Distance-Based Decay in Long-Distance Phonological Processes

Jesse Zymet

1. Introduction

Many long-distance phonological processes exhibit what I call *distance-based decay*: the likelihood of application decreases as transparent distance increases. This article provides a robust model of distance-based decay within Maximum Entropy Harmonic Grammar (Smolensky 1986, Goldwater and Johnson 2003), drawing from thousands of data reflecting three processes across four languages. I reject using distance-specific constraints, and instead posit a decay function that scales the weight of a single AGREE/DISAGREE constraint with increasing distance (Kimper 2011). Though decay rates differ empirically across the four languages, I capture such differences purely with the weight of markedness and the weight of faithfulness without having to fit the decay function to each language. I argue based on statistical measures that distance is best measured in syllables rather than segments, supporting Martin 2005. The decay function takes as input a modified version of syllable count and returns a value that is then multiplied by the weight of AGREE/DISAGREE. It takes the form of a negative power function, which I show performs better than a down-sloping linear function.

2. Long-distance phonological processes with distance-based decay

Distance-based decay is evident in a variety of long-distance assimilatory and dissimilatory processes. This section covers the effect as it arises in three such processes across four languages: rounding dissimilation in Malagasy, liquid dissimilation in Latin and English, and vowel harmony in Hungarian.

In Malagasy, a Western Austronesian language spoken primarily in Madagascar, distance-based dissimilation of rounding on the high vowels (hereafter referred to as rounding dissimilation) can be observed in data extracted from Beaujardière 2004, an online Malagasy dictionary (http://malagasyword.org).

	UR	Passive imperative verb form	Gloss
Faithful items			
(1a)	/bata+u/	[bata-u]	'lift'
(1b)	/fana+u/	[fana-u]	'heat'
Items undergoing local and nonlocal rounding dissimilation			
(2a)	/bab**u**+**u**/	[bab**u**-**i**]	'plunder'
(2b)	/t**u**v+**u**/	[t**u**v-**i**]	'fulfill'
(2c)	/t**u**da+**u**/	[t**u**da-**i**]	'prevent'
(2d)	/g**u**rabah+**u**/	[g**u**rabah-**i**]	'spluttering'
Items with opaque front vowels			
(3a)	/**turi**+**u**/	[**turi**-**u**]	'preach'
(3b)	/**ure**+**u**/	[**ure**-**u**]	'massage'

* Jesse Zymet, UCLA, jzymet@ucla.edu. I wish to thank Bruce Hayes, Kie Zuraw, Robert Daland, Ed Stabler, and audiences at UC Santa Cruz, the Agreement by Correspondence Conference at UC Berkeley, and the 32nd West Coast Conference in Formal Linguistics at the University of Southern California for helpful comments and suggestions. All mistakes are my own.

(1a) and (1b) reveal the underlying form of the passive imperative suffix to be -/u/. (2a) illustrates that if the passive imperative suffix attaches to a stem that ends in /u/, then it dissimilates, surfacing instead as -[i]. (2b) through (2d) illustrate that rounding dissimilation can apply across transparent material. /a/ is transparent to rounding dissimilation, while (3a) and (3b) illustrate that /e/ and /i/ are opaque to the process.

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
(4a)	0 syllables	/ba.b**u**.+**u**/	[ba.b**u**.-**i**]	'plunder'
	0 syllables	/**tu**.v+**u**/	[**tu**.v-**i**]	'fulfill'
	0 syllables	/d**u**.r+**u**/	[d**u**.r-**i**]	'burn'
(4b)	1 syllable	/**ru**.va.+**u**/	[**ru**.va.-**u**]	'palisade'
	1 syllable	/**u**n.da.n+**u**/	[**u**n.da.n-**i**]	'bolster'
	1 syllable	/**tu**.da.+**u**/	[**tu**.da.-**i**]	'prevent'
(4c)	2 syllables	/b**u**.ra.ra.h+**u**/	[b**u**.ra.ra.h-**u**]	'scattered'
	2 syllables	/k**u**.ta.ba.+**u**/	[k**u**.ta.ba.-**u**]	'disorder'
	2 syllables	/g**u**.ra.ba.h+**u**/	[g**u**.ra.ba.h-**i**]	'spluttering'

As the number of transparent syllables increases, the likelihood of rounding dissimilation decreases. The decay effect is displayed in a set of forms extracted from the online dictionary.

Rounding dissimilation in Malagasy			
Transparent syllables	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

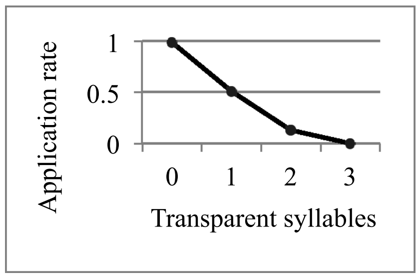

Table 1: *figures for distance-based decay in rounding dissimilation in Malagasy*
Figure 1: *graph of distance-based decay in rounding dissimilation in Malagasy*

Latin as well contains a long-distance phonological process, liquid dissimilation, regulating the distribution of [l] and [r] in particular contexts. The process can be observed in the following data extracted from the Perseus Digital Library (http://www.perseus.tufts.edu/hopper/).

	UR	Adjective form	Gloss
Faithful items			
(5a)	/kib+a:lis/	[kib-a:lis]	'pertaining to food'
(5b)	/fanit+a:lis/	[fanit-a:lis]	'pertaining to a temple'
Items undergoing local and nonlocal liquid dissimilation			
(6a)	/so**l**+a:**l**is/	[so**l**-a:**r**is]	'pertaining to the sun'
(6b)	/wu**l**g+a:**l**is/	[wu**l**g-a:**r**is]	'pertaining to wheat'
(6c)	/**l**a:n+a:**l**is/	[**l**a:n-a:**r**is]	'pertaining to wool'
(6d)	/**l**apid+a:**l**is/	[**l**apid-a:**r**is]	'pertaining to rocks'
Items with opaque /r/			
(7a)	/**l**ito**r**+a:**l**is/	[**l**ito**r**-a:**l**is]	'pertaining to the seashore'
(7b)	/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 (5a) and (5b). If the suffix attaches to a stem ending in /l/, then it dissimilates from the stem, surfacing instead as -[a:ris], as in (6a). (6b) through (6d) illustrate that stem-internal /l/ need not be stem-final in order to trigger liquid dissimilation on -/a:lis/. (7a) and (7b) show that /r/ blocks liquid dissimilation.

The following data illustrate the sensitivity of the process to distance.

	Transparent Syllables	UR	Adjective form	Gloss
(8a)	0 syllables	/so.**l**+aː.**l**is/	[so.**l**-aː.**r**is]	‘pertaining to the sun’
	0 syllables	/mu.**l**+aː.**l**is/	[mu.**l**-aː.**r**is]	‘pertaining to mules’
	0 syllables	/du.p**l**+aː.**l**is/	[du.p**l**-aː.**r**is]	‘pertaining to two’
(8b)	1 syllable	/pa.**l**e.+aː.**l**is/	[pa.**l**e.-aː.**l**is]	‘pertaining to chaff’
	1 syllable	/**l**aː.n+aː.**l**is/	[**l**aː.n-aː.**r**is]	‘pertaining to wool’
	1 syllable	/a.**l**e.+aː.**l**is/	[a.**l**e.-aː.**r**is]	‘pertaining to chance’
(8c)	2 syllables	/**l**ek.tu.+aː.**l**is/	[**l**ek.tu.-aː.**l**is]	‘pertaining to beds’
	2 syllables	/di.**l**u.wi.+aː.**l**is/	[di.**l**u.wi.-aː.**l**is]	‘pertaining to floods’
	2 syllables	/**l**a.pi.d+aː.**l**is/	[**l**a.pi.d-aː.**r**is]	‘pertaining to stone’

Separating forms based on the number of transparent syllables reveals that liquid dissimilation in Latin, like rounding dissimilation in Malagasy, is subject to distance-based decay.

Liquid dissimilation in Latin			
Transparent syllables	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

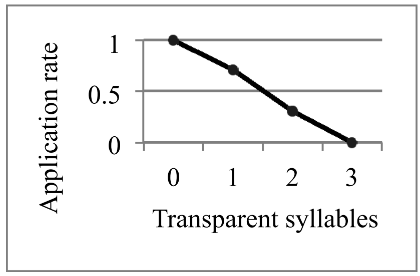

Table 2: *figures for distance-based decay in liquid dissimilation in Latin*
Figure 2: *graph of distance-based decay in liquid dissimilation in Latin*

Data extracted from the Oxford English Dictionary (www.oed.com) illustrate that English inherited long-distance liquid dissimilation from Latin.

	UR	Adjective form	Gloss
Faithful items			
(9a)	/dɪst+əl/	[dɪst-əl]	‘distal’
(9b)	/ejpɪk+əl/	[ejpɪk-əl]	‘apical’
Items displaying local liquid dissimilation			
(10a)	/soʊ**l**+ə**l**/	[soʊ**l**-ə**ɹ**]	‘solar’
(10b)	/vi**l**+ə**l**/	[vi**l**-ə**ɹ**]	‘velar’
(10c)	/kɑndə**l**+ə**l**/	[kɑndə**l**-ə**ɹ**]	‘condylar’
Items showing nonlocal liquid dissimilation			
(10d)	/**l**un+ə**l**/	[**l**un-ə**ɹ**]	‘lunar’
(10e)	/**l**əkjun+ə**l**/	[**l**əkjun-ə**ɹ**]	‘lacunar’
Items with opaque /ɹ/			
(11a)	/f**l**ɔ**ɹ**+ə**l**/	[f**l**ɔ**ɹ**-ə**l**]	‘floral’
(11b)	/æ**l**pɛst**ɹ**+ə**l**/	[æ**l**pɛst**ɹ**-ə**l**]	‘alpestral’

The underlying form of the English adjectival suffix, -/əl/, is revealed in (9a) and (9b). 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 (10a) through (10c). (10d) and (10e) illustrate nonlocal application. (11a) and (11b) demonstrate that /ɹ/ blocks the process.

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

	Trigger-target distance	UR	Adjective form	Gloss
(12a)	Same syllable	/soʊ**l**+ə**l**/	[soʊ**l**-əɹ]	'solar'
	Same syllable	/vi**l**+ə**l**/	[vi**l**-əɹ]	'velar'
	Same syllable	/kɑndə**l**+ə**l**/	[kɑndə**l**-əɹ]	'condylar'
(12b)	Adjacent syl.s	/**l**i.g+ə**l**/	[**l**i.g-ə**l**]	'legal'
	Adjacent syl.s	/p**l**ej.n+ə**l**/	[p**l**ej.n-əɹ]	'planar'
	Adjacent syl.s	/**l**u.n+ə**l**/	[**l**u.n-əɹ]	'lunar'
(12c)	1 transparent syl.	/**l**ej.bi.+ə**l**/	[**l**ej.bi.-ə**l**]	'labial'
	1 transparent syl.	/p**l**u.vi.+ə**l**/	[p**l**u.vi.-ə**l**]	'pluvial'
	1 transparent syl.	/**l**ə.kju.n+ə**l**/	[**l**ə.kju.n-əɹ]	'lacunar'

In this case, the number of transparent syllables is not exactly what divides the data in (12a) through (12c) into distinct classes. In both (12a) and (12b), for example, the trigger and the target are separated by zero transparent syllables, and yet the two kinds of forms are classified for different trigger-target distances. We will put this consideration aside, returning to characterizing the unit of distance for liquid dissimilation in English later on. The figures surrounding application sensitivity to distance are shown below.

Liquid dissimilation in English			
Transparent syllables	Faithful forms	Dissimilated forms	Proportion of dissim'd forms
Same syl.	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

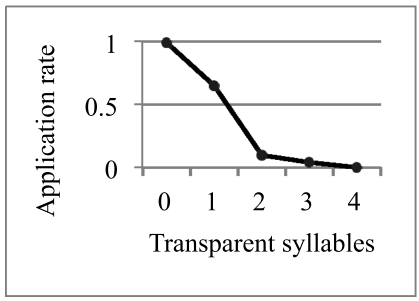

Table 3: *figures for distance-based decay in liquid dissimilation in English*
Figure 3: *graph of distance-based decay in liquid dissimilation in English*

Table 3 and Figure 3 reveal that the number of transparent syllables has an erosive effect on the likelihood of application for liquid dissimilation in English.

Distance-based backness harmony can be observed in Hungarian data (Hayes et al. 2009, http://www.linguistics.ucla.edu/people/hayes/HungarianVH/Index.htm), as shown by the following items.

	UR	Dative form	Gloss
Faithful items			
(13a)	/kɛrt+nɛk/	[kɛrt-nɛk]	'garden'
(13b)	/tsiːm+nɛk/	[tsiːm-nɛk]	'address'
Items that undergo local and nonlocal vowel harmony			
(14a)	/ɔbl**ɔ**k+n**ɛ**k/	[ɔbl**ɔ**k-n**ɔ**k]	'window'
(14b)	/kommunizm**u**s+n**ɛ**k/	[kommunizm**u**s-n**ɔ**k]	'Communism'
(14c)	/ɔpoʃt**o**li+n**ɛ**k/	[ɔpost**o**li-n**ɔ**k]	'apostolic'
(14d)	/b**o**riːteːk+n**ɛ**k/	[b**o**riːteːk-n**ɔ**k]	'envelope'
Items with opaque front rounded vowels			
(15a)	/ʃ**o**f**øː**r-n**ɛ**k/	[ʃ**o**f**øː**r-n**ɛ**k]	'chauffeur'
(15b)	/**ɔ**lbe:rl**øː**-n**ɛ**k/	[**ɔ**lbe:rl**øː**-n**ɛ**k]	'lodger'

Hungarian has a dative suffix, -/nɛk/, whose underlying form is apparent in (13a) and (13b). 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 (14a) and (14b). (14c) and (14d) illustrate nonlocal application across consonants and front unrounded vowels. On the other hand, (15a) and (15b) show that intervening front round vowels block vowel harmony.

As the data below suggest, vowel harmony in Hungarian is sensitive to distance, much like the other processes we have seen.

	Transparent Syllables	UR	Dative form	Gloss
(16a)	0 syllables	/ɔbl**ɔ**k+n**ɛ**k/	[ɔbl**ɔ**k-n**ɔ**k]	‘window’
	0 syllables	/bir**o**:+n**ɛ**k/	[bir**o**:-n**ɔ**k]	‘judge’
	0 syllables	/kommunizm**u**s+n**ɛ**k/	[kommunizm**u**s-n**ɔ**k]	‘Communism’
(16b)	1 syllable	/f**ɔ**se:n+n**ɛ**k/	[f**ɔ**se:n-n**ɛ**k]	‘charcoal’
	1 syllable	/ɔpoʃt**o**li+n**ɛ**k/	[ɔpost**o**li-n**ɔ**k]	‘apostolic’
	1 syllable	/m**a**:rti:r+n**ɛ**k/	[m**a**:rti:r-n**ɔ**k]	‘martyr’
(16c)	2 syllables	/d**o**ktrine:r+n**ɛ**k/	[d**o**ktrine:r-n**ɛ**k]	‘doctrinaire’
	2 syllables	/k**ɔ**libe:r+n**ɛ**k/	[k**ɔ**libe:r-n**ɛ**k]	‘caliber’
	2 syllables	/b**o**ri:te:k+n**ɛ**k/	[b**o**ri:te:k-n**ɔ**k]	‘envelope’

In addition, vowel harmony in Hungarian is sensitive to the height of the rightmost transparent vowel (Hayes and Londe 2006). Hungarian has four front unrounded vowels that contrast for height: /i/, /i:/, /e:/, and /ɛ/. Words with a high transparent vowel (e.g., [ɔpost**o**li-n**ɔ**k] and [m**a**:rti:r-n**ɔ**k]) are likelier to harmonize than words with a mid vowel (e.g., [g**ɔ**lle:r-n**ɔ**k] ‘collar’ but [f**ɔ**se:n-n**ɛ**k] ‘charcoal’), which are themselves likelier to harmonize than words with a low vowel (e.g., [komp**o**nɛns-n**ɛ**k] ‘component’ and [hɔmb**u**rgɛr-n**ɛ**k] ‘hamburger’).

Statistical data for distance-based decay are illustrated below in Table 4[1] and Figure 4.

Vowel harmony in Hungarian			
Transparent syllables	Faithful forms	Harmonized forms	Proportion of harm’d forms
$n = 0$	4.32	6284.68	0.99
$n = 1$	128.04	633.96	0.83
$n = 2$	60.79	17.20	0.22
$n = 3$	8.00	0.00	0.00

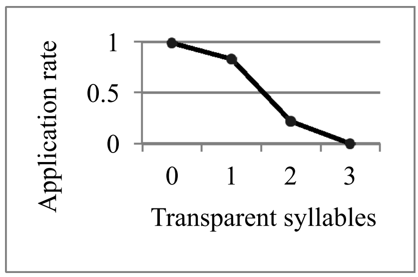

Table 4: *figures for distance-based decay in vowel harmony in Hungarian*
Figure 4: *graph of distance-based decay in vowel harmony in Hungarian*

Though sensitivity to transparent vowel height is not taken into account in the figures above, the model of distance-based vowel harmony will control for the effect.

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 the target decreases as the distance between them increases. We now turn to the task of accounting for the decay effect within 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), a variant of the constraint-based framework Optimality Theory (Prince and Smolensky 1993). Constraints are associated with real-valued weights that signify constraint strength, and outputs are associated with a nonzero probability, a function of the weighted sum of constraint violations. Suppose we have n constraints and are considering m candidate surface forms of a single underlying form. Let w_k be the weight of constraint

[1] The data on distance-based decay are presented in the form of lexical variation, but Hayes et al. 2009 make available a corpus—the results of a Google study on vowel harmony—that features both lexical and within-word variation. Associated with each stem in the corpus is the percentage of tokens of the dative form undergoing vowel harmony. Table 4 tabulates a token-weighted type frequency count of the corpus data: forms such as [ɔbl**ɔ**k-n**ɔ**k] ‘window’-DAT that do not exhibit within-word variation contribute 1 to the count of harmonized forms, while a stem such as /krɔpɛk/ ‘dude’, which takes -[nɔk] 80% of the time and -[nɛk] 20% of the time, contributes 0.80 and 0.20 respectively to the relevant type counts.

k and χ_{ik} be the number of times the ith surface form violates constraint k. The harmony H_i and the probability P_i of the ith surface form are defined in (17a) and (17b) respectively.

(17a) $H_i = \sum_{k=1}^{n} w_k * \chi_{ik}$ (17b) $P_i = e^{-H_i} / \sum_{j=1}^{m} e^{-H_j}$

One approach to accounting for distance-based decay is to posit a family of markedness constraints that penalize trigger-target pairs that occur at specific distances (Hansson 2001, Martin 2005, Hayes and Londe 2006). This approach consigns the learner to acquiring the decay effect—i.e., acquiring higher constraint weights for closer trigger-target pairs and lower ones for farther trigger-target pairs. The account dismisses as coincidental the systematic nature of distance-based decay, failing to rule out the existence of a learner who acquires a language with distance-based "anti-decay", for instance, in which application increases with distance, or a learner who acquires a language with greatest application at intermediate distances. Yet languages possessing such properties are unattested. An account that posits constraints that penalize trigger-target pairs occurring within a particular distance (similar in form to constraints proposed in de Lacy 2002) enforces higher application rates locally and lower rates nonlocally, but is still not restrictive enough: it fails to predict crosslinguistic consistency in the exponential relationship between the weight of markedness and distance (see Section 4). I therefore reject constraint-family approaches due to them being too powerful.

In my account of distance-based decay, I adopt and extend the proposal put forth by Kimper 2011, which introduces a factor that scales the weight of a single SPREAD constraint with increasing distance, yielding non-application of vowel harmony in Hungarian in distal contexts. Unlike a 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 the target and returns a scalar value. When the harmony of the form is calculated, the scalar value is multiplied by the weight of the markedness constraint regulating the long-distance cooccurrence restriction.

More precisely, let C_m be a markedness constraint with weight w_m regulating a long-distance cooccurrence restriction against two (not necessarily distinct) segments a and b. Define a function d that takes a nonnegative integer measure of distance x between the trigger and the target defined by C_m and returns a positive real value $d(x)$. A surface form containing one instance of the subsequence [a...b], where a and b are x distance apart, violates *a...b $d(x)$ times. Assuming that the only two constraints in the grammar are C_m and IDENT, the expression for the harmony of [a...b] is $H_i = w_m * d(x)$. The formula for the harmony of the unfaithful candidate satisfying C_m but violating IDENT remains the same: the weight of faithfulness.

To see an example of how the system works, consider the set of potential forms undergoing liquid dissimilation in Latin, in which the trigger and the target are separated by two transparent syllables. We can define constraints *l…l and IDENT([lat]) so that w_m = w(*l…l) = 10.97 and w_f = w(IDENT([lat])) = 4.53, and instantiate $d(x)$ to be the negative power function $d(x) = 1/x$, where the distance unit x is taken to be the number of transparent syllables plus one. Justification for using such a distance metric will be covered in the following section.

/…**l**σ²+a**l**is/	*l…l w_m = 10.97	IDENT([lat]) w = 4.53	Harmony	Predicted probability	Observed probability
[…**l**σ²-a:**l**is] (e.g., [**l**apid-a:**l**is])	$d(2) = 1/3$ $\approx$ 0.33 violations		$10.97 * 0.33$ $=$ $w_m * d(2)$	$\frac{e^{-10.97*0.33}}{(e^{-10.97*0.33} + e^{-4.53*1})}$ ≈ 0.70	0.69
[…**l**σ²-a:**r**is] (e.g., [**l**apid-a:**r**is])		1	$4.53 * 1$ $= w_f$	$\frac{e^{-4.53*1}}{(e^{-10.97*0.33} + e^{-4.53*1})}$ ≈ 0.30	0.31

Table 5: *tableau representation of liquid dissimilation in Latin with a decay function*

The tableau demonstrates that scaling the weight of markedness can yield predicted probabilities that accurately match observed application rates. The scaling is achieved by replacing integer violations of a cooccurrence constraint with the output of the decay function.

Note that the tableau above uses particular model parameters (i.e., particular values of the weight of markedness and the weight of faithfulness) and a certain decay function, a simple negative power function, to model the decay effect. In the following section, I discuss why a negative power function is desirable for modeling distance-based decay and explain how I determined the parameter values seen in Table 5.

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? As a preliminary study, heuristic distance-based markedness constraints were posited for purposes of exposing how the weight of markedness scales with distance. Data on each of the four languages, language-specific markedness constraints that penalized trigger-target pairs at different distances, and language-specific IDENT constraints were inputted into the MaxEnt Grammar Tool (Hayes and Wilson 2008, http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/). The output weights led to accurate model predictions. A single IDENT constraint was used in each grammar, and the weights of IDENT for the processes in Malagasy, Latin, English, and Hungarian were 8.57, 8.37, 8.46, and 13.83, respectively. Note that the constraint-based model for vowel harmony in Hungarian, as well as all models of the process conducted in this study, contained constraints to control for height sensitivity.

The heuristic distance-based markedness constraints reveal distance sensitivity of long-distance cooccurrence constraints to exhibit a broad pattern that holds across the four languages.

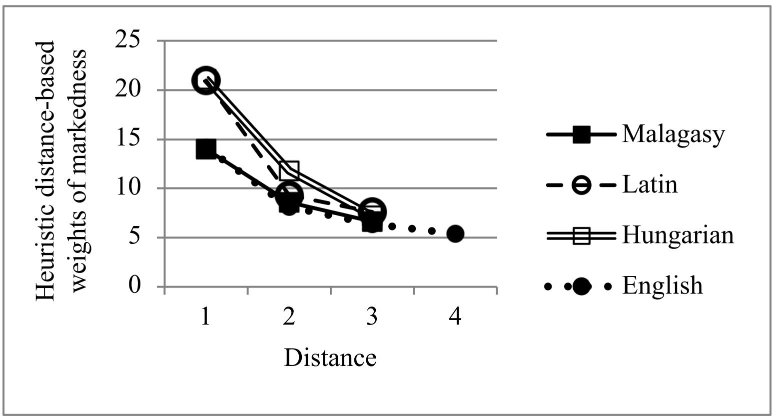

Figure 5: *graph of exponential decrease in the weight of markedness across the four languages*

For all four cases, local weights are high, and tend to zero with increasing distance. In other words, the weight of markedness decreases roughly exponentially with distance. A natural choice for $d(x)$ that would scale the weight of markedness in this way is the negative power function, $d(x) = 1/x^k$, where k is a positive real number. For now, k will be assumed to be a language-specific parameter; nonetheless, I will argue that a single, universal value of k can be used to accurately model the decay effect in the four surveyed languages.

I used the `glmer` function of R's `lme4` package (Bates and Maechler 2011) to fit logistic regression models to each of the four processes, and then used the `Anova` function from the `car` package (Fox 2009) to run likelihood ratio tests for purposes of model comparison. For a given linear model, the likelihood ratio test compared the full model to those that omit each variable in turn, and returned p-values that determined whether the full model was more predictive than each of the smaller models, thereby determining whether each of the variables influenced application rate significantly. I also calculated the A(kaike) I(nformation) C(riterion) of each model, which scored it based on accuracy (i.e., fit to the data) and parsimony (i.e., number of parameters). In cases where one or more variables were not significant as determined by the likelihood ratio test, I eliminated the variable with the highest p-value, reran the likelihood ratio test, and recalculated the AIC. I iterated this process until the remaining variables all had p-values below $p = 0.1$. The AIC of the final model was always lower than those of the predecessor models.

The primary finding is that the number of transparent syllables, rather than the number of transparent segments, is the superior predictor of the likelihood of process application. For each language, the best model with only syllable count had a lower AIC than the best model with only segment count (compare 769.40 and 891.81 in Malagasy, 143.44 and 143.53 in Latin, 278.70 and 305.57 in English, and 311.38 and 354.05 in Hungarian). Moreover, the AIC preferred models that contained only syllable count to those that contained both syllable count and segment count. These results support the findings of Martin 2005, which shows that syllable count is a superior predictor to segment count in sibilant harmony in Navajo. In light of these facts, the basic unit of distance x of the decay function $d(x)$ will depend on syllable count.

Since there are no instances where the trigger and the target are in the same location, the decay function $d(x)$ never takes as input distance $x = 0$, and so the smallest distance input is $x = 1$. For the processes in Malagasy, Latin, and Hungarian, distance is easy to calculate: x is the number of transparent syllables plus one (e.g., in Malagasy, [bab**u**.-**u**] incurs a penalty of $d(1)$, while [r**u**.va.-**u**] incurs a penalty of $d(2)$). In these three processes, the target is in onset or nucleus position, and the trigger and the target are local when they are in adjacent syllables. In English, the target of liquid dissimilation is in coda position, producing a complication: the trigger and the target are local to one another when they are in the same syllable, as substantiated by categorical forms such as *solar* and *condylar*, while adjacent syllables constitute a nonlocal environment, as revealed by lexical variation in forms such as *lunar* but also *filial* and even *algal*. Syllable count alone is an inadequate metric because the same count—zero transparent syllables—classifies both local and nonlocal environments. In my model, this discrepancy is resolved as follows: in all cases, local violations incur a penalty of $d(1)$ (e.g., **solal* in English, and *[bab**u**.-**u**] in Malagasy), with the distance count increasing by the syllable. In English, forms with trigger-target pairs in adjacent syllables incur a penalty of $d(2)$, while those with one transparent syllable incur a penalty of $d(3)$, and so on. In the other cases, counting is the same as stated above: distance increases by the syllable.

Values for three model parameters must be found: the weight of markedness w_m, the weight of faithfulness w_f, and the decay parameter k of the decay function $d(x) = 1/x^k$. I used Microsoft Excel's Solver, which has nonlinear curve-fitting capabilities (see below), to find suitable parameter values for the expression for the probability of process application as a function of distance:

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

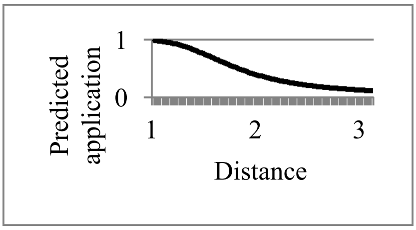

Figure 6: *graph of P*(x) *with* w_m = 9, w_f = 5, *and k* = 1

The probability of faithfulness is simply 1 - $P(x)$. The expression shown above is the same as the expression for the probability process application in Maximum Entropy Harmonic Grammar, except that the weight of markedness w_m is now multiplied by $d(x)$, scaling its value to grow smaller with increasing distance. $P(x)$ is shaped like an asymmetrical sigmoid curve, mimicking the data distributions seen in Section 2.

The negative log likelihood of the dataset was minimized to produce parameters that yield greatest model accuracy. The surface defined by the model-predicted probabilities of the dataset is convex over the space of constraint weights (Della Pietra, Della Pietra, and Lafferty 1997), making it possible for the Solver to find optimal values for w_m and w_f given a fixed value of k without "getting stuck" in a local optimum. The Solver determines parameter values using Newton's Method, an iterative descent algorithm, which maximizes the model-predicted probability of the dataset and minimizes the model-predicted probability of unobserved forms.

Table 6, shown below, contains the results of letting the Solver optimize over the constraint weights as well as k. Average error is here defined as the absolute difference between observed application rate and model-predicted likelihood of application, averaged over the violation profiles.

Rounding dissimilation in Malagasy			Liquid dissimilation in Latin		
	Language-specific k	Universal k		Language-specific k	Universal k
w_m:	12.12	10.24	w_m:	16.72	9.00
w_f:	6.61	4.66	w_f:	0.40	2.63
k :	0.9	**1.1**	k :	3.2	**1.1**
Avg. err.:	0.00	0.01	Avg. err.:	0.00	0.02
Vowel harmony in Hungarian			**Liquid dissimilation in English**		
	Language-specific k	Universal k		Language-specific k	Universal k
w_m:	44.36	27.04	w_m:	9.39	10.90
w_f:	37.08	19.76	w_f:	3.73	5.28
k :	0.3	**1.1**	k :	1.4	**1.1**
Avg. err.:	0.06	0.06	Avg. err.:	0.00	0.01

Table 6: *distance-based decay modeling results*[2] *using a negative power function with language-specific* w_m, w_f, *and either variable or language-invariant k*

Table 6 shows that fitting the decay function to each language by letting *k* vary on a language-by-language basis yields highly accurate model predictions.[3]

Nonetheless, a question persists: are three language-specific parameters necessary to ensure model accuracy? The weight of markedness and the weight of faithfulness are acquired by the learner based on the forms they are exposed to, and are therefore expected to be language-specific. On the other hand, nothing has been said about whether *k* should be expected to vary; i.e., whether the learner needs to acquire a particular decay function—a particular value of *k*—depending, for instance, on the decay rates they are exposed to. One can imagine that differences in decay rate lead to strikingly different optimal values of *k*. On the contrary, Table 6 shows that we can fix *k*, thereby holding $d(x)$ constant across languages, and still accurately model language-specific decay rates by varying only the constraint weights.[4] Based on these results, I conclude that distance-based decay can be modeled effectively with a fixed scale, an invariant decay function.

Zymet 2014 demonstrates that the negative power function performs as well as or better than other candidate models. Consider another model that lacks k altogether, one with a simple down-sloping linear function $d(x) = -x$; i.e., $P(x) = \frac{1}{1+e^{-(w_m*(-x)+w_f)}}$, a more commonplace sigmoid curve in which negative distance is equal to the number of markedness violations. The typical parameters of the linear function are here absorbed by the constraint weights: the slope is absorbed by the weight of markedness while the intercept is absorbed by the weight of faithfulness. Substituting the negative power function with a linear function results in at least tripled average error in all cases except for vowel harmony in Hungarian, in which the two models perform equally (presumably because both models are allowed to fit height-based constraints). Summed AICs over all four languages reveal that

---

[2] Logistic regressions show that the position of the trigger within the syllable affects application rates in English and Latin significantly or nearly significantly. I decided to model only the forms in either language in which the trigger was in onset-initial position, as there were few data on the other pertinent kinds of forms.

[3] Note that the higher average error for vowel harmony in Hungarian is due to the model's inability to predict the high rate of application in BTe: forms (where B is the trigger and T is a transparent vowel). Hayes and Londe 2006 show that the high application rate is an abnormality of the lexicon: native speakers show lower application rates in a wug test (Berko 1958).

[4] To ensure that the Solver was not "getting stuck" in local optima when it was optimizing over constraint weights *and* the decay parameter *k*, *k-basins* were formed. For each language, *k* was fixed to an arbitrary value, and the Solver was set to minimize the negative log likelihood by varying only the constraint weights. The process was iterated for different fixed values of *k* between 0.2 and 8.0 and in increments of 0.1, leading to the formation of the basins. For liquid dissimilation in Latin and vowel harmony in Hungarian, the basins were flat, demonstrating that a wide range of values of *k* lead to minimal modeling error, and explaining why the models of those languages can tolerate $k = 1$ without losing much accuracy, despite distant optimal values of *k*. In contrast, rounding dissimilation in Malagasy and liquid dissimilation in Latin are comparatively delicate: few values of *k* lie at the bottom of the two basins, with optimal values at $k = 0.9$ and $k = 1.4$, respectively. These facts suggest there are universal values of *k* that lead to accurate model predictions across the four languages, but the range of values is restricted, with the optimal value located around $k = 1.1$.

the reduction in parameter count in adopting a linear function is not worth the decrease in modeling accuracy: the summed AIC of the negative power function (1357.8) is roughly fifty points lower than that of the linear function (1408.0). This demonstrates that the negative power function is superior to the linear function in modeling distance-based decay.

On the other hand, the same cannot be said about the inverse-exponential decay function $d(x) = 1/k^x$, the scaling factor proposed in Kimper 2011 for modeling vowel harmony in Hungarian. The inverse-exponential function is so similar in shape to the negative power function $d(x) = 1/x^k$ for lower values of x that the two models perform roughly equally in all cases except for liquid dissimilation in English, in which the negative power function performs slightly better.

5. Conclusion

In summary, distance-based decay results from the weight of AGREE/DISAGREE decreasing exponentially with increasing distance, as measured in syllables. While the constraint-family approach to distance-based decay fails to predict this relationship exclusively, the scale-based approach introduced in Kimper 2011 can be used to model the decay effect. I adopt a negative power function to scale the weight of AGREE/DISAGREE, demonstrating that it need not be fit to each language in order to capture language-specific differences in decay rates present in the data; rather, these differences can be accounted for with an invariant decay function by varying constraint weights alone. Positing a negative power function adds one additional parameter, the decay constant k whose optimal setting is k = 1.1. Getting rid of the decay constant by substituting the negative power function with a linear function results in a significant decrease in modeling accuracy.

References

Bates, Douglas & Maechler, Martin. 2011. Package ‘lme4’. R.

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

de La Beaujardière, Jean-Marie. 2004. Malagasy Dictionary and Encyclopedia of Madagascar.

de Lacy, Paul. 2002. *The formal expression of markedness*. Doctoral dissertation, University of Massachusetts, Amherst. Amherst, MA: GLSA Publications. ROA 542.

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

Fox, John. 2009. Package “car”. R.

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

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

Hansson, Gunnar. 2001. Theoretical and typological issues in consonant harmony. Ph.D. dissertation, UC Berkeley.

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

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

Hayes, Bruce, Zuraw, Kie, Siptar, Peter & Londe, Zsuzsa. 2009. Natural and Unnatural Constraints in Hungarian Vowel Harmony. *Language* 85: 822-863.

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

Prince, Alan & Smolensky, Paul. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. Rutgers Center for Cognitive Science, Piscataway, New Jersey.

Smolensky, Paul. 1986. Information processing in dynamical systems: Foundations of harmony theory. In David E. Rumelhart, John 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.

Zymet, Jesse. 2014. Distance-based decay in long-distance phonological processes: a probabilistic model for Malagasy, Latin, English, and Hungarian. Master’s thesis, UCLA.

Proceedings of the 32nd West Coast Conference on Formal Linguistics

edited by Ulrike Steindl, Thomas Borer, Huilin Fang, Alfredo García Pardo, Peter Guekguezian, Brian Hsu, Charlie O'Hara, and Iris Chuoying Ouyang

Cascadilla Proceedings Project Somerville, MA 2015

Copyright information

ISBN 978-1-57473-466-9 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Zymet, Jesse. 2015. Distance-Based Decay in Long-Distance Phonological Processes. In *Proceedings of the 32nd West Coast Conference on Formal Linguistics*, ed. Ulrike Steindl et al., 72-81. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #3158.