Distance-based decay in long-distance phonological processes

Jesse Zymet
UCLA

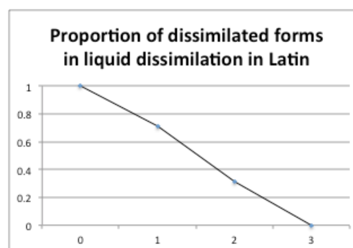
I. Overview

There are long-distance phonological processes throughout the world's languages that are variable within and across words, and that decay in how often they apply as the number of transparent syllables between trigger and target increases (a phenomenon that I call *distance-based decay*). The phenomenon has been observed in several languages, including Arabic (Frisch, Pierrehumbert, and Broe 2004), Sanskrit (Walker and Mpiranya 2006), and Hungarian (Hayes and Londe 2006). Distance-based decay poses a challenge for approaches to assimilation and dissimilation in which distance does not play a role, such as spreading, OCP with autosegmental representations, and ABC with no counting of distance. This poster presents a unified account of distance-based decay, drawing from cross-linguistic data.

II. Data

The data distribution given below, for example, reflects distance-based decay in liquid dissimilation in Latin (n is the number of transparent syllables between trigger and target). This analysis accounts for the decay effect present in liquid dissimilation in Latin, vowel dissimilation in Malagasy, vowel harmony in Hungarian, and retroflex harmony in Sanskrit. Data were gathered from a Malagasy language database at www.malagasyword.org, Latin language databases from the Perseus Digital Library and Linguae Latinae, a Hungarian language database at http://linguistics.ucla.edu/people/hayes/HungarianVH/Index.htm, and a Sanskrit language database at the Digital Corpus of Sanskrit.

Liquid dissimilation in Latin: /…lσ^n+a:lis/ → […lσ^n-a:ris]			
Transparent syllables n	Faithful forms	Dissimilated forms	% of dissim'd forms
0	0	131	100%
1	20	49	71.0%
2	29	13	31.0%
3	4	0	0%



III. Account of distance-based decay

The analysis of distance-based decay is grounded in the maximum entropy framework (cf. Hayes and Wilson 2008). A distance function $d(x)$ (see next section) takes as its argument a measure of distance between trigger and target. Violations of a dissimilation or assimilation constraint are then multiplied by $d(x)$. In Latin, for example, we can define constraints *l…l and IDENT(lateral) with weights w(*l…l) = 9.27 and w(IDENT(lateral)) = 8.37, and $d(x) = 1/x^{0.29}$, where x is the number of transparent syllables.

/…lσ^2+a:lis/ e.g., /lamin+a:lis/	*l…l w = 9.27	IDENT(lateral) w = 8.37	Harmony	Predicted probability	Observed probability
[…lσ^2a:lis]	* → $1/(2^{0.29})$ = 0.82 violations		$e^{-9.27\cdot 0.82}$	$e^{-9.27\cdot 0.82}/(e^{-9.27\cdot 0.82} + e^{-8.37\cdot 1}) \approx 0.69$	.69
[…lσ^2a:ris]		*	$e^{-8.37\cdot 1}$	$e^{-8.37\cdot 1}/(e^{-9.27\cdot 0.82} + e^{-8.37\cdot 1}) \approx 0.31$	.31

The grammar predicts that [...lσ^2-aːlis] surfaces 69% and [...lσ^2-aːris] 31% of the time, whichis quite accurate. This analysis is an extension of that of Kimper 2011a, in which distance-based decay in Hungarian vowel harmony was treated with an exponential scaling factor.

IV. Determining the properties of the distance function

Logistic regression revealed that the number of transparent syllables—but not the number of transparent segments—significantly influences process application. Hence, argument x of the distance function $d(x)$ is taken to be the number of intervening transparent syllables.

With few exceptions, the processes examined apply categorically in strictly local environments, i.e., when there are no transparent syllables intervening. As the number of transparent syllables increases, the likelihood that the process applies tends to zero. Therefore, $d(x)$ is taken to be represented as a monotonically decreasing inverse exponential function, $d(x) = 1/x^k$, where k is a positive, real-valued parameter.

Values for two parameters are needed for the above cases: the difference between markedness and identity constraints and the exponent k of the distance function. As a heuristic, a grammar was first fitted with separate constraints for each distance point. Latin, for example, used *l0l (which penalizes forms with zero syllables intervening between [l]s), *l1l, *l2l, and *l3l. The fitted weights were w(*l0l) = 21.02, w(*l1l) = 9.27, w(*l2l) = 7.57, w(*l3l) = 0, and IDENT(lateral) = 8.57. (Since few words had three transparent syllables, w(*l3l) was then excluded.) Because $1/x^k$ is 1 when $x = 1$, the weight of the general constraint *l...l is w(*l1l)= 9.27. Setting $9.27 \cdot d(2) = 9.27/2^k = w$(*l2l) implies $k = \lg(9.27/7.57) = 0.292$. The distance function is thus $d(x) = 1/x^{0.29}$. $d(x)$ came out to be $1/x^{0.37}$ for vowel dissimilation in Malagasy, $1/x^{0.63}$ for vowel harmony in Hungarian, and $1/x^{0.91}$ for retroflex harmony in Sanskrit. The models—now augmented with a distance function—accurately predict the observed distributions of data provided above with minimal error. That the exponents of the distance functions all lie within a small range may suggest that $d(x)$ is universal, taking as its exponent a value somewhere around 0.5.

V. References

Frisch, S., Pierrehumbert, J., & Broe, M. B. 2004. Similarity Avoidance and the OCP. *Natural Language & Linguistic Theory* 22(1): 179-228.

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.

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

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.