on the likelihood of application for liquid dissimilation in English.

Hungarian has a system of vowel harmony that displays distance-based decay. The vowel inventory is shown below:

	[-back, -round]	[-back, +round]	[+back, +round]
[+high]	/i/ and /iː/	/y/ and /yː/	/u/ and /uː/
[-high, -low]	/eː/	/ø/ and /øː/	/o/ and /oː/
[+low]	/ɛ/		/ɔ/ and /aː/

Table 5: *Hungarian vowel inventory*

Distance-based backness harmony can be observed on the vowels, as shown by the following items:

	UR	Dative form	Gloss
Faithful items			
(15a)	/kɛrt+nɛk/	[kɛrt-nɛk]	'garden'
(15b)	/tsiːm+nɛk/	[tsiːm-nɛk]	'address'

Items that undergo local and nonlocal vowel harmony

(16a)	/ɔbl**ɔ**k+n**ɛ**k/	[ɔbl**ɔ**k-n**ɔ**k]	'window'
(16b)	/kommunizm**u**s+n**ɛ**k/	[kommunizm**u**s-n**ɔ**k]	'Communism'
(16c)	/ɔpoʃt**o**li+n**ɛ**k/	[ɔpost**o**li-n**ɔ**k]	'apostolic'
(16d)	/b**o**riːteːk+n**ɛ**k/	[b**o**riːteːk-n**ɔ**k]	'envelope'

Items with opaque front rounded vowels

(17a)	/ʃ**o**f**øː**r-n**ɛ**k/	[ʃ**o**f**øː**r-n**ɛ**k]	'chauffeur'
(17b)	/**ɔ**lbeːrl**øː**-n**ɛ**k/	[**ɔ**lbeːrl**øː**-n**ɛ**k]	'lodger'

Hungarian has a dative suffix, -/nɛk/, whose underlying form is apparent in (15a) and (15b).[6] If the dative suffix -/nɛk/ attaches to a stem whose final vowel is [+back], then it undergoes vowel harmony, surfacing instead as -[nɔk], as in (16a) and (16b). (16c) and (16d) show that vowel harmony can apply at a distance, so long as the triggers and targets are separated only by consonants or front unrounded vowels. On the other hand, (17a) and (17b) show that intervening front round vowels block vowel harmony.

Hayes and Londe 2006 show that vowel harmony in Hungarian is subject to a height effect: the likelihood of vowel harmony is directly related to the height of the last transparent vowel. Hungarian has four front unrounded vowels that contrast for height: /i/, /iː/, /eː/, and /ɛ/. Consider the following data fixed to a distance of one transparent vowel:

	UR	Dative form	Gloss
Items with /i/ *and* /iː/			
(18a)	/ɔpoʃt**o**li+n**ɛ**k/	[ɔpost**o**li-n**ɔ**k]	'apostolic'
	/b**u**li+n**ɛ**k /	[b**u**li-n**ɔ**k]	'party'
	/gr**ɔ**fit+n**ɛ**k /	[gr**ɔ**fit-n**ɔ**k]	'graphite'
(18b)	/ɛksp**ɔ**nsiːv+n**ɛ**k /	[ɛksp**ɔ**nsiːv-n**ɔ**k]	'expansionary'
	/f**ɔ**kiːr+n**ɛ**k /	[f**ɔ**kiːr-n**ɔ**k]	'poor'
	/m**a**ːrtiːr+n**ɛ**k /	[m**a**ːrtiːr-n**ɔ**k]	'martyr'

[6] Further evidence that the dative suffix takes the underlying form -/nɛk/ comes from constructions such as [nɛk-ɛm] 'me'-DAT (Vago 1976). A few investigators nevertheless contend that the dative suffix vowel is underspecified for [back] (see Hayes and Londe 2006 for discussion); for the purposes analyzing the decay effect arising in vowel harmony, I simply assume -/nɛk/.

Items with /eː/

(18c)	/f**ɔ**seːn+n**ɛ**k /	[f**ɔ**seːn-n**ɛ**k]	‘charcoal’
	/ɔd**ɔ**leːk+n**ɛ**k/	[ɔd**ɔ**leːk-n**ɔ**k]	‘datum’
	/g**ɔ**lleːr+n**ɛ**k/	[g**ɔ**lleːr-n**ɔ**k]	‘collar’

Items with /ɛ/

(18d)	/komp**o**nɛns+n**ɛ**k/	[komp**o**nɛns-n**ɛ**k]	‘component’
	/hɔmb**u**rgɛr+n**ɛ**k/	[hɔmb**u**rgɛr-n**ɛ**k]	‘hamburger’
	/kr**ɔ**pɛk+n**ɛ**k/	[kr**ɔ**pɛk-**nɔk**][7]	‘dude’

Forms with the transparent vowel is /i/ or /iː/ are likelier to undergo vowel harmony than those with /eː/, and those with /eː/ are likelier to undergo vowel harmony than those with /ɛ/ as the transparent vowel.

Vowel harmony in Hungarian, much like the other processes we have seen, displays distance-based decay:

	Transparent Syllables	UR	Dative form	Gloss
(19a)	0 syllables	/ɔbl**ɔ**k+n**ɛ**k/	[ɔbl**ɔ**k-n**ɔ**k]	‘window’
(19b)	0 syllables	/bir**o**ː+n**ɛ**k/	[bir**o**ː-n**ɔ**k]	‘judge’
(19c)	0 syllables	/kommunizm**u**s+n**ɛ**k/	[kommunizm**u**s-n**ɔ**k]	‘Communism’
(19d)	1 syllable	/f**ɔ**seːn+n**ɛ**k/	[f**ɔ**seːn-n**ɛ**k]	‘charcoal’
(19e)	1 syllable	/ɔpoʃt**o**li+n**ɛ**k/	[ɔpost**o**li-n**ɔ**k]	‘apostolic’
(19f)	1 syllable	/m**a**ːrtiːr+n**ɛ**k/	[m**a**ːrtiːr-n**ɔ**k]	‘martyr’
(19g)	2 syllables	/d**o**ktrineːr+n**ɛ**k/	[d**o**ktrineːr-n**ɛ**k]	‘doctrinaire’
(19h)	2 syllables	/k**ɔ**libeːr+n**ɛ**k/	[k**ɔ**libeːr-n**ɛ**k]	‘caliber’
(19i)	2 syllables	/b**o**riːteːk+n**ɛ**k/	[b**o**riːteːk-n**ɔ**k]	‘envelope’

Hayes et al. 2009 make available a corpus—the results of a Google study—of stems that are token-weighted for vowel harmony. Associated with each stem in the corpus is the percentage of tokens of its dative form that undergo vowel harmony. For example, the stem /krɔpɛk/ ‘dude’ takes -[nɔk] 80% of the time and -[nɛk] 20% of the time.

[7] In fact, /kr**ɔ**pɛk+n**ɛ**k/ surfaces as [kr**ɔ**pɛk-n**ɛ**k] in one out of five instances.

The table shown below tabulates a token-weighted type frequency count of the corpus data. Forms such as [ɔblɔk-nɔk] ‘window’-DAT that do not exhibit within-word variation contribute 1 to the relevant type count. [krɔpɛk-nɔk] and [krɔpɛk-nɛk], on the other hand, will contribute respectively .8 and .2 to the pertinent type counts. The leftmost column divides words based on distance and the height of the last transparent vowel. B represents the triggering back vowel and T represents a transparent vowel.

Vowel harmony in Hungarian: /$\begin{bmatrix} \text{+syl} \\ \text{+back} \end{bmatrix}$ σ^n+nɛk/ → [$\begin{bmatrix} \text{+syl} \\ \text{+back} \end{bmatrix}$ σ^n-nɔk]			
Type of stem	Faithful forms	Harmonized forms	Proportion of harmonized forms
B	4.32	6284.68	0.99
Bi	4.40	467.60	0.99
Bi:	1.05	51.95	0.98
Be:	18.65	101.35	0.84
Bɛ	103.94	13.06	0.11
BTi	27.72	9.28	0.25
BTi:	5.14	2.86	0.36
BTe:	6.94	5.06	0.42
BTɛ	20.99	0.00	0.00
BTTi	2.00	0.00	0.00
BTTi:	0.00	0.00	N/A
BTTe:	4.00	0.00	0.00
BTTɛ	2.00	0.00	0.00

Table 6: *figures for distance-based decay in vowel harmony in Hungarian*

The four line graphs shown below displays distance-based decay in the number of transparent syllables and were generated based on the height of the last vowel:

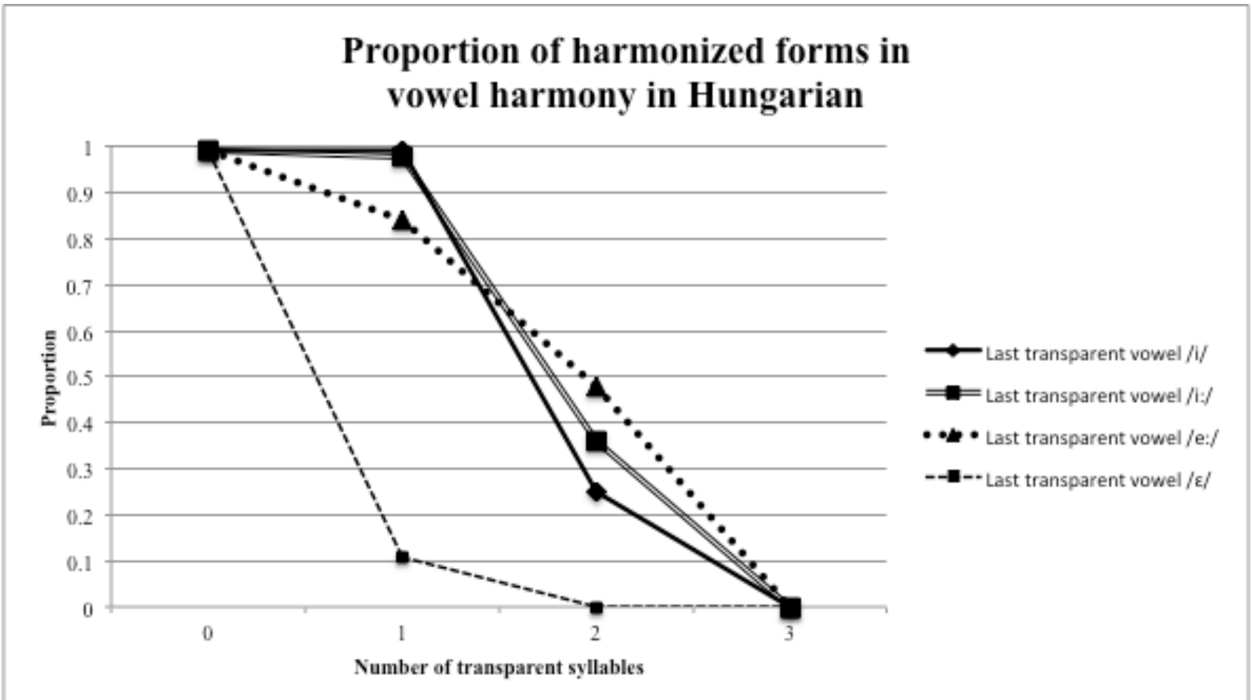


Figure 4: *graph of distance-based decay in vowel harmony in Hungarian*

As the table and graph show, the number of transparent syllables has an erosive effect on the likelihood of application.

The curves in Figure 3 and Figure 4 are strikingly similar to sigmoid curves (see Zuraw 2010 for other instances of sigmoid curves in phonology). As we will see later, the curves of distance-based application rates of the surveyed processes can be modeled by sigmoid curves, even though Figure 1 and Figure 2 do not look as sigmoid-like as Figure 3 and Figure 4. Ultimately, while decay rates across the surveyed languages differ to some extent, all of the cases show that the amount of interaction between the trigger and target decreases as the amount of transparent distance increases. We now turn to the task of accounting for the decay effect in the theory of phonology.

3 Rationale behind the account of distance-based decay

The account of distance-based decay will be grounded in the framework of Maximum Entropy Harmonic Grammar (Smolensky 1986, Goldwater and Johnson 2003, Hayes and Wilson 2008). Maximum Entropy Harmonic Grammar is a variant of Optimality Theory (Prince and Smolensky 1993), which specifies that the grammar of any given language is a set of conflicting constraints ranked in a strict hierarchy to produce a set of categorical outputs (i.e., surface forms) from a set of inputs (i.e., underlying forms). Maximum Entropy Harmonic Grammar departs from Optimality Theory in that the former is probabilistic: each output is associated with a nonzero probability that is a function of numerical constraint weights—real values associated with each constraint that signify the strength of the constraint.

Suppose we have *n* constraints and are considering *m* surface forms of single underlying form. If w_k is the weight of constraint *k* and C_{ik} is the number of times the *i*th surface form violates constraint *k*, then the harmony H_i of the *i*th surface form is defined as:

(20a) $H_i = \sum_{k=1}^{n} w_k * C_{ik}$

The probability P_i of the *i*th surface form is taken to be the inverse exponent of H_i normalized by the sum of the inverse exponents of the harmonies of each of the surface forms:

(20b) $P_i = e^{-H_i} / \sum_{j=1}^{m} e^{-H_j}$

To see how Maximum Entropy Harmonic Grammar works specifically for the purposes of accounting for long-distance phonological processes, we can consider the general case of long-distance dissimilation. Suppose we have an underlying form /*s*...*s*/ containing two instances of a segment *s*, as well as the surface forms [*s*...*s*] and [*s*...*s*′], where *s*′ is a segment distinct from *s*. Let **s*...*s* be a markedness constraint that is violated by [*s*...*s*] once, and let its weight be w_m (i.e., the weight of markedness). Let IDENT be a faithfulness constraint that is violated by [*s*...*s*′]

once, and let its weight be w_f (i.e., the weight of faithfulness). Dissimilation in Maximum Entropy Harmonic Grammar works as follows:

/s…s/	*s…s $w = w_m$	IDENT $w = w_f$	Harmony	Predicted Probability
[s…s]	1		$1 * w_m = w_m$	$e^{-w_m}/(e^{-w_m} + e^{-w_f})$
[s…s′]		1	$1 * w_f = w_f$	$e^{-w_f}/(e^{-w_m} + e^{-w_f})$

Table 7: *tableau representation of dissimilation in Maximum Entropy Harmonic Grammar*

Since each candidate violates exactly one constraint only once, the harmony for each candidate is simply the weight of the constraint it violates: [*s*…*s*] violates **s*…*s* and thus incurs a harmony of w_m, while [*s*…*s*′] violates IDENT and incurs a harmony of w_f. The harmony of [*s*…*s*′] is *wf* and the probability of it surfacing—i.e., the probability of dissimilation—is as follows:

(21) $$P_{[s\ldots s']} = \frac{e^{-w_f}}{e^{-w_m}+e^{-w_f}} = \frac{1}{1+e^{-(w_m-w_f)}}$$

The probabilities of [*s*…*s*] and [*s*…*s*′] sum to 1, and so the probability of the faithful candidate surfacing is $P_{[s\ldots s]} = 1 - P_{[s\ldots s']}$. The expression on the rightmost side of the equation in (21) will be useful later for when we model distance-based decay.

Hayes and Wilson 2008 provide the Maxent Grammar Tool (http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/), a computer program that takes as input a set of underlying forms, a set of surface forms for each underlying form, the frequencies of the surface forms, a set of constraints, and the violation vectors of each of the surface forms given the underlying forms and constraints. It uses the conjugate gradient method to find the set of constraint weights that maximize the probability of the dataset[8]—that is, the

[8] In fact, the Maxent Grammar Tool finds the constraint weights that maximize the probability of the dataset minus a regularization penalty. The penalty is set to trivial values for this study.

model-predicted probability that the surface forms take on the frequencies they do given the underlying forms and constraints.

One approach to accounting for distance-based decay arising in long-distance phonological processes is to posit a family of markedness constraints that penalize trigger-target pairs at different distances (Hansson 2001). Take distance-based decay as it arises in liquid dissimilation in Latin, for example. Leaving aside the fact that the process only applies to the -/aːlis/ suffix, we can posit as a preliminary study a constraint-family account for the decay effect shown in (22).

(22)

$*l\sigma^0l$:	penalize an output if it contains a pair of [l]s that are in adjacent syllables.
$*l\sigma^1l$:	penalize an output if it contains a pair of [l]s that are in syllables that are separated by one transparent syllable.
$*l\sigma^2l$:	penalize an output if it contains a pair of [l]s that are in syllables that are separated by two transparent syllables.
IDENT([lat]):	corresponding segments in the input and output match for the value of [lateral].

The Maxent Grammar Tool, upon taking as input the above constraints as well as the data in Table 3 on liquid dissimilation in Latin, returns the following constraint weight values:

(23)

$w(*l\sigma^0l) =$ 21.02
$w(*l\sigma^1l) =$ 9.27
$w(*l\sigma^2l) =$ 7.57
IDENT([lat]) = 8.57

We find that the outputted distance-based markedness weights are decreasing with greater numbers of transparent syllables. This can account for the decay effect in liquid dissimilation: the decrease in weights is the learner's response to a decrease in rate of application.

It is in this response that lies the problem with positing a constraint family account for distance-based decay. Distance-based decay is a crosslinguistically and crossprocessually general phenomenon, its erosive nature represented across the languages and long-distance processes we have seen. Under the constraint family account, the learner would thus coincidentally learn decreasing weights with increasing transparent distance across all languages that show the effect. On the other hand, such approach dismisses this systematicity as coincidental. It does not rule out the existence of a learner who acquires distance-based "anti-decay", for instance, in which the weights of the constraints increase with transparent distance, or a learner who acquires a language with greatest application at intermediate distances. Yet languages that possess such properties are unattested. We therefore reject this approach due to its being too powerful.

In my account of distance-based decay, I adopt and extend the proposal put forth by Kimper 2011, which scales the weight of markedness to produce the decay effect. Unlike the constraint-family account, a scale-based account relieves the learner of the task of acquiring the decay effect. Scaling can be achieved by positing a decay function *d* that takes a measure of distance between the trigger and target of a long-distance phonological process and returns a nonnegative real number. When the harmony of the form is calculated, the scalar value is then multiplied by the weight of the markedness constraint that regulates the long-distance cooccurrence restriction.

More precisely, let C_m be a markedness constraint with weight w_m regulating a long-distance restriction against the cooccurrence of the two (not necessarily distinct) segments *a* and b, and let some nonnegative integer $x \in \mathbb{Z}^*$ be the measure of transparent distance between *a* and b in a surface form with a subsequence $[aSb]$, where $[S]$ is itself a subsequence of transparent segments. Then define the following:

(24) $d: \mathbb{Z}^* \to (0, \infty)$: takes a measure of transparent distance *x* between the trigger and target defined by C_m and returns a real value $d(x) \in (0, \infty)$.

A surface form that contains the subsequence $[aSb]$ violates **a...b* $d(x)$ times, while a surface form that contains both the subsequences $[aS_1b]$ and $[aS_2b]$ violates **a...b* $\left(d(x_1) + d(x_2)\right)$ times, and so on.[9] I will restrict attention to the case where **a...b* is violated once, as this is the only case on which I have data. We see from the tableau in (22) that the harmony of the faithful candidate is equal to the weight of markedness. Assuming that the only two constraints in the language are C_m and IDENT, the formula for the harmony of surface form *i* containing $[aSb]$—which violates C_m—is as follows:

(25) $H_i = w_m * d(x)$

The formula for the harmony for unfaithful candidate satisfying C_m but violating IDENT remains the same: the weight of faithfulness.

To see an example of how the system works, we consider the set of candidate forms for liquid dissimilation in Latin in which the trigger and target are separated by two transparent syllables. Suppose we define constraints *l…l and IDENT([lat]) so that *w*(*l…l) = 10.97 and *w*(IDENT([lat])) = 4.53, and instantiate *d*(*x*) to be the negative power function $d(x) = 1/(x+1)$, where *x* is the number of transparent syllables. Justification for using a decay function like this one will be covered in the following section.

---

[9] Just to cover a more complicated case, suppose that a language penalized $[aSb]$ sequences even when another instance of *a* or *b* occurs in *S*; for example, $[S] = [S_1bS_2]$. Then the subsequence $[aSb] = [aS_1bS_2b]$ incurs a violation of $\left(d(x_1) + d(x_2')\right)$, where x_1 is the transparent distance between *a* and the first instance of *b*, and x_2' is the transparent distance between *a* and the second instance of *b*. This case does not arise in the data at hand.

/…**l**σ²+a**l**is/ (e.g., /**l**apid+a:**l**is/)	*l…l w = 10.97	IDENT([lat]) w = 4.53	Harmony	Predicted probability	Observed probability
[…**l**σ²-a:**l**is] (e.g., [**l**apid-a:**l**is])	1/3 ≈ 0.33 violations		$10.97 * 0.33 = w_m * d(x)$	$\frac{e^{-10.97*0.33}}{(e^{-10.97*0.33} + e^{-4.53*1})} \approx 0.70$	0.69
[…**l**σ²-a:**r**is] (e.g., [**l**apid-a:**r**is])		1	$4.53 * 1 = w_f$	$\frac{e^{-4.53*1}}{(e^{-10.97*0.33} + e^{-4.53*1})} \approx 0.30$	0.31

Table 8: *tableau representation of liquid dissimilation in Latin with a decay function*

The tableau—and, in particular, the score of the faithful form—demonstrates that augmenting the grammar through scaling the weight of markedness can produce predicted probabilities that accurately match the observed ones. The scaling is achieved by multiplying the number of markedness violations by the output of a decay function that takes as input the number of transparent syllables.

The tableau above shows that our model works well in predicting distance-based application rates given particular parameter values (i.e., particular values of the weights of markedness and faithfulness) and a particular decay function, a simple power function. In the following section, I explain how I determined the parameter values seen in (28) and discuss why a power function is desirable for modeling distance-based decay.

4 Modeling distance-based decay

The task is to find an accurate yet crosslinguistically and crossprocessually robust decay function for modeling distance-based decay. What should $d(x)$ look like? With few exceptions, the processes examined apply categorically in strictly local environments, and as the amount of transparent distance increases, the rate of application tends to zero. Therefore, the effect of multiplying the weight of markedness by the decay function should be such that the scaled weight of markedness should be high for forms that violate markedness locally and lower for forms that violate markedness at a distance. Moreover, the scaled weight of markedness tends to zero with increasing distance.

A natural choice for $d(x)$ that has these properties is the negative power function, $d(x) = 1/x^k$, where k is a positive real number. For now, we will assume that k is a language-specific parameter; nonetheless, Section 4.2.2 will argue that a single value of k can be used to accurately model the decay effect in the four surveyed languages. For all integer measures of units of transparent material $x > 0$ we have that that $0 < d(x) < d(x-1)$, thus producing the erosive effect that greater transparent distance will have. Between the values of $1 \leq x \leq 4$, the decay function would therefore look roughly as follows:

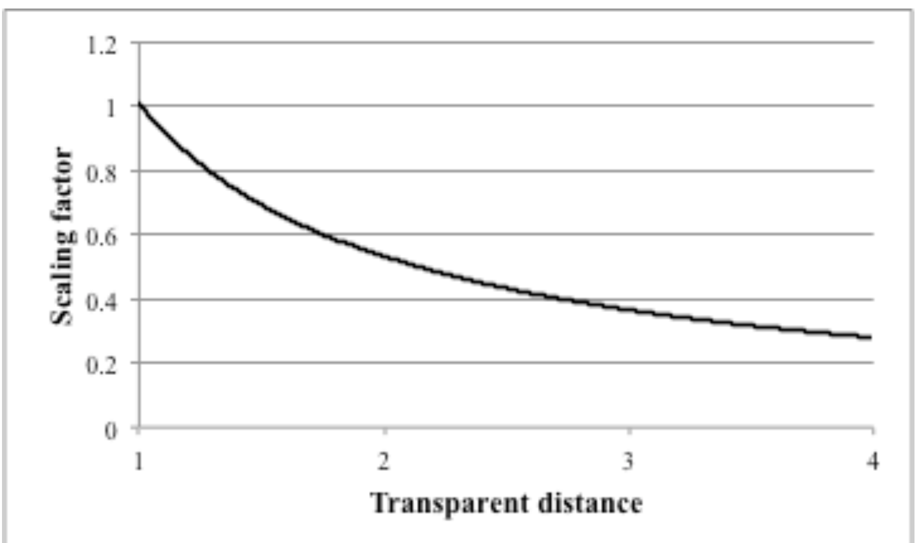


Figure 5: *graph of the negative power function* $d(x) = 1/x$

One complication of using a negative power function is that $d(0)$ is undefined, and $d(x)$ tends to positive infinity as we approach zero from the right-hand side of the y-axis within the first quadrant. I avoid this problem by only multiplying the weight of markedness by $d(0.1)$ in cases where I should be multiplying it by $d(0)$. As we will see, we will only need to consider $d(0.1)$ for one case: the case of local liquid dissimilation in English (e.g., forms like [soʊ.ləɹ]). Otherwise, *x* takes on nonzero values.

There are alternative candidate decay functions that could have been used; for example, we could use a linear function to model the effect instead of a negative power function. We will see later that a linear model does not fit to the data as well as the negative power function. For now, however, we put this consideration aside, focusing instead on elucidating properties about the input to the decay function as well as its parameters.

4.1 Determining an appropriate measure of distance

Previously, we discussed distance-based decay arising when the number of transparent syllables increases. What makes transparent syllables special? Why not use a finer unit of distance such as segments? This section determines a single unit of distance that is significant in producing the decay effect in the above languages. This will be the unit of measure *x* of the distance function, *d*(*x*).

I used R's `lme4` package and the `Anova()` function to run likelihood ratio tests comparing generalized linear models of the data for the four covered languages. For a given linear model, the likelihood ratio test compared the full model to models that omit each variable in turn, and returned *p*-values that determined whether the full model was better than each of the smaller models. The likelihood ratio test thereby determined whether each of the variables was significant in influencing application rate. I then calculated the A(kaike) I(nformation) C(riterion) of the model, which calculated overall fit of the model to the data while penalizing models with more parameters. In cases where one or more variables were not significant, I eliminated the variable with the highest *p*-value, reran the likelihood ratio test, and recalculated the AIC, which (typically[10]) yielded a lower value than that of the first model. I iterated this process until the remaining variables all had *p*-values below *p* = 0.1, and the AIC was lower than those of models containing variables that were not significant. For example, the fullest model of labial dissimilation in Malagasy—before any variables were eliminated—is shown below:

```
Call:
glm(formula = Dissimilation ~ Syllables + Segments + Labial +
Coronal + uCount + Interveningei, family = "binomial", data = MyData)

Coefficients:
              Estimate Std. Error z value
(Intercept)     4.5751     0.4037  11.334
Syllables      -3.7277     0.3653 -10.204
Segments       -0.1677     0.1257  -1.334
Labial          0.1107     0.2240   0.494
Velar          -0.3757     0.3402  -1.104
uCount         -0.2527     0.1676  -1.508
Interveningei  -4.9616     0.7178  -6.912
---
AIC: 772.05
```

The model resulting from applying the algorithm mentioned above is shown below:

```
Call:
glm(formula = Dissimilation ~ Syllables + Interveningei, family = "binomial",
data = MyData)

Coefficients:
              Estimate Std. Error z value
(Intercept)     4.0396     0.2379  16.983
Syllables      -3.9778     0.2548 -15.614
Interveningei  -4.9508     0.7167  -6.908
---
AIC: 769.4
```

The primary finding is that the number of transparent syllables, rather than the number of transparent segments, is a significant predictor of the likelihood of process application. In particular, when one puts both counts into the same model, the model that is yielded by applying the algorithm mentioned at the beginning of this section is one in which syllable count is

[10] Sometimes an "intermediate stage" AIC had a slightly higher value than the AIC of the model before it (e.g., in the following derivation, the model with syllable count, segment count, and intervening front vowel count had a slightly higher AIC than the model with syllable count, segment count, intervening front vowel count, and trigger count). In the following derivation, for example, a slight rise in the AIC of an intermediate model is not a cause for concern because the final model has a lower AIC its predecessors.

preserved and segment count is eliminated. Furthermore, while syllable count and segment count are highly correlated in all cases ($r = 0.91$ for Malagasy, $r = 0.94$ for Latin, $r = 0.95$ for English, and $r = 0.80$ for Hungarian), the best model that is yielded from starting with a full model containing only syllable count has a lower AIC than the best model that is yielded from starting with only segment count (compare 769.40 and 891.81 in Malagasy, 143.44 and 143.53 in Latin[11], 278.70 and 305.57 in English, and 311.38 and 354.05 in Hungarian). The results of using this method are shown in the table below:

Phonological process	*Factors that are significant or nearly significant and their coefficients*	*Factors that are not significant*
Vowel dissimilation in Malagasy	**- Number of transparent syllables is significant in producing the decay effect ($p < 0.001$)** - Intervening front round vowels block application ($p < 0.001$)	**- Number of transparent segments** - Number of triggers - Presence of intervening labial or coronal consonants
Liquid dissimilation in Latin	**- Number of transparent syllables is significant in producing the decay effect ($p = 0.03$) (see footnote 8)** - If the trigger was in onset-noninitial position, application was significantly less likely ($p < 0.001$) - Whether the trigger was in coda position is nearly significant ($p = 0.06$) in tending to provoke application	**- Number of transparent segments (see footnote 8)** - Presence of intervening labial or velar consonants - Number of triggers

[11] Latin looks to be a borderline case: the best model according to AIC (142.47) was one with both segments (p = 0.03) and syllables (p = 0.07) in it. Crucially, however, this model had intervening labials in it, a variable which was not significant according to the ANOVA; as a result, I eliminated intervening labials, and in the subsequent model, only syllables were significant, but not segments. Segments were then eliminated, yielding the model that was kept for the analysis at hand. The model that was kept had an AIC of 143.44. These results make me suspicious of the segment-syllable comparison in Latin, and may suggest that we should explore using some counting method other than syllables/intervening target positions (see Section 4.2.1 and Section 5.1) in the future.

Vowel harmony in Hungarian	**- Number of transparent syllables is significant in producing the decay effect ($p < 0.001$)** - Intervening front round vowels block application ($p < 0.001$) - Height of the last transparent vowel is significantly related to application rate ($p < 0.001$) - Number of triggers is nearly significant in provoking application ($p = 0.064$)	**- Number of transparent segments**
Liquid dissimilation in English	**- Number of transparent syllables is significant in producing the decay effect ($p < 0.001$)** - Intervening /ɹ/ blocks application ($p < 0.001$) - If the trigger was in coda position, application was significantly less likely ($p < 0.001$) - If the trigger was in onset-noninitial position, application was significantly less likely ($p = 0.04$) - Intervening velar consonants were significant in blocking application ($p = 0.02$)	**- Number of transparent segments** - Number of triggers - Intervening labial consonants

Table 9: *factors that are significant/not significant in influencing application rate*

None of the full models included interactions between variables. Cser 2010 argues that application rate of liquid dissimilation in Latin is negatively influenced by the presence of intervening labial and velar consonants; nonetheless, my findings show that they were not significant in producing the decay effect. Furthermore, even though the number of triggers was significant in predicting the application rate of geminate devoicing in Japanese (Kawahara and Sano 2013), the number of triggers was not significant in predicting application rate for each of the processes I surveyed (though it was nearly significant in predicting vowel harmony in Hungarian). In vowel dissimilation in Malagasy, the number of transparent syllables is significant ($p < 0.001$) in producing the decay effect, while the number of transparent segments is not. Likewise, the number of transparent syllables is significant in liquid dissimilation in Latin ($p = 0.03$) and English ($p < 0.001$), and vowel harmony in Hungarian ($p < 0.001$). The trigger's position in the syllable turned out to be significant to rate of application in liquid dissimilation in Latin and English. In either case, the trigger being in onset-noninitial position was associated with significantly lower application rates. In addition, whether or not trigger position matches target position matters to some extent: in Latin, triggers that were in the syllable coda do not match the target for position, and as a result are potentially associated with higher application rate ($p = 0.08$); in English, triggers that were in the syllable coda match the target for position, and are associated with lower application rate ($p < 0.001$). It may be the case that candidates for long-distance consonant dissimilation display greater-than-chance faithfulness if the trigger and target match for syllable position. Nevertheless, since coda position is only nearly significant in Latin, and since we only have two cases (i.e., English and Latin), such result is only tentative.

Liquid dissimilation in English is an interesting case when it comes to counting transparent syllables. Consider syllable count as a measure of distance for a process in which the target is in coda position:

(26a) Categorical liquid dissimilation:
[soʊ.ləɹ]
[rɛ.gjə.ləɹ]
[æl.vi.jə.ləɹ]
[vɛn.tɹɪk.jə.ləɹ]

(26b) Variable liquid dissimilation:
[li.gəl]
[loʊ.kəl]
[lu.nəɹ]
[lʌm.bəɹ]

In (26a), liquid dissimilation applies to the words categorically: the trigger and target are in the same syllable, and thus there are zero transparent syllables separating the two. In (26b), application is variable (applying to roughly only four out of ten words, shown in Table 4 and Figure 3): the trigger and target are in adjacent syllables and the trigger is in onset position. Strikingly, however, zero transparent syllables separate the trigger from the target. Transparent syllable count is inadequate as a predictor of the decay effect because the same count of distance—zero transparent syllables—classifies two groups of data that behave differently from one another.

More broadly, the comparison between (26a) and (26b) demonstrates that transparent syllable count is an inadequate notion of distance when we are accounting for processes in which the trigger can be in onset position while the target is in coda position. We could simply posit that the decay function takes different units of measurement depending on whether or not a process enables the configurations seen in the above examples, but then we would have to explain why learners of different long-distance processes could have different ways of distinguishing classes of surface forms based on transparent distance.

Consider these four data points across the four processes we have seen:

Malagasy	English
Vowel dissimilation with trigger and target in the nuclei of adjacent syllables: CATEGORICAL APPLICATION	Liquid dissimilation with trigger in onset and target in coda of the *same syllable*: CATEGORICAL APPLICATION
(27a) /ba.bu.+u/ → [ba.bu.-**i**]	(27b) /soʊ.l+əl/ → [soʊ.l-**əɹ**]
Latin	English
Liquid dissimilation with trigger and target in nuclei of adjacent syllables: CATEGORICAL APPLICATION	Liquid dissimilation with trigger in onset and target in coda of *adjacent syllables*: OPTIONAL APPLICATION
(27c) /so.l+aː.lis/ → [so.l-**a**ː.**ris**]	(27d) /loʊ.k+əl/ → [loʊ.k-əl] (27e) /lu.n+əl/ → [lu.n-**əɹ**]

As seen in (27a) and (27b), the number of transparent segments is different between data that reflect local vowel dissimilation and local liquid dissimilation even though both process are categorical. This is undesirable, since we do not want multiple distinct counts of distance to predict categoricality across processes. The number of intervening moras (Kimper 2011), rimes, or syllable boundaries is inadequate for the same reason: such units yield differing counts between data that reflect local vowel dissimilation (as in (27a)) and local liquid dissimilation (as in (27c)). Reiterating, the number of intervening syllables is the same between the subsets of the data that reflect local liquid dissimilation and nonlocal liquid dissimilation in English, as the contrast between (27a) and the pair (27b) and (27c) reveal. This is undesirable, since we want the distance count to distinguish classes of data that reflect categorical application from optional application.

In light of these facts, two options stand out. The first option is to alter slightly the way we count syllables. Instead of measuring distance simply by the number of transparent syllables, we can define distance as follows:

(28) The distance between the trigger and target is 0 if the they are in the same syllable, 1 if they are in adjacent syllables, 2 in cases with one transparent syllable, 3 in cases with two transparent syllables, and so on.

Notice that this is different than simply counting transparent syllables, because it distinguishes between cases in which the trigger and target are in the same syllable from those in which the trigger and target are in adjacent syllables. Hereafter, we will call the metric in (28) *syllabic distance*. Defining distance in this way means that forms that violate liquid dissimilation in English locally (e.g., /soʊ.l-əl/ > [soʊ.l-əɹ]) incur a violation of *d*(0), while the forms that violate the long-distance processes most locally in the other languages (e.g., /ba.bu.-u/ > [ba.bu.-i] in Malagasy) get a count of *d*(1). As we will see later, the effect that defining distance in this way will have is that the weight of markedness will be lower than that of faithfulness for liquid dissimilation in English alone, but not for the rest of the cases.

A second option would be to take seriously the fact that the units that arise commonly in phonology fail to provide a consistent count of transparent distance in the data in (32a-e), and our only recourse is to posit a new way of counting transparent distance. I propose a new such unit: intervening target positions.

(29) Let the target of the process be in syllable position *P*, where *P* is one of onset, nucleus, or coda. Then define the distance between the trigger and target to be the number of positions *P* that come in between trigger and target, plus one.[12]

Hereafter, we will call this metric *position distance*. The data in (26a) and (26b), reproduced below as (30a) and (30b), on liquid dissimilation in English are thus accounted for as follows:

(30a) Categorical liquid dissimilation:
[(soʊ).(ləɹ)]
[(rɛ).(gjə).(ləɹ)]
[(æl).(vi).(jə).(ləɹ)]
[(vɛn).(tɹɪk).(jə).(ləɹ)]

Position distance: 1

(30b) Variable liquid dissimilation:
[(li_).(gəl)]
[(loʊ_).(kəl)]
[(lu_).(nəɹ)]
[(lʌm).(bəɹ)]

Position distance: 2

As the target is in the coda position, we are counting the number of intervening coda positions. Note that a coda position need not be filled in order for it to be counted, as we see in (30b). The number of intervening target positions in each of the data in (30a) is zero, so the distance is one. Furthermore, the number of intervening target positions in each of the data in (30b) is one, so the distance is two. For Malagasy, the distance count works as follows:

(31a)	(ba).(bu).(u)	Intervening target positions (nuclei): 0	Position distance = 1
(31b)	(ru).(va).(u)	Intervening target positions (nuclei): 1	Position distance = 2
(31c)	(ku).(ta).(ba).(u)	Intervening target positions (nuclei): 2	Position distance = 3

[12] We note the "plus one" here because the trigger and target are never in the same target position, and so we never have a case where the distance is 0 by this definition. The metric provided in (33) (i.e., syllabic distance) can take on a value of 0, since the trigger and target pairs can be in the same syllable.

Notice here that the target is in the nucleus position, and so we count intervening nucleus positions. As all syllables have nuclei, the number of intervening target positions is always equal to the number of transparent syllables. For Latin, the distance count works as follows:

(32a)	(so).(la:).(ris)	Intervening target positions (onsets): 0	Position distance = 1
(32b)	(la).(na:).(ris)	Intervening target positions (onsets): 1	Position distance = 2
(32c)	(lek).(t u).(a:).(lis)	Intervening target positions (onsets): 2	Position distance = 3

In the examples above, we also find the number of intervening targets is equal to the number of transparent syllables. Notice that the number of intervening target positions is equal to the number of intervening syllables in processes in which the target is in nucleus or onset position. In other words, syllabic distance and position distance are always the same when the target is in nucleus or onset position. This is demonstrated in the table below:

For a long-distance phonological process /AXB/ → [AXB′]			
Target position	Surface form structure	Number of intervening target positions	Number of transparent syllables
Onset	[(AVC_0).(C_0VC_0)^n.(B′VC_0)]	n	n
Onset	[(C_0VA).(C_0VC_0)^n.(B′VC_0)]	n	n
Nucleus	[(C_0AC_0).(C_0VC_0)^n.(C_0B′C_0)]	n	n
Coda	[AVB′]	0	0
Coda	[(AVC_0).(C_0VC_0)^n.(C_0VB′)]	$n + 1$	n
Coda	[(C_0VA).(C_0VC_0)^n.(C_0VB′)]	n	n

Table 10: *comparing syllabic distance versus position distance over different types of surface forms*

The only case where we must make a distinction between the number of intervening target positions and the number of transparent syllables is when we are accounting for forms in which the trigger is in the onset positions and the target in the coda position, as is the case in liquid dissimilation in English. Furthermore, one should note that, while the position distance metric in (29) and the syllabic distance metric in (28) are highly correlated as variables for liquid dissimilation in English (the correlation value was 0.96), the best model with only position distance had an AIC of 278.70 while the best model with syllabic distance had an AIC of 296.60, suggesting that position distance is the counting mechanism we should adopt.

Furthermore, as we will find out in the following section, it is beneficial to adopt intervening target positions as our unit count. In addition to finding optimal parameter values and comparing decay functions of different shapes, the following section will compare models that use syllabic distance against ones that use position distance. We will see that position distance yields models that better fit the data at hand.

4.2 Determining the parameter values of the decay function

Three parameters for the decay function need to be found: the weight of markedness w_m, the weight of faithfulness w_f, and the decay parameter *k*. Section 4.2.1 finds optimal language-specific values for the three parameters. Section 4.2.2 shows that distance-based decay can be modeled accurately across the four languages with a single value for the decay parameter *k*.

4.2.1 Decay functions with language-specific values for *k*

In the beginning of section 4, we came to reject an approach which accounts for distance-based decay with a distance-based constraint family, instead opting for an account that utilizes a single markedness constraint that is scaled by the decay function *d*(*x*). Nevertheless, as we will see, the weights of heuristic distance-based constraints suggests that *d*(*x*) is inverse-exponential shape.

Recall the distance-based constraints that we had posited to account for liquid dissimilation in Latin. Based on our new definition of distance, we restate them as follows:

(33)

MaxEnt input:		MaxEnt output:
*DISAGREE[lat]-DIST(1):	penalize adjacent [l]-segments.	w = 21.02
*DISAGREE[lat]-DIST(2):	penalize pairs of [l]-segments that are one transparent syllable away from each other.	w = 9.27
*DISAGREE[lat]-DIST(3):	penalize pairs of [l]-segments that are two transparent syllables away from each other.	w = 7.57
IDENT([lat]):	corresponding segments in the input and output match for the value of [lateral].	w = 8.57

The values of the weights for the distance-based weights decrease in a roughly inverse-exponential fashion with distance, starting high and tending to zero:

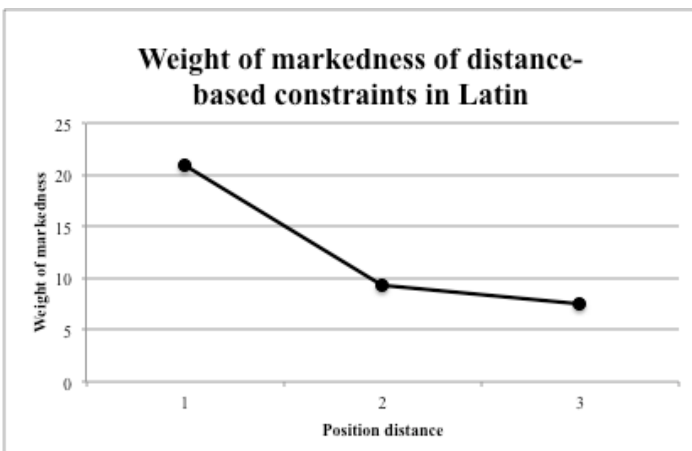


Figure 6: *inverse-exponential decrease in the weight of markedness in liquid dissimilation in Latin*

We see that the difference in weight between *DISAGREE[lat]-DIST(1) and *DISAGREE[lat]-DIST(2) is greater than that between *DISAGREE[lat]-DIST(2) and *DISAGREE[lat]-DIST(3). Notice that the graph above resembles that of a negative power function, and in particular, the graph of $d(x) = 21/x$, where x is position distance (which, again, is equal to the number of transparent syllables plus one in this case).

The weights outputted by corresponding heuristic grammars for the four languages covered in Section 2 are shown in the table below. Average error is defined here as the absolute difference between observed and predicted probabilities averaged over the violation profiles in each language.

Vowel dissimilation in Malagasy: /**u**σ^n+**u**/ → [**u**σ^n-**i**]	
w(*DISAGREE[rd]-DIST(1)):	14.08
w(*DISAGREE[rd]-DIST(2)):	8.59
w(*DISAGREE[rd]-DIST(3)):	6.62
IDENT([rd]):	8.57
Average error:	0.00

Liquid dissimilation in Latin: /**l**σ^n+aː**l**is/ → [**l**σ^n-aː**r**is]	
w(*DISAGREE[lat]-DIST(1)):	21.02
w(*DISAGREE[lat]-DIST(2)):	9.27
w(*DISAGREE[lat]-DIST(3)):	7.57
IDENT([lat]):	8.37
Average error:	0.00

Vowel harmony in Hungarian: /$\begin{bmatrix}\text{+syl}\\\text{+back}\end{bmatrix}$ σ^n+n**ɛ**k/ → [$\begin{bmatrix}\text{+syl}\\\text{+back}\end{bmatrix}$ σ^n-n**ɔ**k]	
w(*AGREE[bk]-DIST(1)):	21.11
w(*AGREE[bk]-DIST(2)):	11.75
w(*AGREE[bk]-DIST(3)):	7.21
w(*$\begin{bmatrix}\text{+syl}\\\text{+back}\end{bmatrix}$ ... i$\begin{bmatrix}\text{+syl}\\\text{-back}\end{bmatrix}$)	6.06
w(*$\begin{bmatrix}\text{+syl}\\\text{+back}\end{bmatrix}$ ... iː$\begin{bmatrix}\text{+syl}\\\text{-back}\end{bmatrix}$)	6.02
w(*$\begin{bmatrix}\text{+syl}\\\text{+back}\end{bmatrix}$ ... eː$\begin{bmatrix}\text{+syl}\\\text{-back}\end{bmatrix}$)	4.06
w(*$\begin{bmatrix}\text{+syl}\\\text{+back}\end{bmatrix}$ ... ɛ $\begin{bmatrix}\text{+syl}\\\text{-back}\end{bmatrix}$)	0
IDENT([bk]):	13.83
Average error:	0.07

Liquid dissimilation in English: /**l**σ^n+ə**l**/ → [**l**σ^n-ə**ɹ**]	
w(*DISAGREE[lat]-DIST(0)):	14.16
w(*DISAGREE[lat]-DIST(1)):	8.14
w(*DISAGREE[lat]-DIST(2)):	6.36
w(*DISAGREE[lat]-DIST(3)):	5.37
IDENT([lat]):	8.46
Average error:	0.00

Table 11: *distance-based decay modeling results using distance-based markedness constraints*

The language corpora for these languages display a paucity of forms on distances greater than two transparent syllables, and so I only provided weights for constraints up to three units of distance. Since English has forms with zero units of distance, the data suffice to motivate a fourth constraint.

The distance-based constraints for Hungarian are controlled for vowel height: they are the returned weights of a grammar that included the distance-based constraints shown in Table 11 as well as the height-based constraints $*\begin{bmatrix}\text{+syl}\\ \text{+back}\end{bmatrix} \ldots \text{i}\begin{bmatrix}\text{+syl}\\ \text{-back}\end{bmatrix}$, $*\begin{bmatrix}\text{+syl}\\ \text{+back}\end{bmatrix} \ldots \text{i:}\begin{bmatrix}\text{+syl}\\ \text{-back}\end{bmatrix}$, $*\begin{bmatrix}\text{+syl}\\ \text{+back}\end{bmatrix} \ldots \text{e:}\begin{bmatrix}\text{+syl}\\ \text{-back}\end{bmatrix}$, and $*\begin{bmatrix}\text{+syl}\\ \text{+back}\end{bmatrix} \ldots \varepsilon\begin{bmatrix}\text{+syl}\\ \text{-back}\end{bmatrix}$, which regulated decreasing transparency with decreasing height of the rightmost transparent vowel. The average error was minimal for each of the heuristic grammars with the exception of Hungarian. The higher error value is due to the BTi forms and the BTe: forms. For the BTe: forms, the lexicon displays a 42% rate of application, while the model predicts only a 10% rate of application. As Hayes and Londe 2006 note, there are very few data on this kind of violation profile. Furthermore, the investigators showed that Hungarian speakers applied vowel harmony far less to BTe: forms when they were given a wug test. The lexicon shows that BTi forms harmonize 25% of the time, but the distance-based weights predict that they harmonize 35% of the time; this discrepancy is not something I can currently explain.

As Figure 6 reveals, the heuristic distance-based markedness constraints for liquid dissimilation in Latin display inverse-exponential decay: the difference between the weight in distance constraint 1 and distance constraint 2 is greater than that between distance constraint 2 and distance constraint 3.

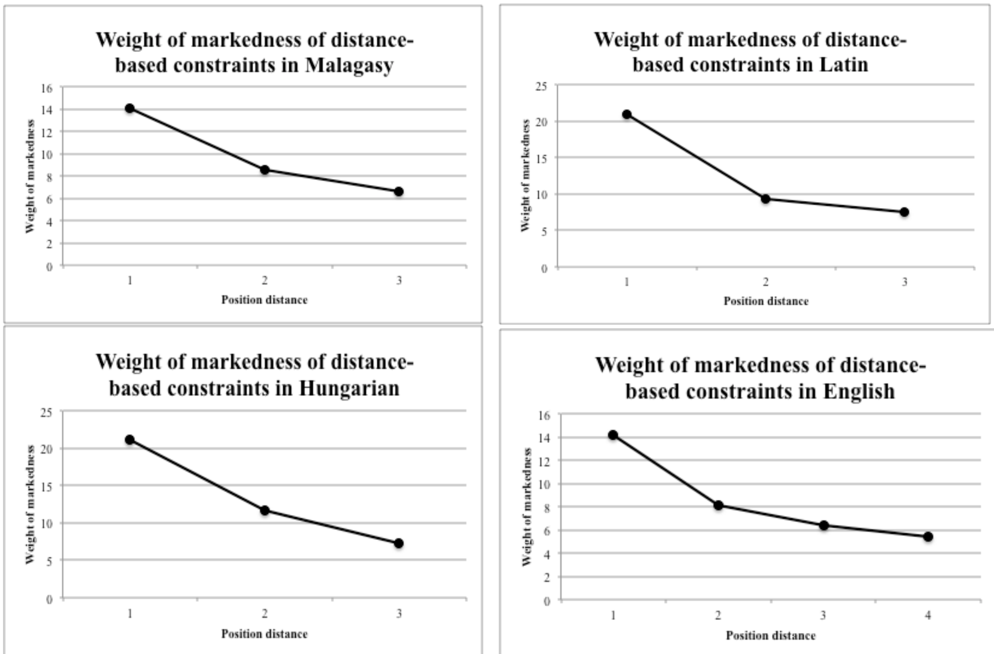


Figure 7: *inverse-exponential decrease in the weight of markedness across the four languages*

The graphs shown above reveal that the weight of markedness appears to decrease inverse-exponentially with greater transparent distance across all four languages. As we will see further in section 4.3., these plots suggest that a negative power function gives a better fit to the data than a linear function does in scaling the weight of markedness.

Recall that we want to find good values for three parameters: the weight of markedness w_m, the weight of faithfulness w_f and the decay parameter *k* of the decay function $d(x) = 1/x^k$. To do this, I used Microsoft Excel's Solver, which has nonlinear curve-fitting capabilities (see below). The expression that we are finding optimal parameter values for is the formula for the probability of process application, posed as a function of distance:[13]

[13] Note that the expression in (41) is slightly more complicated in the case of vowel harmony in Hungarian since our model of it includes constraints that regulate the height effect.

(34)

$$P\big(d(x)\big) = \frac{1}{1 + e^{-(w_m * d(x) - w_f)}} = \frac{1}{1 + e^{-(\frac{w_m}{x^k} - w_f)}}$$

The probability of faithfulness is simply 1 - *P*(*d*(*x*)). Note that the formula in (34) is the same as the formula in (21), except that now the weight of markedness w_m is multiplied by *d*(*x*), thus scaling its value to grow smaller with increasing distance. The function shown above is crucially shaped like an asymmetrical sigmoid curve, mimicking roughly the data distributions seen in Section 2. Its sigmoidal shape is shown in the figure below:

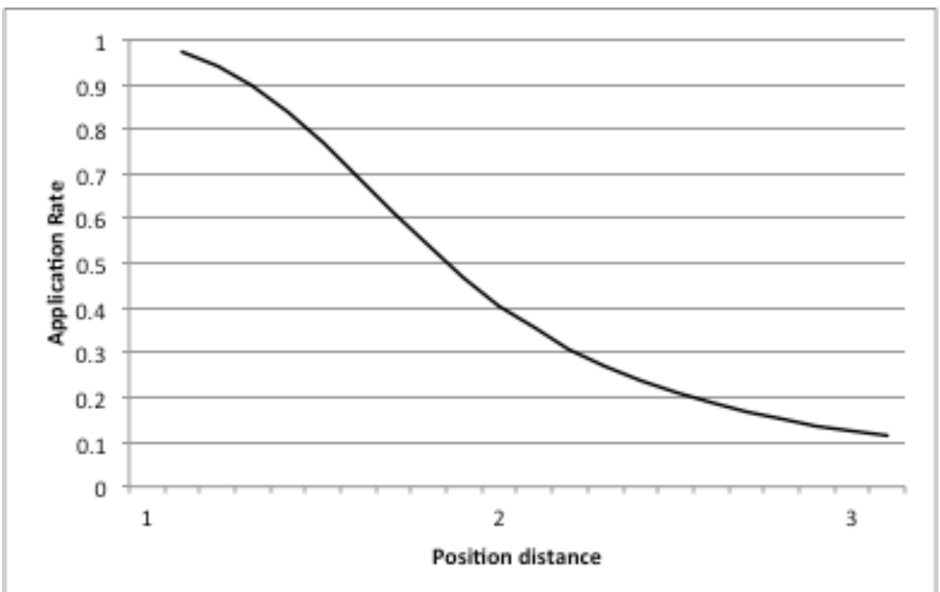


Figure 8: *graph of P*(*d*(*x*)) *with* w_m, w_f, *and k set to 9, 5, and 1 respectively*

We will see later that a more typical sigmoid curve such as $\frac{1}{1+e^{-(w_m * x - w_f)}}$ (where w_m is negative) does a poorer job fitting the data.

To optimize the model, I calculated the natural logarithm of each of the probabilities predicted by the model, and multiplied the result by the frequency count of the type of surface form (e.g., surface forms containing an [lσ^{2}l] subsequence in Latin). I then summed the resulting values over all of the violation profiles, yielding a negative number with a large absolute value (hereafter called the *log likelihood*). Microsoft Excel's Solver then determines the values of w_m, w_f, and *k* that minimize the absolute value of the log likelihood using Newton's Method. Doing this determines the parameter values that maximize the model-predicted probability of the observed forms and minimize the probabilities of the unobserved forms. Della Pietra, Della Pietra, and Lafferty 1997 show that the surface defined by the probabilities of sets of surface forms is convex over the space of constraint weights. In other words, the space is such that an iterative descent algorithm would not get stuck in a local optimum.

The Solver requires the user to input reasonable starting estimations of the model parameters w_m, w_f, and *k*. I used as an estimate the weight of the most local heuristic, distance-based constraint, as well as the weight of the heuristic faithfulness constraint, and an arbitrary decay parameter value, $k = 1$. The Solver in turn outputted the following:

Vowel dissimilation in Malagasy	
w_m:	12.12
w_f:	6.61
k :	0.9
Average error:	0.000

Liquid dissimilation in Latin	
w_m:	16.72
w_f:	0.40
k :	3.2
Average error:	0.000

Vowel harmony in Hungarian	
w_m:	44.36
w_f:	37.08
k :	0.3
Average error:	0.057

Liquid dissimilation in English using syllabic distance	
w_m:	17.89
w_f:	18.19
k :	0.1
Average error:	0.003

Liquid dissimilation in English using position distance	
w_m:	9.39
w_f:	3.73
k :	1.4
Average error:	0.004

Table 12: *distance-based decay modeling results using a negative power function with language-specific* w_m, w_f, *and k*

Provided that the model is free to fit *k* to each language, we find based on the table above that there is little to no distinction in fit between using syllabic distance (defined in (28)) and using position distance (defined in (29)) as a metric. Again, error here is defined as the absolute difference between the observed probabilities and the predicted probabilities averaged over the violation profiles of each of the languages. Since the position of the trigger within the syllable was significant in affecting application rates in Latin and English, I decided only to model the forms in either language in which the trigger was in onset-initial position (since there were few data on forms with the trigger in onset-noninitial or coda position). Note that the model for Hungarian also contains the height-based constraints posited in Table 11, whose weights were allowed to vary freely as the Solver fit the model to the data. For the case of local liquid dissimilation in English, the weight of markedness was multiplied by *d*(0.1). Observe that the scale yields just as good a fit as the distance-based constraint family does.

Recall that Kimper 2011 scales the weight of a spread constraint for vowel harmony in Hungarian by multiplying it by a constant in the interval (0, 1) *x* times over, where *x* is transparent distance. Kimper's method is effectively the same as scaling the weight of markedness by the decay function $d(x) = 1/k^x$, where *x* is transparent distance and k is a positive real-valued parameter. Using such a decay function yields the following results:

Vowel dissimilation in Malagasy		Liquid dissimilation in Latin	
w_m:	23.82	w_m:	138.29
w_f:	3.05	w_f:	0.07
k :	2.7	k :	10.0
Average error:	0.000	Average error:	0.000
Vowel harmony in Hungarian		**Liquid dissimilation in English using syllabic distance**	
w_m:	33.59	w_m:	9.28
w_f:	11.03	w_f:	2.31
k :	1.8	k :	4.6
Average error:	0.056	Average error:	0.011
		Liquid dissimilation in English using position distance	
		w_m:	12.75
		w_f:	4.39
		k :	4.6
		Average error:	0.020

Table 13: *distance-based decay modeling results using* $d(x) = 1/k^x$
with language-specific w_m, w_f, *and k*

A comparison between Table 12 and Table 13 reveals that $d(x) = 1/x^k$ fits the data either equally or even better than $d(x) = 1/k^x$, regardless of whether we adopt syllabic distance or position distance as our distance metric. While there is not a great deal of distinction between the accuracy of the two models, I will continue to use the negative power function as the decay function for the rest of the analysis.

This section shows that having a rather unconstrained model—one with three language-specific parameters—yields good fit to the data at large. We now turn to the next section, which demonstrates that we can constrain the model by reducing the number of parameters while still maintaining great accuracy.

4.2.2 Decay functions with universal settings for *k*

While the model with three language-specific parameters—the weight of markedness, the weight of faithfulness, and the decay parameter *k*—model the decay effect across the four languages with great accuracy, one wonders if a model with fewer language-specific parameters can also provide accurate predictions. Within the Maximum Entropy Harmonic Grammar, the weight of markedness and the weight of faithfulness are acquired by the learner based on the forms that they are exposed to, and are thus expected to be language specific. On the other hand, why should *k* be assumed language-specific? Thus far, we have not said anything about whether the learner needs to acquire a particular decay function (i.e., a particular value for *k*) depending on the forms they are exposed to. As we had seen in the previous section, distance-specific weights of markedness decrease with distance in inverse-exponential fashion across the four surveyed languages. We therefore posited a decay function—a negative power function—to capture the four language patterns. Beyond this, decay rates differed to some degree on a language-by-language basis, in that the best-fit values for *k* differed across languages. In spite of this, could we account for differences in decay rate using only the weight of markedness and faithfulness? In other words, while letting constraint weights vary freely, can we fix the value of *k* across the four languages—yielding a universal decay function—and still adequately model the data in each of them?

To answer this question, I ran several optimizing trials using the Solver. For each language, I fixed the value of *k* and set the Solver to minimize the absolute value of the log likelihood only by varying the weight of markedness and the weight of faithfulness. The trials led to the formation of *k-basins*, shown below in Figure 9 and 10. The *y*-axis is the absolute difference between observed and predicted probabilities averaged over violation profiles.

Suppose we used position distance as our distance metric. The resulting *k*-basins are shown below:

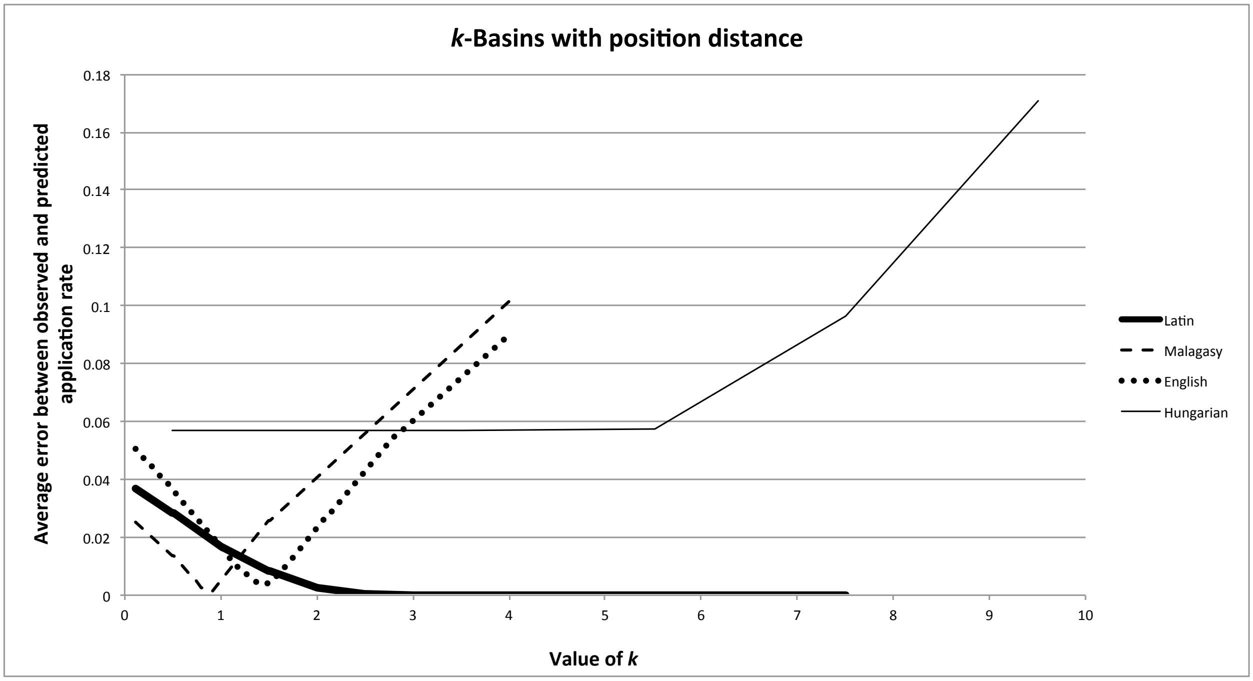


Figure 9: *k-basins with position distance*

For fixed values of *k* between 0.2 and 8 and in increments of 0.1, the Solver determined the weights that minimize the absolute value of the log likelihood.[14] I plotted the fixed value of *k* of a particular trial against the averaged error, or the absolute difference between observed and predicted probabilities averaged over the number of violation profiles in the language. The *k*-basins created for the four languages show that for liquid dissimilation in Latin and vowel harmony in Hungarian, a wide range of values for *k* lead to relatively minimal error. (Once again, note that the error for Hungarian is higher due to the BTi and BTe: forms). On the other

[14] For values of *k* at or below 0.2., the Solver was unable to converge upon finite values for the weights.

hand, vowel dissimilation in Malagasy and liquid dissimilation in English are more delicate: few values of *k* lie at the bottom of the two basins. This restricts the set of possible values for *k* that we can draw from that are satisfactory for modeling the data across the four languages. Nevertheless, the graph suggests that there exist values for *k* such that, if the function were to take on such value, it would predict the probabilities of the observed data reasonably well. Such *k* resides around *k* = 1.1.

I set the Solver to find the single value of *k* that minimizes the sum of the absolute values of the log likelihoods for all of the languages, letting constraint weights vary. The results were as follows:

Vowel dissimilation in Malagasy		Liquid dissimilation in Latin	
w_m:	10.24	w_m:	9.00
w_f:	4.66	w_f:	2.63
k :	**1.1**	k :	**1.1**
Average error:	0.010	Average error:	0.015
Vowel harmony in Hungarian		**Liquid dissimilation in English**	
w_m:	27.04	w_m:	10.90
w_f:	19.76	w_f:	5.28
k :	**1.1**	k :	**1.1**
Average error:	0.057	Average error:	0.006

Table 14: *distance-based decay modeling results using a negative power function, position distance, and a universal setting of k to k = 1.1*

As the table shows, we can fix a single value of *k* at *k* = 1.1 at fairly little cost: each of the models is still able to predict observed data quite accurately (see next section for a plot).

Now suppose that we used syllabic distance as the distance metric. The resulting *k*-basins are shown below:

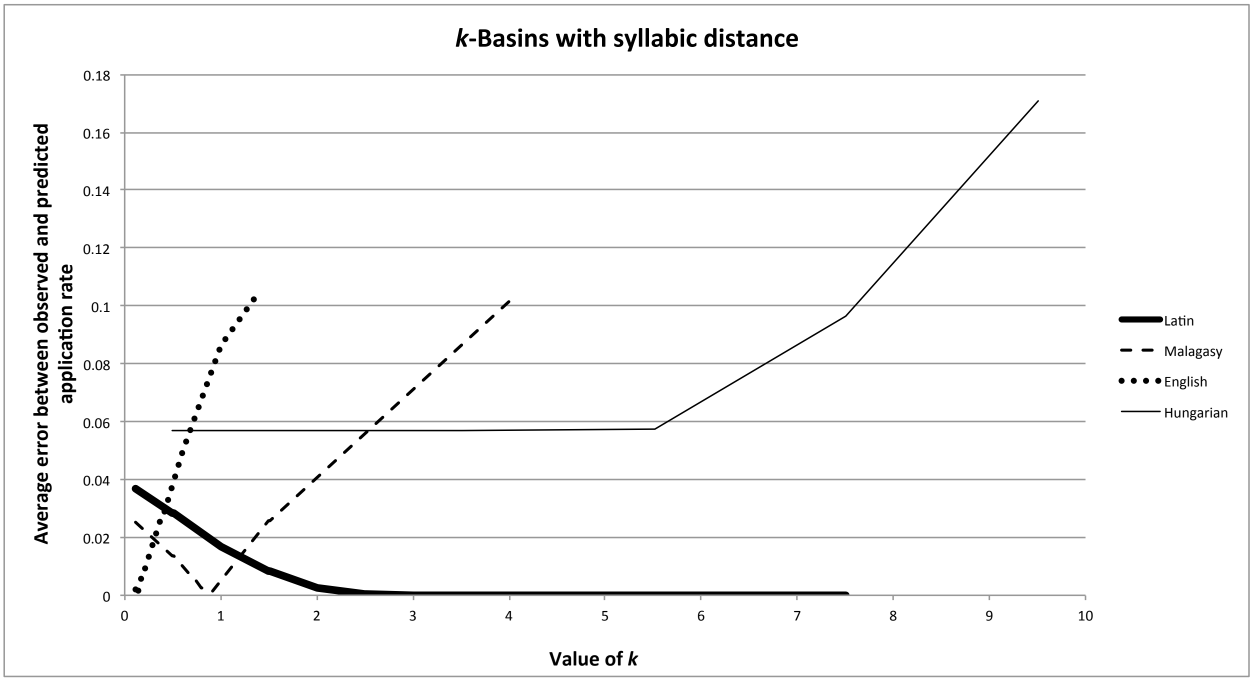


Figure 10: *k-basins with syllabic distance*

The *k*-basin for liquid dissimilation in English migrated to the left, with the bottom of the basin hovering around $k = 0.1$. This, presumably, is due to the fact syllabic distance treats local application of liquid dissimilation in English differently than it does local application of the other three processes: while local violations of markedness constraints incur a violation of $d(0.1)$ in liquid dissimilation in English, local violations incur a violation of $d(1)$ in the other languages. There is a striking decrease in accuracy in fitting the data from English as *k* takes on larger values, and as a result, constraining the model so as to fix *k* crosslinguistically yields overall decrease in modeling accuracy across languages:

Vowel dissimilation in Malagasy		Liquid dissimilation in Latin	
w_m:	26.71	w_m:	18.23
w_f:	21.39	w_f:	13.01
k :	**0.3**	k :	**0.3**
Average error:	0.020	Average error:	0.033
Vowel harmony in Hungarian		**Liquid dissimilation in English**	
w_m:	46.84	w_m:	6.12
w_f:	39.56	w_f:	6.54
k :	**0.3**	k :	**0.3**
Average error:	0.056	Average error:	0.020

Table 15: *distance-based decay modeling results using a negative power function, position distance, and a universal setting of k to k = 1.1*

The Solver determined $k = 0.3$ to be the best crosslinguistically-fit k. The decrease in accuracy of using syllabic distance with fixed k is thus observable from comparing average error in Table 15 with average error in Table 14:

Vowel dissimilation in Malagasy		Liquid dissimilation in Latin	
Average error using position distance:	0.010	Average error using position distance:	0.015
Average error using syllabic distance:	0.020	Average error using syllabic distance:	0.033
Vowel harmony in Hungarian		**Liquid dissimilation in English**	
Average error using position distance:	0.074	Average error using position distance:	0.006
Average error using syllabic distance:	0.074	Average error using syllabic distance:	0.020

Table 16: *comparing of models with position distance against those with syllabic distance*

In each of the cases except for vowel harmony in Hungarian, the average error that results from using syllabic distance is at least double that which results from using position distance. I therefore take position distance to be the superior unit of input to the decay function $d(x)$.

Summing up, we find thus far that using $d(x) = 1/x^{1.1}$, where x is position distance, leads to an accurate model of the decay effect across the four surveyed languages. Since the best fixed k is strikingly close to 1, using $d(x) = 1/x$ still yields an overall good fit to the data:

Vowel dissimilation in Malagasy		Liquid dissimilation in Latin	
w_m:	11.06	w_m:	9.29
w_f:	5.51	w_f:	3.15
k :	**1.0**	k :	**1.0**
Average error:	0.005	Error:	0.017
Vowel harmony in Hungarian		Liquid dissimilation in English	
w_m:	27.25	w_m:	10.78
w_f:	19.97	w_f:	5.57
k :	**1.0**	k :	**1.0**
Error:	0.057	Error:	0.016

Table 17: *distance-based decay modeling results using a negative power function, position distance, and a universal setting of k to k = 1*

This shows that simply dividing the weight of markedness by the determined count of transparent distance produces a fairly accurate representation of distance-based decay.

4.3 Comparing performance of the inverse exponential function to a linear function

McPherson and Hayes (submitted) observe that the rate of vowel harmony in Tommo So decreases with morphological distance—that is, harmony is less likely to apply to a verbal suffix that attaches farther from the stem than one that attaches closer to it. McPherson and Hayes scale the weight of markedness using a linear function in order to derive the morphological decay effect. Let us consider a model of application rate based on transparent distance when it is augmented with a simple linear function; i.e., suppose the decay function for distance-based decay is taken to be $d(x) = x$:

(35)

$$P\big(d(x)\big) = \frac{1}{1 + e^{-(w_m * x - w_f)}}$$

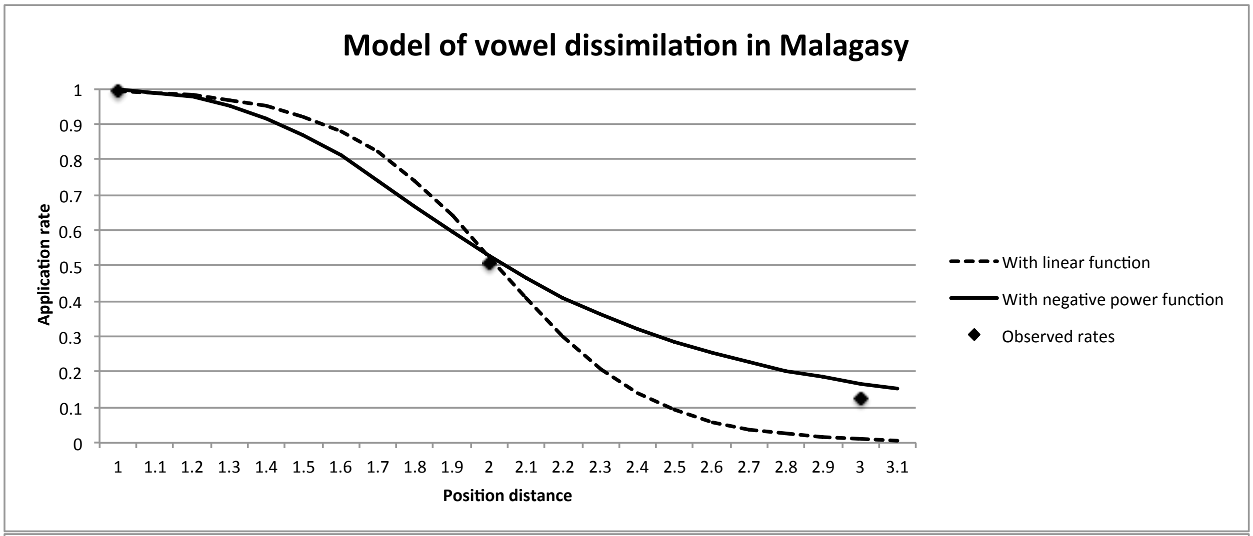
Notice that this function lacks a third parameter, *k*. Instead, properties of the linear function are absorbed by constraint weights: the intercept of the linear function is absorbed by the weight of faithfulness, while its slope (presumably negative) is absorbed by the weight of markedness. The model thus has two parameters. How does a model that lacks *k* entirely fare against one that has crosslinguistically-fixed *k*?

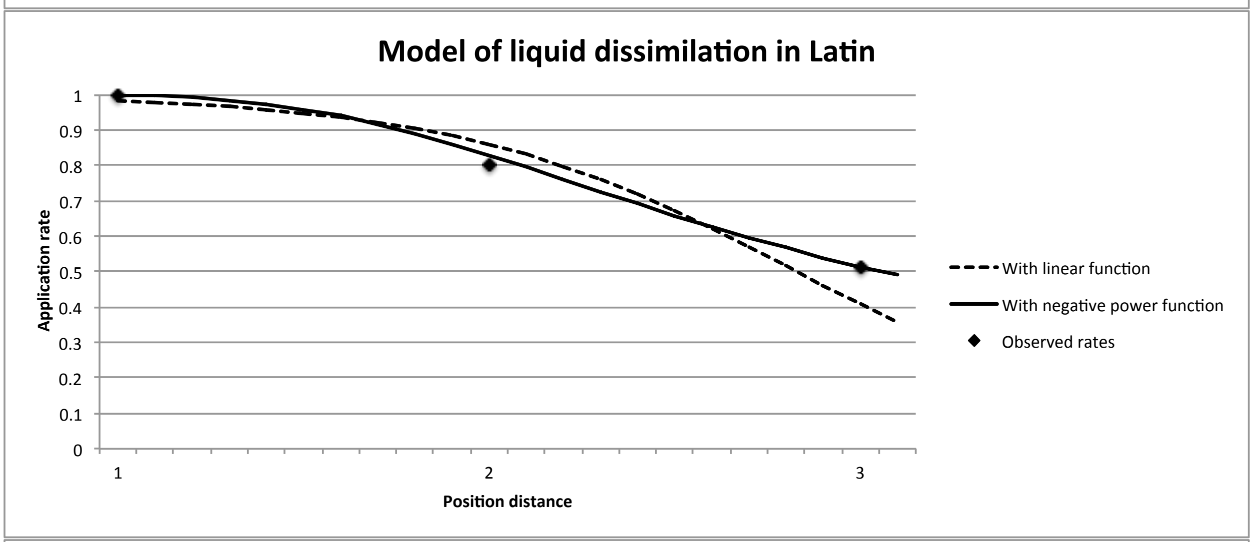
We can replicate the optimizing process by setting the Solver to find the values of the weight of markedness and faithfulness that minimize the absolute value of the log likelihood of the dataset. The results are shown in the table below:

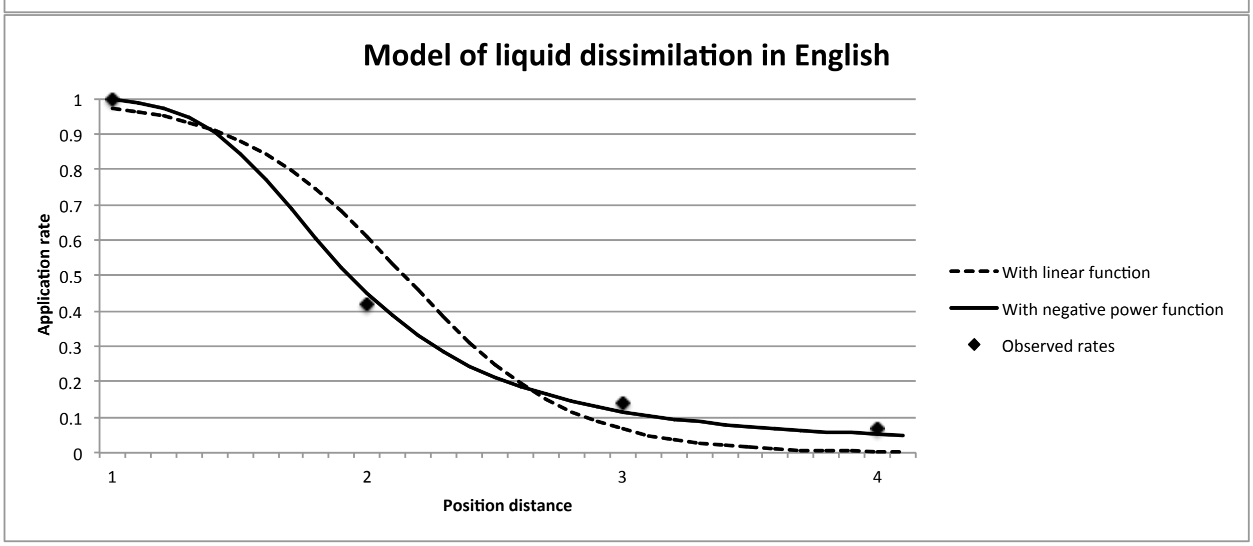
Vowel dissimilation in Malagasy		Liquid dissimilation in Latin	
w_m:	4.75	w_m:	2.18
w_f:	9.60	w_f:	6.27
Average error:	0.045	Average error:	0.058
Vowel harmony in Hungarian		Liquid dissimilation in English	
w_m:	4.54	w_m:	3.09
w_f:	11.83	w_f:	6.63
Average error:	0.057	Average error:	0.088

Table 18: *distance-based decay modeling results using a linear function and position distance*

The table above shows that replacing the negative power function for a linear one—thus getting rid of *k* entirely—yields too simple of a model. The linear function does a poor job predicted observed probabilities in vowel dissimilation in Malagasy and liquid dissimilation in Latin and English, especially relative to the negative power function. (I do not know why it does well with vowel harmony Hungarian. Perhaps it is because the model also contains the height-based constraints for Hungarian, allowing the Solver more freedom to fit the data.) The table shows that abandoning the negative power function for a linear one results in a striking decrease in accuracy. The decrease in accuracy is as well revealed in the following plots, with one model of application rate containing the negative power function and the other containing the linear function:


Model of vowel dissimilation in Malagasy
Application rate
Position distance
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2 2.1 2.2 2.3 2.4 2.5 2.6 2.7 2.8 2.9 3 3.1
With linear function
With negative power function
Observed rates


Model of liquid dissimilation in Latin
Application rate
Position distance
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 2 3
With linear function
With negative power function
Observed rates


Model of liquid dissimilation in English
Application rate
Position distance
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 2 3 4
With linear function
With negative power function
Observed rates

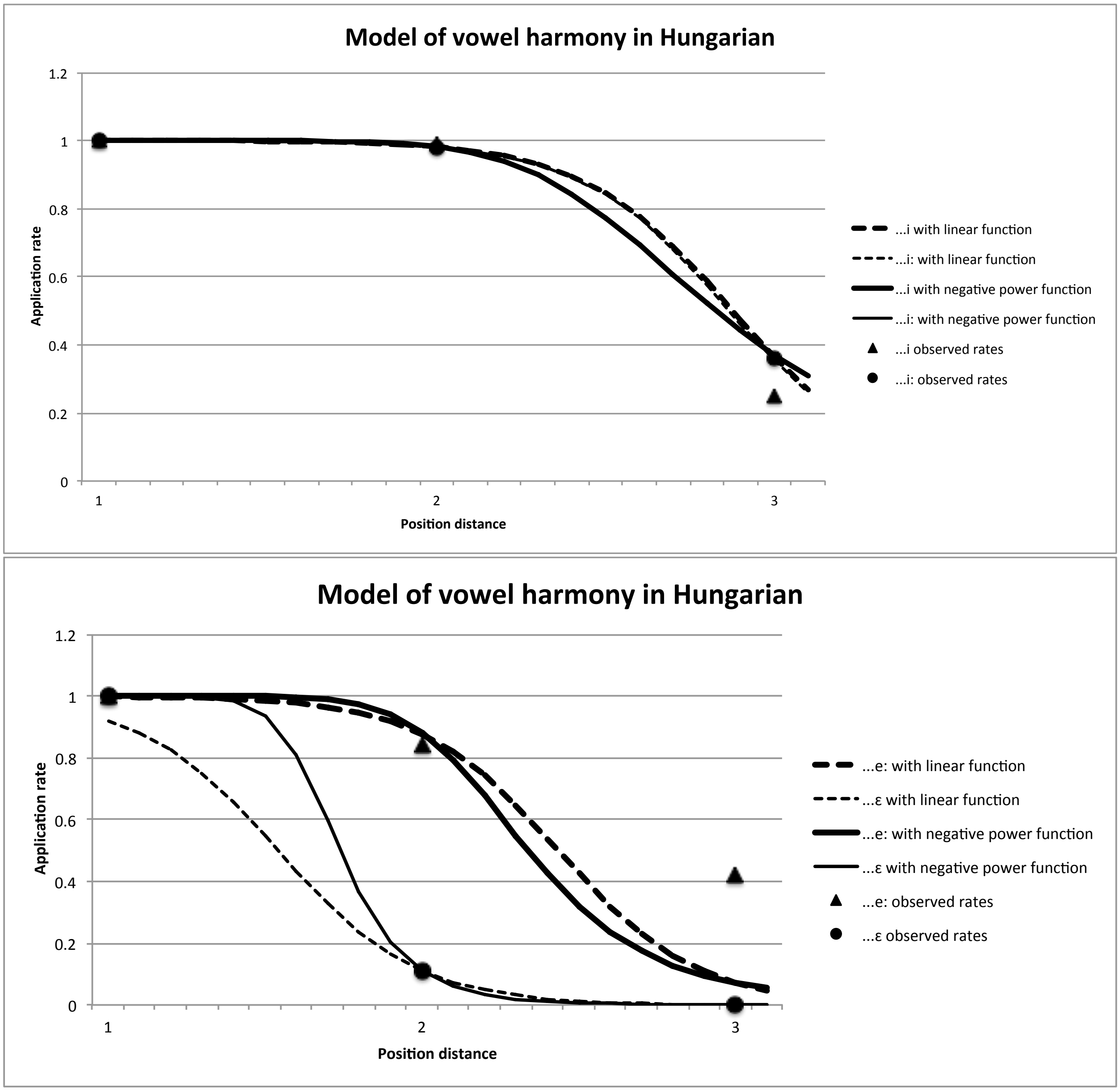


Figure 11: *plots that compare models using the linear function against those using the negative power function where k is set to k = 1.1*

The plots show that the model using negative power function performs better than the model using the linear function for all the covered processes except vowel harmony in Hungarian, and in such case the two models perform roughly equally. Note that Hungarian was broken up into two plots: one with transparent vowel heights /i/ and /i:/ and another with transparent vowel heights /e:/ and /ɛ/. For the first of these plots, the graphs of the height-based models with the negative power function overlap entirely, as do the graphs of the height-based models with the linear function.

Finally, we can use the AIC to determine whether or not the increase in fit upon adopting a negative power function is worth having the extra parameter, the decay constant *k*. The sum of log likelihoods that result from fitting the linear model to each of languages is -52.4 + (-321.5) + (-116.5) + (-201.6) = -692.0. Since linear model has two language-specific parameters (namely, the weight of markedness and the weight of faithfulness) to cover the four processes as well as the four weights to capture the height effect in Hungarian, the AIC measure for the linear model is thus 2(2*4+4) – 2(-692.0) = 1408.0. Furthermore, the sum of log likelihoods that result from fitting the negative power function to each of languages is -48.8 + (-313.4) + (-102.2) + (-201.5) = -665.9. The negative power function has an additional parameter (i.e., the invariant decay constant *k*), and so its AIC measure is 2(2*4+4+1) – 2(-665.9) = 1357.8. Comparing AICs for the two models, we find that the AIC for the negative power function is lower than that of the linear model. Based on these results, I reject the linear function as an adequate model of distance-based decay, and conclude that the effect is best accounted for using a negative power function. Why a linear model suffices as an account of morphological distance in vowel harmony in Tommo So, but not of transparent distance in the various processes surveyed here, is a question that I leave to further research.

5 Conclusions and future directions

Toward the beginning of this thesis, we surveyed three different long-distance phonological processes in four languages, and found that each of them displays distance-based decay. We opted not to posit distance-based constraint families, since it is too powerful, and dismisses as coincidental the learner's bias for learning high markedness weights for local constraints and successively lower markedness weights with greater distance. Instead, we follow Kimper 2011 in positing a scaling factor that scales the weight of markedness with increasing transparent distance.

Likelihood ratio tests over generalized linear models of the data revealed that the number of transparent syllables, rather than the number of transparent segments, is significant in producing the decay effect for the surveyed processes. Furthermore, heuristic distance-based constraints reveal that the weight of markedness decreases in inverse-exponential fashion. In turn, we posited for our scaling factor a decay function $d(x)$—a negative power function—taking the form $d(x) = 1/x^k$, where *x* is position distance and *k* is a positive real-valued parameter.

Explorations with model fitting show that the decay parameter *k* need not be language-specific; in fact, fixing *k* at roughly $k = 1.1$ yields a model that is able to accurately predict observed probabilities in all the surveyed languages. Since decay rates differ across languages, the ability of models with fixed *k* to accurately predict crosslinguistic data suggests that decay rate differences need not be attributed to language-specific decay parameters, but rather to language-specific differences between the weight of markedness and the weight of faithfulness. Nevertheless, *k* is necessary to ensure good fit between model predictions and the observed data: opting to replace the negative power function with the simple linear function $d(x) = x$ comes at the price of strikingly reduced accuracy in model predictions.

5.1 On the unit of distance: retroflex assimilation in Sanskrit

In Sanskrit, long-distance retroflex harmony can be observed. Shown below are data on how the nominal/adjectiveal suffix -/ana/ is affected by the process:

	UR	Nominal/adjectival form	Gloss
Faithful items			
(52a)	/naj+ana/	[naj-ana]	‘leading’
(52b)	/devajadʒ+ana/	[devajadʒ-ana]	‘worshipping of gods’
Items that display retroflex assimilation			
(53a)	/rak**ʂ**+a**n**a /	[rak**ʂ**-a**ɳ**a]	‘protection’
(53b)	/k**r̩**p+a**n**a /	[k**r̩**p-a**ɳ**a]	‘inclined to grieve’
(53c)	/ak**r**am+a**n**a /	[ak**r**am-a**ɳ**a]	‘attack’
(53d)	/k**ʂ**aj+a**n**a/	[k**ʂ**aj-a**ɳ**a]	‘annihilating’
Items with opaque segments			
(54a)	/ce**ʂʈ**+a**n**a/	[ce**ʂʈ**-a**n**a]	‘stirring’
(54b)	/**r**o**c**+a**n**a/	[**r**o**c**-a**n**a]	‘shining’

The triggers of retroflex assimilation are /ʂ/ and /r/ (the latter of which can be consonantal or syllabic, as comparing (53a) and (53b) shows), and the only target is /n/. /n/ can retroflect into [ɳ] if it occurs after a trigger. Schein and Steriade 1986 claim there are a variety of opaque intervening segments, including dentals, retroflexes, and palatal consonants. Finally, the process can apply over suffix boundaries to a variety of suffixes, as well as across root boundaries within compounds.

Data from the Digital Corpus of Sanskrit (http://kjc-fs-cluster.kjc.uni-heidelberg.de/dcs/) show that retroflex assimilation in Sanskrit displays distance-based decay, the figures for which are shown below:

Retroflex assimilation in Sanskrit: /{ʂ,r}σ^n+ana/ → [{ʂ,r}σ^n-**aɳa**]			
Transparent syllables *n*	Faithful forms	Harmonized forms	% of harmon’d forms
n = 0	1	905	0.99
n = 1	19	422	0.96
n = 2	54	35	0.39
n = 3	12	4	0.25

Table 19: *figures for distance-based decay in retroflex assimilation in Sanskrit based on transparent syllables*

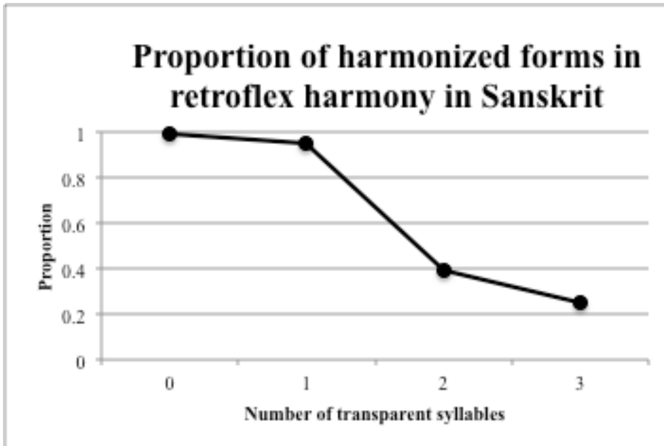


Figure 12: *graph of distance-based decay in retroflex assimilation in Sanskrit based on transparent syllables*

As it turns out, retroflex assimilation is special in that the number of transparent syllables is not significant in producing the decay effect when compared against the number of transparent segments. When I ran likelihood ratio tests over generalized linear models of the data (i.e., the full model and models omitting each variable), I found that if I included both syllable count and segment count in the model, the model that was yielded from applying the algorithm discussed at the beginning of Section 4.1 was one in which segment count was preserved and syllable count was eliminated. Furthermore, the best model that was yielded from starting with a full model containing only syllable count had an AIC of 611.32, whereas the best model that was yielded from starting with only segment count had an AIC of 590.96.

Figures for application rate as a function of segment count are shown below:

Retroflex assimilation in Sanskrit: /{ʂ,r}φ^n+ana/ → [{ʂ,r}φ^n-**aɳa**]			
Transparent segments *n*	Faithful forms	Harmonized forms	% of harmon'd forms
n = 1	2	903	0.99
n = 2	2	78	0.97
n = 3	16	308	0.95
n = 4	9	41	0.82
n = 5	35	29	0.45
n = 6	18	2	0.10
n = 7	10	1	0.09
n = 8	3	0	0.00

Table 20: *figures for distance-based decay in retroflex assimilation in Sanskrit based on transparent segments*

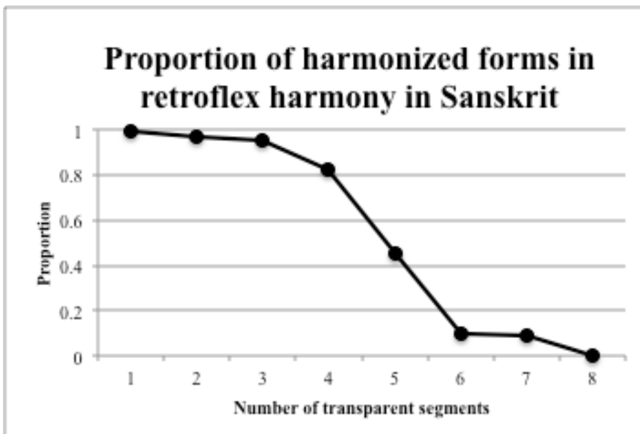


Figure 13: *graph of distance-based decay in retroflex assimilation in Sanskrit based on transparent segments*

Figure 13 shows that application rate of retroflex harmony in Sanskrit decreases roughly sigmoidally as a function of the number of transparent segments.

In addition, triggers in coda position were significantly associated with lower application ($p = 0.02$), but triggers in onset-noninitial position were not significant in influencing application rate. With the exception of triggers and palatal nasals, I found that intervening dentals, retroflexes, and palatal consonants were significant in blocking application (all $p < 0.001$).

It is intriguing that the number of transparent segments, but not transparent syllables, is significant to producing the decay effect in retroflex assimilation, while the number transparent syllables is significant in all the other processes surveyed in this work. One might speculate that it is related to retroflex assimilation in Sanskrit being post-lexical. The three processes covered in the previous sections are non-post-lexical: they apply only over suffix boundaries and target particular suffixes. Sanskrit, on the other hand, can apply over word boundaries, and does not apply to a particular suffix in the exclusion of others. It could be that the learner phonologizes phonetic distance into phonological distance—a discrete count, the number of transparent syllables—when the long-distance process is non-post-lexical in nature, and is sensitive to structure. In post-lexical long-distance phonological processes, phonetic distance is not phonologized as the result of a lack of sensitivity to morphological environment.

6 References

Beaujardière, J. 2004. Online corpus of the Encyclopedia of Madagascar and Malagasy Dictionary.

Bennett, W. 2013. Dissimilation, Consonant Harmony, and Surface Correspondence. Ph.D. Dissertation, Rutgers.

Berko, J. 1958. The child's learning of English morphology. *Word*, 14: 150-177.

Blevins, J. 2001. Evolutionary phonology: The emergence of sound patterns. Cambridge: Cambridge University Press.

Boersma, P. 1997. How we learn variation, optionality, and probability. *Institute of Phonetic Sciences Proceedings* 21: 43-58.

Cser, András. 2010. The -alis/aris- allomorphy revisited. In *Variation and change in morphology: selected papers from the 13th international morphology meeting*, ed. by D. Kastovsky F. Rainer, W. U. Dressler & H. C. Luschützky, 33-51. Amsterdam/Philadelphia: John Benjamins.

Della Pietra, Stephen, Vincent J. Della Pietra, and John D. Lafferty. 1997. Inducing features of random fields. IEEE Transactions on Pattern Analysis and Machine Intelligence 19:380–393.

Digital Corpus of Sanskrit: http://kjc-fs-cluster.kjc.uni-heidelberg.de/dcs/

Goldsmith, J. 1976. Autosegmental Phonology. PhD dissertation, MIT.

Goldwater, S. & Johnson, M. 2003. Learning OT constraint rankings using a maximum entropy model. In the *Proceedings of the Stockholm workshop on variation within Optimality Theory.*

Kimper, W. 2011. Competing triggers: Transparency and opacity in vowel harmony. Ph.D. dissertation, University of Massachusetts, Amherst.

Hansson, G. 2001. Theoretical and typological issues in consonant harmony. Doctoral dissertation, UC Berkeley.

Hayes, B. & Londe, Z. 2006. Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology* 23: 59-104.

Hayes, B. & Wilson, C. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379-440.

Hayes, B., Zuraw, K., Siptar, P. & Londe, Z. 2009. Natural and Unnatural Constraints in Hungarian Vowel Harmony. *Language* 85: 822-863.

Linzen, T., Kasyanenko, S., & Gouskova, M. 2013. Lexical and phonological variation in Russian prepositions. *Phonology* 30(3), 453-515.
McCarthy, J. 1986. OCP Effects: Gemination and Antigemination. *Linguistic Inquiry* 17, 207-264.

Martin, A. 2005. The effects of distance on lexical bias: sibilant harmony in Navajo compounds. Master's thesis, UCLA.

Maxent Grammar Tool: http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/.

McPherson, L. & Hayes, B. Submitted. Relating application frequency to morphological structure: the case of Tommo So vowel harmony. Ms., UCLA.

Nordhoff, S., Hammarström, H., Forkel, R., & Haspelmath, M. 2013. "Malagasy". *Glottolog 2.2*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Oxford English Dictionary: www.oed.com

Perseus Digital Library: http://www.perseus.tufts.edu/hopper/

Prince, A. & Smolensky, P. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. Rutgers Center for Cognitive Science, Piscataway, New Jersey.
Smolensky, P. 1986. Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart, J. L. McClelland & the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Volume 1: Foundations. Cambridge, MA: MIT Press/Bradford Books. 194–281.

Schein, B. & Steriade, D. 1986. On geminates. *Linguistic Inquiry* 17: 691-744.

Smolensky, P. & Legendre, G. 2006. The harmonic mind: From neural computation to optimality-theoretic grammar. Cambridge, MA: MIT Press.

Vago, R. 1976. Theoretical implications of Hungarian vowel harmony. *Linguistic Inquiry* 7: 243-263.

Walker, R. & Mpiranya, F. 2006. On triggers and opacity in coronal harmony. *Proceedings of BLS 31*, ed. by Rebecca T. Cover and Yuni Kim: 383-394.

Zuraw, K. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28(2). 417–472.