Consonant harmony, disharmony, memory and time scales

Adamantios Gafos
University of Potsdam
gafos@uni-potsdam.de

Abstract

I argue that properties of memory, a so far largely neglected source of explanation in phonological patterns, offer a grounding for a number of not well understood traits of long distance consonantal restrictions (both dissimilatory and assimilatory).

1 Introduction

In the great variety of long distance consonantal restrictions met across languages, production and perception factors have been implicated. Here, I consider the neglected factor of memory. I argue that properties of memory, when combined with the other two factors, enable new answers to unresolved questions about long distance consonantal restrictions. Section 2 considers short term memory in dissimilatory restrictions. Long term memory and assimilatory restrictions are taken up in Section 3. Section 4 concludes.

2 Dissimilation

Dissimilations involving laryngeal features are a primary example of long distance co-occurrence restrictions in consonants (MacEachern, 1999; Mackenzie, 2009; Gallagher, 2010). A number of languages ban the presence of repeated ejectives, aspirated or implosive consonants within roots. For example, in Shuswap, roots cannot contain two ejectives, e.g., using C' to denote ejectivization, /k^{w}'alt/ 'to stagger', /qet'/ 'to hoist', /k^{w}up/ 'to push', /qmut/ 'hat', but */k^{w}'alt'/, */q'et'/ and so on are unattested (Kuipers, 1974). Gallagher (2010) proposes a contrast neutralization analysis of such restrictions. The basic generalization captured in that analysis is that roots with two ejectives, as in */k'ap'i/, are unattested because they may not contrast with roots with only one ejective, as in /k'api/ or /kap'i/. An example from this analysis is in tableau (1). The four-way contrast in the distribution of ejectivization in a hypothetical input set {/k'ap'i/, /k'api/, /kap'i/, /kapi/} is neutralized to a three-way contrast. This is because the high ranking *1vs2 constraint bans candidate sets exhibiting a contrast between one versus two ejectives as in set (1i), and neutralization to one ejective as in (1ii) is preferred over neutralization to two ejectives as in (1iii), because the faithfulness constraint to the ejectivization feature FAITH(+CG) outranks the *0vs1 constraint which bans candidate sets exhibiting a contrast between zero versus one ejective ('+CG' stands for constricted glottis).

{/k'ap'i/, /k'api/, /kap'i/, /kapi/}	*1vs2	FAITH(+CG)	*0vs1
i. [k'ap'i, k'api, kap'i, kapi]	**		****
ii. ☞ [k'api, kap'i, kapi]		*	**
iii. [k'ap'i, kapi]		**!	

Table 1. *1vs2 >> FAITH(+CG) >> *0vs1

In support of the ranking *1vs2 >> *0vs1 that plays out in this loss of contrast analysis, Gallagher conducted discrimination experiments where listeners heard pairs of CVCV stimuli and gave a same versus different judgment. The stimuli were disyllabic 'words' containing zero, one, or two ejectives and were paired in three conditions: condition 1vs2, that is, one versus two instances of the ejective as in [kap'i]-[k'ap'i], 0vs1 as in [kapi]-[k'api], and 0vs2 as in [kapi]-[k'ap'i] (and so on for aspirated Cs). Percent of judgment

correct (of the same versus different judgment) was the dependent variable. The main result was that, for both ejectivization and aspiration, the 1vs2 condition was 'more difficult' (i.e., correct responses to same versus different were less accurate) than the 0vs1 condition, which in turn was more difficult than the 0vs2 condition. Gallagher (2010), whose primary concern is synchronic analyses of dissimilation patterns as they play out in their language-particular details, leaves open the source of these discrimination asymmetries ('A major question … is the source of the perceptual asymmetries', ibid.: 106).

Consider what the task in the experiments above involves. A chain of auditory processing, encoding of the stimuli in short term (or working) memory, along with a discrimination decision is implicated. That is, comparison of the stimuli is mediated by representations of their forms in short term memory (Baddeley, 1986). This invites consideration of the extent to which the results may derive from properties of short term memory.

A well-replicated finding about short term memory is that features of items in a list that match features of other items tend to be omitted, a so-called interference effect. The mechanism which gives rise to this effect is dubbed feature overwrite in models of short term memory (Nairne, 1990; Nairne, 2001): in a list of elements (e.g., syllables), a feature F of the memory trace of an element is omitted (with some probability; more on this below) if it matches the feature of a following, adjacent element (Nairne, 1990: 252).

Consider how interference plays out in the different conditions of Gallagher's experiments. In the 0vs1 condition, the stimuli pairs are [kapi-k'api], [kapi-kap'i], [k'api-kapi], and [kap'i-kapi]. No pair is subject to feature overwrite, as the ejective is not repeated. Discriminability of the two words in each pair is function of their similarity. The similarity $s(i,j)$ between words i,j is related to their featural distance $d(i,j)$ by $s(i,j) = e^{-d(i,j)}$ (Shepard, 1987); the larger the distance, the less the similarity. Distance is a function of the number of featural mismatches between the two forms. More precisely, distance is given by $d(i,j) = \frac{\sum b_k M_k}{N}$ where b_k is the weight of the particular feature (some features may contribute more to distance than others), M_k is a counter of the featural mismatches, and N is the number of features. In our case, N is 1 as all pairs in this task (e.g., [kapi-k'api]) are identical except for [+CG] (thus, the weight parameter b_k can also be set to 1 since effectively there is no other feature, given the stimuli in this task). This means that the distance between the two stimuli in any pair of stimuli of the task is given simply by the number of mismatches with respect to the [+CG] feature. Thus, d([kapi], [k'api]) $= 1$ and s([kapi], [k'api]) $= e^{-d} = e^{-1} = 0.37$. The same holds for the other pairs, [kapi-kap'i], [k'api-kapi], [kap'i-kapi]. For each pair, the distance is 1 and therefore the similarity is $e^{-1} = 0.37$. Thus, the average similarity between the two stimuli across all four stimuli pairs in this 0vs1 condition is 0.37.

In the 1vs2 condition, the stimuli are [k'ap'i-k'api], [kap'i-k'ap'i], [k'ap'i-kap'i], [k'api-k'ap'i]. These become after overwrite [k'ap'i-k'api] → [kap'i-k'api], with a resulting similarity s([kap'i], [k'api]) $= e^{-2} = 0.13$ ($d = 2$, s $= e^{-d} = 0.13$), [kap'i-k'ap'i] → [kap'i-kap'i] ($d = 0$, s $= e^{-d} = 1$), [k'ap'i-kap'i] → [kap'i-kap'i] ($d = 0$, s $= e^{-d} = 1$), and finally [k'api-k'ap'i] → [k'api-kap'i] ($d = 2$, s $= e^{-d} = 0.13$). Across all four stimuli pairs, the average similarity between the two stimuli within pairs is $\frac{1}{4}[0.13 + 1 + 1 + 0.13] = 0.57$ (higher than for the 0vs1 condition).

These illustrations assume that interference applies in every experimental trial. To prove a general result, I assume that overwrite applies with probability p and show how to derive the discriminability asymmetries in the above experiments for any p. I begin by expressing the expected similarity for each of the above pairs as a function of this probability p of overwrite ($1 - p$, no overwrite). In the 1vs2 condition, the expected distance for [k'ap'i-k'api] → [kap'i-k'api] is $1(1-p) + 2p = 1 + p$ (read: 1, the distance when overwrite does not apply, multiplied by the probability of no overwrite $(1-p)$, plus 2, the distance when overwrite applies, multiplied by the probability of overwrite p), $1(1-p) + 0p = 1 - p$ for [kap'i-k'ap'i] → [kap'i-kap'i], $1(1-p) + 0p = 1 - p$ for [k'ap'i-kap'i] → [kap'i-kap'i], and $1(1-p) + 2p = 1 + p$ for [k'api-k'ap'i] → [k'api-kap'i]. Hence, the average similarity across the four pairs is $\frac{1}{4}[2e^{-p-1} + 2e^{p-1}] = \frac{1}{2}e^{-1}(e^{-p} + e^{p})$. This expression is lower bounded (at $p = 0$) by the similarity of the

0vs1 condition, $e^{-1} = 0.37$ which does not vary as a function of p. Figure 1 plots the similarities across these conditions. The 1vs2 condition has higher similarity than the 0vs1 condition.

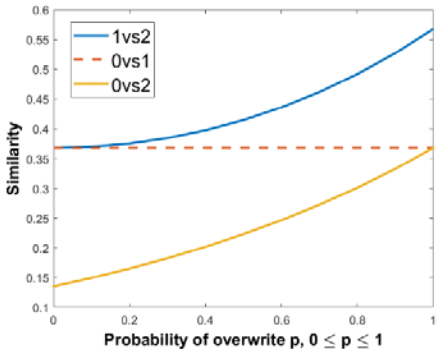

Figure 1. Similarities across the three conditions, 1vs2, 0vs1, 0vs2, as a function of probability p of overwrite.

For the 0vs2 condition, the stimuli pairs are two: [kapi-k'ap'i], [k'ap'i-kapi]. Each has a distance of 2 when no overwrite applies. With overwrite, the pairs turn to [kapi-kap'i] and [kap'i-kapi], each with distance of 1. Given that p is the probability of overwrite, the expected distance for each pair is $2(1-p)+1p = 2-p$. Hence, the average similarity across the two pairs is $s = e^{p-2}$ which is less than or equal to e^{-1} of the 0vs1 condition; the maximum in similarity in this 0vs2 condition is attained for $p = 1$.

Overall, we see that the 0vs2 condition has the lowest similarity (easiest to discriminate), followed by the 0vs1 condition which in turn has lower similarity than the 1vs2 condition. This is in full accordance with the experimental results (Gallagher, 2010:95, Fig. 2; 101, Fig. 6). It is clear that the same account applies to discrimination results on aspiration (cf. Grassmann's law).

In sum, the discrimination asymmetries in the above experiments can be seen to follow from mechanisms of memory (see Appendix A for variants of the account). Evaluation of the claim that these asymmetries constitute a basis for cross-linguistic dissimilation patterns (Gallagher, 2010) extends beyond the present scope, but the following points are worth making in this regard. Ultimately, a full account of long distance dissimilatory patterns must rely on additional memory principles. Specifically, in addition to feature omission, also the positional encoding of features is subject to interference. Features can swap positions: /…C_i^F…C_j…/ (read: feature F is associated to consonant in position i) changes to /…C_i…C_j^F…/ and vice versa (Nairne, 1991; Estes, 1972; Lee and Estes, 1977). Features are mobile, in other words, under certain conditions. This offers a basis for another systemic constraint, additional to those in tableau (1), penalizing contrasts with respect to the position of a feature, that is, *{C^F-C, C-C^F} (Gallagher, 2010: 133). Note the parallelism between swap and coarticulation. Swap relocates a feature. Coarticulation extends the presence of a feature (on how this leads to similarity-conditioned identity, see 3.3). Both swap and coarticulation potentially reduce or eradicate contrasts, in different channels, memories versus vocal tracts. In swap, /…C_i^F…C_j…/ changes to /…C_i…C_j^F…/ and vice versa (transpositions are bidirectional). Coarticulation eliminates the contrast between /tVʈ/ (the second C is retroflex) and /ʈVʈ/ (both Cs are retroflex). Now, not all features are equally amenable to memory interference (swap) and coarticulation. Ejectivization plays out in dissimilatory as well as assimilatory patterns, but not so for retroflexion; there are no (undisputed) cases of dissimilation in retroflexion (Arsenault, 2012). This difference relates to the nature of the involved features. The phonetic cues to ejectivization are separable: these cues, a long VOT and an intense burst amplitude, are the same regardless of the other features of their segmental hosts. In contrast, retroflexion is a so-called integral feature (Garner, 1974). Its phonetic cues, as I review in 3.3, depend on other features of the segmental host. Thus, in contrast to ejectivization, which can swap position in /kap'i/, retroflexion in, say, /paʈi/ cannot swap as the cues to retroflexion on /p/ are different from those on coronals; in fact, retroflexion has no acoustic consequences on labials (this is the less interesting case of what it means for retroflexion to be an integral feature; see 3.3 for the more interesting case). The point here is that interference mechanisms do not apply to retroflexion. Hence, if dissimilation has a basis in memory interference mechanisms, the absence of dissimilation for retroflexion follows.

Finally, consider: who or what system is doing the optimization in tableau (1)? The computation shown therein is meant as a fragment of a grammar internalized in an idealized speaker-listener (Chomsky, 1980: 220) or a fragment of I-

Language (Chomsky, 1986: 22). But it employs systemic constraints on contrast, as in *1vs2, whereas I have argued that the forces applying within the individual, that is, the short term memory mechanisms, are unconcerned with any systemic forces on contrast. These mechanisms just do what they do. But their effects give rise to contrast-related pressures at the systemic level. If optimizations along the lines of tableau (1) tell us something useful about languages, as I believe they do, this is the business of another (not individual but) supra-individual system. I leave discussion of this issue to future work.

3 Identity

The broad class of long distance consonantal identity phenomena (Archangeli and Pulleyblank, 2007) splits into three subclasses: total C identity across vowels along with the related subspecies of so-called across-the-board effects (3.1), feature identity due to active spreading (3.2), and static feature identity holding over lexica (3.3). It is in the latter case where consideration of memory, along with perception and production factors, illuminates certain heretofore not well understood properties of these static feature identity patterns.

3.1 Total identity and ATB effects (copying)

A well-known case of total identity is found in Arabic, [ħabab-tu], [malil-tu] and so on (‘to love’, ‘be weary’, 1.Sg.Perfect) where consonants are doubled stem-finally. In Chomsky and Halle (1968), the grammar mechanism underlying all assimilation (including harmony) was feature change. Later, assimilation and harmony were reanalyzed as spreading, the extension of a single autosegment or assimilating feature from trigger to target (Clements, 1976; van der Hulst, 1985). As autosegmentalism gained traction, a range of other phenomena were analyzed with the same mechanism. This is how the total identity pattern in Arabic came to be treated as spreading, in McCarthy (1979, 1981).

On the basis of parsimony considerations (due to special language-particular representational assumptions required to apply spreading in these identity cases) and typological properties, Gafos (1996a, 1996b[1999], 1998, 2003, 2018) argued that such cases of C identity across vowels in Arabic, Chaha, Temiar, Sierra Miwok and other languages are cases of copying, not spreading. In this reanalysis, the formal mechanism effecting the copying (leading to identity) is the Optimality Theoretic notion of correspondence (McCarthy and Prince, 1995). When two segments stand in correspondence, they are required to agree in their features. An agreement imperative over segments standing in correspondence is then another way to achieve feature identity. In contrast to spreading, where there is only a single feature whose domain extends to encompass the agreeing segments, in correspondence-based agreement each segment has its own instance of the agreeing feature; hence copying (in parallel to spreading) is a convenient shorthand for the formal grammar mechanism in the so-effected identity patterns. ‘Agreement’ is sometimes used to refer to the same mechanism but the term also serves as a decriptive label for identity patterns some of which arguably involve spreading as the formal mechanism. This can lead to confusion (see 3.2).

The same analysis, not spreading but copying, was argued in Gafos (1998) to extend to a class of what are referred to as ‘across the board’ (ATB) effects which target specific features. In the classic ATB pattern, a duplicated consonant in a root must echo any featural modification the other consonant undergoes in its local context. For example, certain Chaha morphological categories are expressed by labialization of the rightmost labializable consonant (velar or labial), e.g., from the verbal impersonal, /dänäg/ → [dänäg^{w}] ‘hit’ and /mäsär/ → [m^{w}äsär] ‘seem’. When the verb ends in two identical consonants, labialization appears in both instances: /säkäk/ → [säk^{w}äk^{w}] ‘plant in ground’, /gämäm/ → [gäm^{w}äm^{w}] ‘chip the rim’. Furthermore, when a root is duplicated, labialization appears again on both instances of the (labializable) consonant, e.g., /sexäsäx/ → [sex^{w}äsäx^{w}] ‘shell by grinding’. Previous analyses (McCarthy, 1983) posited two mechanisms, non-local spreading in [säk^{w}äk^{w}] but copying in [sex^{w}äsäx^{w}]. Gafos (1998) argued that both should be accounted for with a single mechanism, copying, expressed formally via correspondence.

3.2 Feature identity due to spreading

In contrast to whole consonant copying and ATB effects, Gafos (1996b[1999]) argued that certain other cases of long distance featural identity should be analyzed using strictly local spreading.

Strictly local means that, in contrast to alleged cases of spreading where consonants were thought to spread from C-to-C skipping the vowel (3.1), C-to-C spreading should not be possible 'except in the case of a consonantal gestural parameter which is able to propagate through the vowel and thus affect the consonant on the other side of the vowel' (Gafos, 1996b[1999]: 176). From the perspective of locality considerations, spreading abides to strict locality; copying does not, for non-unique reasons: multiple mechanisms sow the seeds for the potential generation of long distance (non-)identity: reduplication, planning (see 3.3), and, as per the new proposal of this paper, storage-based effects (see 2, 3.3). Coarticulation, one such mechanism, takes place in vocal tracts and is local in space-time. The other mechanisms (e.g., reduplication) are not (necessarily) local. For example, as argued in 2 and will be further argued in 3.3, properties of memory can give rise to non-local effects.

Strict locality of spreading only concerns spreading. Strictly local spreading implies, for example, that a C cannot spread across a V, e.g., /kap/ → /pap/ is impossible. The phonetic basis of spreading is coarticulation and spreading of /p/ in this example would require that /p/ propagates through the intervening V, eradicating the vowel. But other consonantal properties, such as mid-sagittal or cross-sectional postures of the tip-blade, can conceivably spread through a vowel.

I illustrate this for apicality-laminality. Figure 2 depicts mid-sagittal tongue shapes at three time points during the words [kʌs] (left) and [kʌlps] (right). The speaker is facing to the right. Within each panel, the top trace is the speaker's palate. Below the palate, appear the positions of four sensors placed on the tongue tip, blade, medio-dorsum, and dorsum. The three time points at which the tongue shape is shown are, from top to bottom, the onset of modal voicing of the vowel, the midpoint during the vowel, and the acoustic offset of the vowel. The profiles on the left of Figure 2 should be compared to those on the right, which illustrates coarticulation in global tongue shape originating from /l/ in [kʌlps]. It can be seen that at vowel onset (top), the tip-blade has already assumed a different posture (from that in [kʌs]) which propagates through [ʌ] (middle) in anticipation of the apical posture (tip is up relative to the blade) for /l/ of this speaker (bottom).

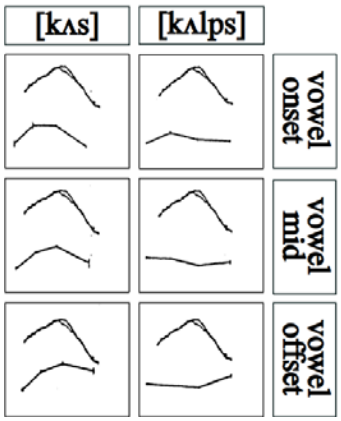

Figure 2. Tongue shapes in [kʌs] (left) versus [kʌlps] (right) at three time points, vowel onset (top), midpoint (middle), and vowel offset (bottom).

Several authors have argued that phonologized cases of C-to-C coarticulation through vowels are attested in various languages. Whitney (1889: (§189a), Flemming (1995), Steriade (1995ab), Gafos (1996b[1999]), Wiltshire and Goldstein (1997), Ní Chiosáin and Padgett (2001), Hamann (2003), Whalen et al. (2011), and Whalen and Tiede (2020) offer theoretical discussion and or phonetic plausibility arguments. A particularly relevant case is found in Walker et al. (2008)'s study on Kinyarwanda. The language shows a retroflexion harmony where stem /s, z/ change to their retroflex versions before suffix /-iiʂ/, e.g., /ku-sooz-a/ → [gusooza] 'finish', /ku-sooz-iiʂ-a/ → [guʂoozʐiiʂa] 'cause to finish'. Intervening segments were thought to be transparent but Walker et al.'s results indicate that tip-blade posture during such segments (/m/, /k/ and so on) is not different from that of the trigger fricative. Walker et al. (2008) conclude that this harmony case supports a strictly local spreading analysis.

There is potential for confusion here. Strictly local spreading has been interpreted to predict that 'non-coronal consonant harmony should not exist' (Hansson, 2001[2010]: 3). The confusion arises because 'harmony' refers to different phenomena for different authors. For Gafos (1996b[1999]), as per autosegmental ideas, the mechanism of assimilation and harmony is spreading, and thus 'consonant harmony' refers to those phenomena where spreading seems viable and testable by asking whether the spreading parameter goes through intervening segments. In that sense of harmony, phenomena involving labials as in /kap/ → [pap] or /bad/ → [bab] (from child language)

are not harmony (Gafos, 1996b[1999]). Similarly, consider the four examples of 'dorsal harmony' in Hansson (2001[2010]). Gafos (1996b[1999]) would not refer to these as harmony (as they involve dorsal-uvular place of articulation features that cannot propagate through vowels). Terms can confuse but are not essential. What is essential is that strict locality does not predict that these phenomena should not exist; it only predicts that if they exist they could not have originated (solely) in coarticulation in vocal tracts and spreading should not be the mechanism of their analysis: different phenomena, different mechanisms (spreading vs. copying), reflecting their distinct origins and concomitantly distinct typological patterning.

In fact, when all is taken into account, Hansson (2001[2010])'s survey of 'consonant harmony' phenomena validates this prediction: certain cases of long distance identity involve spreading whereas others involve a different mechanism. That Hansson (2001[2010]) considers the latter cases to be the prototypical cases of 'consonant harmony' is inessential. To wit, Kinyarwanda is 'clear evidence that … harmony is achieved by means of strictly-local feature spreading (gestural extension)' (Hansson 2001[2010]: 168), but it 'is different in kind from other instances of consonant harmony' (Hansson 2001[2010]: 195). Consider also Sanskrit: 'We may safely conclude that Vedic Sanskrit n-retroflexion does involve (local) spreading, and that it is thus distinct from the other phenomena that are categorized as consonant harmony in this work' (Hansson 2001[2010]: 192). Also, 'It is possible that certain other coronal harmony phenomena based on retroflexion are also cases of spreading rather than agreement. For example, some of the coronal stop harmonies (primarily in Australian and Dravidian languages) discussed by Steriade (1995b) and Gafos (1999) may well be of this type' (Hansson, 2001[2010]: Footnote 48). 'Agreement,' in the last excerpt, means (not spreading but) correspondence-based identity, the same mechanism used in Gafos (1996a, 1996b[1999], 1998) in reduplicative and also in non-reduplicative contexts (Gafos, 2003, 2018). All in all, then, Sanskrit, Kinyarwanda, 'certain other coronal harmony phenomena,' and 'some of the coronal stop harmonies (primarily in Australian and Dravidian languages)' are not prototypical cases of 'consonant harmony' for Hansson (2001[2010]), but they are or could be cases where a strictly local spreading (not an agreement) analysis is conceded. In other words, when we collate all the, for Hansson (2001[2010]), 'different in kind' cases, a generalization without any exceptions emerges: all such cases instantiate precisely what is predicted in Gafos (1996b[1999]: 176) – evidence for strictly local spreading *only* for properties that can propagate through the vowel – and referred to as consonant harmony therein and elsewhere (Flemming, 1995; Steriade, 1995ab; Ní Chiosáin and Padgett, 2001; Walker et al., 2008).

I conclude by addressing a related point of confusion: 'Under the hypothesis of Strict Locality (Gafos 1999), all cases of LDA are reduced to spreading operations' (Heinz, 2010: 641). LDA stands for 'long distance agreement' ('agreement' is used here in its descriptive sense, not in the grammar mechanism sense). That all instances of LDA are due to spreading is a possible thesis, but it is not the one advocated in Gafos (1996b[1999]). Once again, Gafos (1996b[1999]) argues that spreading should not be the means of achieving identity for entire classes of LDA, both whole segment but also feature-only identity (see 3.1). Instead, copying is argued to be involved. Abiding to strictly local spreading is not the same as saying that all cases of long distance identity are due to spreading.

3.3 Feature identity not due to spreading

Walker (2000) takes up phenomena displaying long distance feature identity and argues for a correspondence-based (not spreading) analysis of these effects, beyond the total segmental identity and feature-specific identity (ATB effects) cases seen in 3.1 where Gafos' (1996b, 1998) original correspondence-based reanalysis of former spreading phenomena were developed (see also Rose and Walker, 2004; Arsenault and Kochetov, 2008; Hansson, 2001; Arsenault, 2012; Danis, 2019). For example, in Ngbaka, tautomorphemic homorganic stops must have the same voicing (/pɛpu/ 'vent', /babã/ 'companion'), but when heterorganic stops or consonants of other manners are involved, this constraint is lifted (/bata/ 'three', /gapa/ 'to divide', /tolo/ 'strike').

Identity of voicing in Ngbaka and the analyses of other cases exhibiting such patterns is effected via a correspondence relation between the consonants hosting the agreeing feature. As clarified in 3.2, strictly local spreading does not predict that Ngbaka-like identity cases should be unattested; it only precludes spreading as the

mechanism of their analysis. Strict locality precludes a spreading analysis of such patterns because voicing is not known to coarticulate from C-to-C through vowels (see Pearce 2005 for an illuminating reanalysis of a presumed case of C-to-C spreading of voicing over vowels in Kera). These patterns, then, could not have originated (solely) in coarticulation in vocal tracts. I will argue here for a storage-based discriminability approach to how such patterns emerge.

A striking property of long distance identity is that the hosts of the agreeing feature F must be similar. For instance, in Ngbaka, only homorganic stops agree in voicing and in Indus Kohistani only obstruents of the same manner (examples follow) agree in retroflexion. Walker (2000) formally expressed this using correspondence constraints projected from similarity scales. For instance, the constraint enforcing identity between homorganic (more similar) stops is higher ranked than that between heterorganic (less similar) stops.

Whence the similarity prerequisite? Rose and Walker (2004: 489) suggest pressures from speech planning are implicated, citing studies on similar consonant misproductions or 'slips of the tongue' as in [s…ʃ] → [ʃ…ʃ] (see also Walker, 2000; Hansson, 2001[2010]; Tilsen, 2019 for a different take and modeling). To render plausible how errors might give rise to identity effects in terms of different features, Rose and Walker (2004) cite evidence that some errors are not audible and thus that errors may occur also in non-sibilant contexts. For example, Goldstein et al. (2007) show that in fast repetitions of /kop top/ an extra tongue body gesture may appear during /t/. When this happens, there is no audible effect. This raises a concern. If the speaker knows the intended lexical representation and the error (the extra tongue body gesture in /kop top/) is not audible (as would be required for the listener to adopt the error by changing her lexical representation eventually), how does the error get transmitted? Note that this does not challenge the plausibility of Rose and Walker's proposal for sibilants. The error is sufficiently audible to be caught even by the unaided ear (Fromkin, 1971) in the [s…ʃ] → [ʃ…ʃ] case; but not so for stops as in the /kop top/ example above. The concern thus is with the generalizability of the errors proposal to the rest of the cases.

Consider a lexicon with CVC roots, with the Cs freely drawn from all manners (stops, fricatives, and so on). Roots with two stops are a subset of that set. Members of this subset are more similar to one another (than to roots outside this subset). As the set of roots shrinks to the smaller set of more similar CVCs (both Cs are stops), maximizing discrimination (using a 0vs2 contrast, C-C vs. C^F-C^F), becomes more pressing in this narrower subset. This is because, as I will show, narrow subsets amplify discrimination pressures (due to coarticulation) compared to less narrow subsets.

In fleshing out this approach, I will adopt, from models of lexical access, the so-called recall likelihood or, more generically, the sampling probability of a word (from a set of stored words) given a phonetic representation: the sampling probability of word $/i/$ given phonetic representation $[j]$, $P(/i/, [j])$, is given by a ratio of similarities, $P(/i/, [j]) = \frac{s(i,j)}{\sum_k s(j,k)}$ (Luce and Pisoni, 1988; Luce, 1963; Shepard, 1957). Similarity $s(i, j)$ between the phonetic form $[j]$ and the memorized form $/i/$ is in the nominator (because recall likelihood of $/i/$ from phonetic form $[j]$ is directly proportional to the similarity between $/i/$ and $[j]$) but it is divided by the sum of the similarities between $[j]$ and all stored forms /k/. Similarities are computed as in Section 2.

Observe now that, from set {/kaki/, /gaki/} to set {/kaki/, /gagi/}, the likelihood ratio improves. This is because the ratio's denominator decreases relative to the numerator in the latter compared to the former set. For concretness, compare set {/kaki/, /gaki/}, illustrating the 0vs1 contrast, with the Ngbaka-like {/kaki/, /gagi/} set, illustrating the 0vs2 contrast. To be compared, more specifically, is the likelihood ratio $P(/\text{kaki}/, [\text{kaki}])$ of word /kaki/ given the phonetic form [kaki] across the two sets. For {/kaki/, /gaki/}, the similarities are s(/kaki/, [kaki]) = e^{-0} = 1 (since d(/kaki/, [kaki]) = 0) and s(/gaki/, [kaki]) = e^{-1} = 0.37 (since d(/gaki/, [kaki]) = 1). Therefore, $P(/\text{kaki}/, [\text{kaki}]) = \frac{e^{-0}}{e^{-0}+e^{-1}}$ = 0.73. For {/kaki/, /gagi/}, s(/kaki/, [kaki]) = e^{-0} = 1 and s(/gagi/, [kaki]) = e^{-2} = 0.13 (since d(/gagi/, [kaki]) = 2), with $P(/\text{kaki}/, [\text{kaki}]) = \frac{e^{-0}}{e^{-0}+e^{-2}}$ = 0.88, higher than in the 0vs1 contrast set.

This offers a glimpse of why the 0vs2 contrast set is preferred over the 0vs1 contrast set, but does not yet by itself provide any direct insight on the similarity requirement. To better understand the

basis for this requirement, one must consider the forces of coarticulation and how they, in concert with (but at different time scales from) the lexical storage factors identified here, lead lexica in the direction of evolving an agreement imperative.

In illustrating the idea, I will use a case involving coronals. For coronals, a great number of long distance identity phenomena *look* like cases of strictly local spreading (3.2), but exhibit a fossilized lexical character. As there is no evidence for active spreading, a copying correspondence-based analysis is called for here. For example, in Indus Kohistani (Zoller, 2005; Arsenault, 2012), within roots, coronals of the same manner must agree in retroflexion; see (2) for stops and (3) for sibilants. No agreement imperative is in effect when different manners are involved, e.g., [dùːʂ, siʈìː, sìːʈʰ, ʒèːʈʰ, ʃòʈʰ] (‘sin’, ‘whistle’, ‘rich’, ‘name of a month’, ‘a bump’).

a.	tʌtʰ	‘hot; heat’
	dítʰi	‘given’
b.	ʈaːʈʰ	‘a small rug’
	ɖíʈʰi	‘span of hand’

Table 2. Retroflex agreement between stops.

a.	sʌzúː	‘sister’s son’
	zʰʌnzéːr	‘a kind of bird’
b.	ʂìʂ	‘a head’
	ʐàːʈʂ	‘a grape’

Table 3. Retroflex agreement between sibilants.

What is going on here? In the CVC(V) context, different factors conspire for the same result. The first is the tip-blade’s coarticulatory ability to extend through vowels. The second is reduction in discriminability between forms in pairs such as /sVʂ/-/ʂVʂ/, /sVʃ/-/ʃVʃ/. These pairs exemplify the 1vs2 contrast (one vs. two retroflexes in /sVʂ/-/ʂVʂ/ and one vs. two alveopalatals in /sVʃ/-/ʃVʃ/) with the first member in each pair realized variably (due to the first factor) as [ʂVʂ], [ʃVʃ] respectively, hence the discriminability reduction. Perhaps, as discussed earlier, a third factor, errors in planning is also at play. The point is that pressures from different systems (coarticulation, discrimination, planning) conspire for the same outcome, agreement in [±retroflex] or [±anterior] on the consonants in the CVC context. This convergence seems to explain why long distance identity is so robustly reported to involve coronals as opposed to other consonants. The present focus, however, is not on comparing frequencies of long distance identity patterns for different features (e.g., [±retroflex] vs. [±voice]), which requires discussion of processes other than coarticulation in other consonant classes (see concluding paragraph), but rather to identify a more robust basis for the similarity requirement. To return to this issue, using our running example, the key question is: why does an agreement imperative develop in the same-manner {/sVʂ/, /ʂVʂ/} or {/tVʈ/, /ʈVʈ/} sets but not in the mixed-manner {/tVʂ/, /ʈVʂ/} set?

Consider the consequences of regressive coarticulation in terms of the feature F in a CVC^{F} context, where F is [±retroflex] or [±anterior]. For concreteness, consider the case where F is [±retroflex] and the form is /CVʂ/ with a final retroflex fricative. The key point is that, as F propagates through the V to affect the C on the other side, the acoustic consequences of this coarticulatory effect are most salient when that C is also a fricative. In /sVʂ/ → [ʂVʂ], regressive coarticulation (‘→’ denotes that [ʂVʂ] is the phonetic coarticulated manifestation of /sVʂ/) results in the same and hence most salient, due to being repeated, acoustic consequences on the two fricatives. Regressive coarticulation is in effect also in /tVʈ/ → [ʈVʈ], with a crucial difference: the gesture propagating in /tVʈ/ is not the same as that in /sVʂ/. Retroflexion is an integral feature, that is, it is expressed differently between stops and fricatives. Articulatorily, Keating (1991) and Shridhar (1990) document distinct mid-sagittal postures between stop and fricative retroflexes. Acoustically, short term noise spectra for fricatives exhibit a lower first peak and a lower starting point than stops and sonorants (Hamann, 2003; Zygis and Hamann, 2003).

Thus, in /tVʂ/, the articulatory configuration of retroflexion originating from the fricative and propagating through V to affect the /t/ and change it to [t^{+}] in [t^{+}Vʂ] (the coarticulated phonetic form of /tVʂ/) is not the same as that of a /ʈ/. Hence, effects of coarticulation neutralize distinctions with other roots *within* each manner class, so that /sVʂ/ → [ʂVʂ] merges with /ʂVʂ/ and /tVʈ/ → [ʈVʈ] merges with /ʈVʈ/, but not so in roots from different manner classes. That is, /tVʂ/ → [t^{+}Vʂ] does not merge with /ʈVʂ/. In [t^{+}Vʂ], the retroflexion originating from the fricative /ʂ/

leaves some auditory cues on the stop [t], denoted here by [t$^+$], but these are different from the cues to retroflexion on [ʈ].

We can now fill in the specifics of the memory-based account sketched earlier. Compare lexical set {/tVʂ/, /ʈVʂ/}, where roots contain consonants from different manner classes, to set {/sVʂ/, /ʂVʂ/}, where roots contain consonants from the same manner class. Both sets exhibit a non-agreement pattern. We wish to identify a basis for why an agreement imperative develops in the latter but not in the former set. To do so, we consider the sampling probability $P(/tVʂ/, [tVʂ])$ of stored word /tVʂ/ given the phonetic form [tVʂ], the coarticulated phonetic form deriving from /tVʂ/. To make clear the intended source of any phonetic form, instead of $P(/tVʂ/, [tVʂ])$, I henceforth write $P(/tVʂ/, [tVʂ]\leftarrow/tVʂ/)$. For set {/tVʂ/, /ʈVʂ/}, the similarities of [tVʂ] to the two stored words are s(/tVʂ/, [tVʂ]) = e^{-0} = 1 (since d = 0) and s(/ʈVʂ/, [tVʂ]) = e^{-1} = 0.37 (since d = 1). Hence, $P(/tVʂ/, [tVʂ]\leftarrow/tVʂ/) = \frac{e^{-0}}{e^{-0}+e^{-1}}$ = 0.72.

Consider now the same-manner lexical set {/sVʂ/, /ʂVʂ/}. The problem met by this non-agreement observing set is that coarticulated /sVʂ/ → [ʂVʂ] is more similar to the other lexical item /ʂVʂ/ than it is to the co-existing /sVʂ/ in this set. The probability $P(/sVʂ/, [ʂVʂ]\leftarrow/sVʂ/)$ of /sVʂ/ given phonetic form [ʂVʂ] (the coarticulated phonetic form corresponding to /sVʂ/) is computed on the basis of the similarities: s(/ʂVʂ/, [ʂVʂ]) = e^{-0} = 1 (since d = 0) and s(/sVʂ/, [ʂVʂ]) = e^{-1} = 0.37 (since d = 1). Hence, $P(/sVʂ/, [ʂVʂ]\leftarrow/sVʂ/) = \frac{e^{-1}}{e^{-1}+e^{-0}}$ = 0.27 which is lower than the corresponding ratio for the {/tVʂ/, /ʈVs/} set. Correct recall likelihood decreases in this set (and incorrect recall $P(/ʂVʂ/, [ʂVʂ]\leftarrow/sVʂ/)$ increases as the reader can verify) compared to the {/tVʂ/, /ʈVʂ/} set. Both sets {/tVʂ/, /ʈVs/} and {/sVʂ/, /ʂVʂ/} exhibit the 1vs2 contrast but their correct recall likelihood ratios are quite distinct. The same-manner set amplifies discriminability losses compared to the different-manner set.

The assumption that the coarticulated output of /sVʂ/ is [ʂVʂ] and the assignment of a 0 difference between it and /ʂVʂ/ versus the assignment of a difference of 1 between the coarticulated output [t$^+$Vʂ] of /tVʂ/ and /ʈVʂ/ are non-essential. The only crucial assumption is that the coarticulated output of /sVʂ/ is more similar to [ʂVʂ] than coarticulated [t$^+$Vʂ] is to /ʈVʂ/. This assumption was argued to be valid on the basis of differences in the articulatory and acoustic realization of retroflexion between stops and fricatives.

In sum, using concepts established in models of lexical storage, I have shown that correct recall for lexical set {/tVʂ/, /ʈVʂ/} is higher than for set {/sVʂ/, /ʂVʂ/}. This is the proposed basis for the similarity prerequisite and in particular for the asymmetry that an agreement imperative is in effect in roots with consonants from the same manner class but not in roots with consonants from different manner classes.

Although extension of the approach to other features is beyond the present scope, the general take should be clear. There is a trade-off between more inclusive contrast sets and correct recall. Long distance identity phenomena are language-specific ways of resolving this trade-off. Take voicing, for instance. Consider the lexical set {/dVtV/, /dVdV/, /tVtV/, /tVdV/} vs. the Ngbaka-like set{/tVtV/, /dVdV/}. The latter set shows improved correct recall ratios over the more inclusive set. Instead of regressive coarticulation, intervocalic stop voicing (as /tVtV/ → [tVdV] and /dVtV/ → [dVdV]) seems to offer a reasonable candidate for phonetic pressures leading to contrast reduction here, combined perhaps with selection pressures for identity from learning mechanisms (Colavin et al., 2014; Gallagher, 2013). Finally, another property of long distance phenomena is that the identity imperative weakens with the distance between trigger and target sites. This property too seems to have a basis in memory (see Appendix C).

4 Conclusion

A great deal has been learned about phonological patterning by considering forces deriving from two sources: production and perception (Ohala, 1981, 1983; Beddor, 2009; Lindblom et al., 1995). Here, I have considered another so far largely neglected source: memory (see also Appendix B). I have argued, in particular, that the diversity and specific heretofore not well understood properties of long distance consonantal (non-)identity can be more fully explained when in addition to production and perception, the role of storage is considered.

Acknowledgments

This paper benefited greatly from the feedback of two anonymous reviewers. Thanks are also due to Maria Lialiou, Manfred Pastätter and Stephan Kuberski for numerous fruitful discussions, to Louis Goldstein for access to a dataset, originally collected by Cathe Browman and Louis Goldstein at the NIH X-Ray Microbeam facility to study coordination in syllables, from which I extracted the examples depicted in Figure 2, and to Kevin Ryan for information on the Kalevala. Preparation of this manuscript was supported, during its early stages by an ERC Advanced Grant 249440, which provided a focused time period for research, and in its later stages by a German Research Foundation (DFG) grant, 317633480, C04.

References

Diana Archangeli and Douglas Pulleyblank. 2007. Harmony. In Paul de Lacy (ed.), *The Cambridge Handbook of Phonology.* Cambridge University Press, pages 353-378.

Paul E. Arsenault. 2012. *Retroflex consonant harmony in South Asia*. Doctoral dissertation, University of Toronto.

Paul E. Arsenault and Alexei Kochetov. 2008. Retroflex harmony in Kalasha: agreement or spreading? In *Proceedings of NELS 39, the thirty-ninth annual meeting of the North East Linguistic Society (Volume 1).* GLSA, pages 56-56. https://doi.org/10.7282/T3Z60KZF.

Alan D. Baddeley. 1986. *Working memory*. Clarendon Press, Oxford.

Patrice S. Beddor. 2009. A coarticulatory path to sound change. *Language*, 85(4):407-428. https://doi.org/10.1353/lan.0.0165.

Noam Chomsky. 1980. *Rules and Representations*. Columbia University Press.

Noam Chomsky. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. Praeger.

Noam Chomsky and Morris Halle. 1968. *The Sound Pattern of English*. New York, Harper and Row.

Nick G. Clements. 1976. *Vowel harmony in nonlinear generative phonology: an autosegmental model*. Indiana University Linguistics Club, IN.

Rebecca Colavin, Roger Levy and Sharon Rose. 2014. Modeling OCP-place in Amharic with the maximum entropy phonotactic learner. *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 46, pages 27-41.

Nelson Cowan. 1995. *Attention and memory: An integrated framework*. New York: Oxford University Press.

Nelson Cowan and AuBuchon M. Angela. 2008. Short-term memory loss over time without retroactive stimulus interference. *Psychonomic Bulletin & Review,* 15:230-235. https://doi.org/10.3758/PBR.15.1.230.

Nick Danis. 2019. Long-distance major place harmony. *Phonology*, 36(4):573-604. https://doi.org/10.1017/S0952675719000307.

William K. Estes. 1972. An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (eds.), *Coding processes in human memory*. Washington, DC, pages 161-190.

Edward Flemming. 1995. Vowels undergo consonant harmony. Handout of talk presented at the Annual Trilateral Phonology Weekend, University of California, Santa Cruz, December 11, 1995.

Stephan Frisch. 1996. *Similarity and frequency in phonology*. Doctoral dissertation. Northwestern University.

Victoria A. Fromkin. 1971. The non-anomalous nature of anomalous utterances. *Language*, 47(1):27-52. https://doi.org/10.2307/412187.

Adamantios Gafos. 1996a. Correspondence in Temiar: No need for long distance spreading here. In W. de Reuse and S. Chelliah (eds.), *Papers from the Fifth Annual Meeting of the South East Asian Linguistics Society*, Tucson, AZ, pages 30-47.

Adamantios Gafos. 1996b. *The articulatory basis of locality in phonology*. Doctoral dissertation, Johns Hopkins University. [Published 1999, Outstanding Dissertations in Linguistics, Routledge.]

Diamandis Gafos. 1998. Eliminating long-distance consonantal spreading. *Natural Language & Linguistic Theory*, 16(2):223-278. https://doi.org/10.1023/A:1005968600965.

Adamantios Gafos. 2003. Greenberg's asymmetry in Arabic: a consequence of stems in paradigms. *Language*, 79(2):317-357. https://doi.org/10.1353/lan.2003.0116.

Adamantios Gafos. 2018. Stems and paradigms in Modern Standard Arabic and dialects. In Elabbas Benmamoun and Reem Bassiouney (eds.), *The Routledge Handbook of Arabic Linguistics*, Routledge, pages 62-86. https://doi.org/ 10.4324/9781315147062-5.

Gillian E. S. Gallagher. 2010. *The perceptual basis of long-distance laryngeal restrictions*. Doctoral dissertation, Massachusetts Institute of Technology.

Gillian Gallagher. 2013. Learning the identity effect as an artificial language: bias and generalisation. *Phonology*, 30(2):253-295. https://doi.org/10.1017/S0952675713000134.

Wendell R. Garner. 1974. *The processing of information and structure*. New York: Wiley.

Louis Goldstein, Marianne Pouplier, Larissa Chen, Elliot Saltzman, and Dani Byrd. 2007. Dynamic action units slip in speech production errors. *Cognition*, 103(3): 386-412. https://doi.org/10.1016/j.cognition.2006.05.010.

Sharon Goldwater and Mark Johnson 2003. Learning OT constraint rankings using a Maximum Entropy model. In Jennifer Spenador, Anders Eriksson & Östen Dahl (eds.) *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. pages 111-120.

Silke Hamann. 2003. *The phonetics and phonology of retroflexes*. Doctoral dissertation, University of Utrecht.

Gunnar Hansson. 2001. *Theoretical and typological issues in consonant harmony*. Doctoral dissertation, University of California, Berkeley.

Jonathan Harrington, Felicitas Kleber, Ulrich Reubold, Florian Schiel, Mary Stevens. 2018. Linking cognitive and social aspects of sound change using agent-based modeling. *Topics in Cognitive Science*, 10(4):707-728. https://doi.org/10.1111/tops.12329.

Bruce Hayes and Londe Zsuzsa. 2006. Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology*, 23:59-104. https://doi.org/10.1017/S0952675706000765.

Bruce Hayes and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39:379-440. https://doi.org/10.1162/ling.2008.39.3.379.

Jeffrey Heinz. 2010. Learning long-distance phonotactics. *Linguistic Inquiry*, 41(4):623-661. https://doi.org/10.1162/LING_a_00015.

Harry Hulst van der. 1985. Vowel harmony in Hungarian: A comparison of segmental and autosegmental analysis. In Harry van der Hulst and Norval Smith (eds.), *Advances in nonlinear phonology*. Dordrecht, Foris, pages 267-303. https://doi.org/10.1515/9783110869194-015.

Patricia A. Keating. 1991. Coronal Places of Articulation. In Paradis Carole and Jean F. Prunet (eds.), *The Special Status of Coronals: Internal and External Evidence*. Academic Press, New York, pages 29-48. https://doi.org/10.1016/B978-0-12-544966-3.50008-7.

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

Stephan R. Kuberski and Adamantios Gafos. 2019. The speed-curvature power law in tongue movements of repetitive speech. *Plos One*, https://doi.org/10.1371/journal.pone.0213851.

Aert Kuipers. 1974. *The Shuswap Language*. The Hague, Mouton.

Catherine L. Lee and Estes K. William. 1977. Order and position in primary memory for letter strings. *Journal of Memory and Language*, 16(4):395-418. https://doi.org/10.1016/S0022-5371(77)80036-4.

Stephan Lewandowsky and Simon Farrell. 2008. Phonological similarity in serial recall: Constraints on theories of memory. *Journal of Memory and Language*, 58(2):429-448. https://doi.org/10.1016/j.jml.2007.01.005.

Björn Lindblom, Guion Susan, Hura Susan, Seung-Jae Moon, and Raquel Willerman. 1995. Is sound change adaptive? *Rivista di Linguistica*, 7(1):5-37.

Duncan R. Luce. 1963. Detection and recognition. In Luce R. Duncan, Robert. R. Bush, and Eugene Galanter (eds.), *Handbook of mathematical psychology (Volume 1)*. Wiley, pages 103-189.

Paul A. Luce and David B. Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19(1):1-36. https://doi.org/10.1097/00003446-199802000-00001.

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

Margaret MacEachern. 1999. *Laryngeal co-occurrence restrictions.* New York: Garland.

Sara Mackenzie. 2009. Contrast and similarity in consonant harmony processes. Doctoral dissertation, University of Toronto.

John J. McCarthy. 1979. *Formal problems in Semitic phonology and morphology*. Doctoral dissertation, MIT. [Published 1985, New York, Garland Publishers]

John J. McCarthy. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry*, 12:373-418. https://www.jstor.org/stable/4178229.

John J. McCarthy. 1983. Consonantal morphology in the Chaha verb. In *Proceedings of WCCFL* (Volume 2). Somerville, MA, Cascadilla Press, pages 176-188.

John J. McCarthy and Alan Prince. 1995. Faithfulness and reduplicative identity. *UMOP*, 18:249-384. https://doi.org/doi:10.7282/T31R6NJ9.

Joyce McDonough. 1991. On the representation of consonant harmony in Navajo. In *Proceedings of the Tenth West Coast Conference on Formal Linguistics*, pages 319-335.

Robert S. McLean and Lee W. Gregg. 1967. Effects of induced chunking on temporal aspects of serial recitation. *Journal of Experimental Psychology*, 74(4):455-459. https://doi.org/10.1037/h0024785.

Ger-Jan Mensink and Jeroen G. W. Raaijmakers. 1988. A model for interference and forgetting. *Psychological Review*, 95:434-455. https://doi.org/10.1037/0033-295x.95.4.434.

Bennet B. Murdock. 1982. A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89:609-626. https://doi.org/10.1037/0033-295X.89.6.609.

Bennet B. Murdock. 1997. Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review*, 104:839-862. https://doi.org/10.1037/0033-295X.104.4.839.

James S. Nairne. 1990. A feature model of immediate memory. *Memory and Cognition*, 18:251-269. https://doi.org/10.3758/BF03213879.

James S. Nairne. 1991. Positional uncertainty in long-term memory. *Memory and Cognition*, 19(4):332-340. https://doi.org/10.3758/BF03197136.

James S. Nairne. 2001. A functional analysis of primary memory. In Henry L. Roediger, James S. Nairne, Ian Neath, and Aimee M. Surprenant (eds.), *The nature of remembering: Essays in honor of Robert G. Crowder*. Washington, DC, American Psychological Association, pages 282-296. https://doi.org/10.1037/10394-015.

Ian Neath and Aimée M. Surprenant. 2003. *Human memory: An introduction to research, data, and theory*. Second edition. Belmont, CA: Wadsworth.

Máire Ní Chiosáin and Jaye Padgett. 2001. Markedness, segment realization, and locality in spreading. In Linda Lombardi (ed.), *Segmental phonology in Optimality Theory: constraints and representations*. Cambridge, pages 118-156. https://doi.org/10.1017/CBO9780511570582.005.

John J. Ohala. 1981. The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, and M. F. Miller (eds.), *Papers from the Parasession on Language and Behavior*, Chicago: Chicago Linguistic Society, pages 178-203. https://doi.org/10.1075/cilt.323.05oha.

John J. Ohala. 1983. The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.), *The production of speech*. New York: Springer, pages 189-216. https://doi.org/10.1007/978-1-4613-8202-7_9.

Mary Pearce. 2005. Kera tone and voicing. *UCL Working Papers in Linguistics* 17. pages 61-82. https://doi.org/10.1016/j.lingua.2007.10.023.

Janet Pierrehumbert. 1993. Dissimilarity in the Arabic verbal roots. *Proceedings of NELS 23*. Amherst: GLSA Publications. pages 367-381.

Sharon Rose and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language*, 80:475-531. https://doi.org/ 10.1353/lan.2004.0144.

Kevin Ryan. 2016. Strictness functions in meter. Handout of MIT Phonology Circle, March 28, 2016.

Roger N. Shepard. 1957. Stimulus and response generalization: A stochastic model relating generalization to psychological space. *Psychometrika*, 22:325-345. https://doi.org/10.1007/BF02288967.

Roger N. Shepard. 1987. Toward a universal law of generalization for psychological science. *Science*, 237:1317-23. https://doi.org/10.1126/science.3629243.

Paul Smolensky and Géraldine Legendre. 2006. *The harmonic mind: From neural computation to optimality-theoretic grammar*. Cambridge, MA: MIT Press.

Shikaripur N. Sridhar. 1990. *Kannada*. London, Routledge.

Donca Steriade. 1995a. Underspecification and Markedness. In John Goldsmith (ed.), *Handbook of Phonological Theory*, Blackwell, pages 114-174.

Donca Steriade. 1995b. *Positional neutralization*. Ms. University of California, Los Angeles.

Sam Tilsen. 2019. Motoric mechanisms for the emergence of non-local phonological patterns. *Frontiers in Psychology*, 10:2143, https://doi.org/10.3389/fpsyg.2019.02143.

Simon Todd, Janet B. Pierrehumbert, and Jennifer Hay, 2019. Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition*, 185: 1-20. https://doi.org/10.1016/j.cognition.2019.01.004.

Rachel Walker. 2000. Long-distance consonantal identity effects. In *Proceedings of WCCFL (Volume 19).* Somerville, MA, Cascadilla Press, pages 532-545.

Rachel Walker, Dani Byrd, and Fidèle Mpiranya. 2008. An articulatory view of Kinyarwanda coronal harmony. *Phonology*, 25(3):499-535. https://doi.org/10.1017/S0952675708001619.

Adam Wayment. 2009. *Assimilation as attraction: Computing distance, similarity, and locality in phonology*. Doctoral dissertation, The Johns Hopkins University.

Andrew B. Wedel. 2006. Exemplar models, evolution and language change. *The Linguistic Review*, 23(3):247-274. https://doi.org/10.1515/TLR.2006.010.

Wayne A. Wickelgren. 1970. Time, interference, and rate of presentation in short-term recognition memory for items. *Journal of Mathematical Psychology*, 7(2):219-235. https://doi.org/10.1016/0022-2496(70)90045-3.

Douglas H. Whalen, Patricia Shaw, Aude Noiray, and Reethee Antony. 2011. Analogs of Tahltan consonant harmony in English CVC syllables. In *Proceedings of the 17th International Congress of Phonetic Sciences*. pages 2129-2132. https://doi.org/ 10.13140/RG.2.1.2966.8961.

Douglas H. Whalen and Mark K. Tiede. 2020. Testing predispositions in consonant harmony. Abstract presented at Ultrafest IX, 2020, October 21-24.

William D. Whitney. 1889. *Sanskrit Grammar.* London: Oxford University Press.

Caroline Wiltshire and Louis Goldstein. 1997. Tongue tip orientation and coronal consonants. In *Proceedings of the Eastern States Conference on Linguistics (Volume 14)*, pages 216-225.

Marzena Zygis and Silke Hamann. 2003. Perceptual and acoustic cues of Polish coronal fricatives. In *Proceedings of the 15th International Conference of Phonetic Sciences*. Barcelona, pages 395-398.

Ariel Zylberberg, Stanislas Dehaene, Gabriel B. Mindlin and Mariano Sigman. 2009. Neurophysiological bases of exponential sensory decay and top-down memory retrieval: a model. *Frontiers in Computational Neuroscience*. https://doi.org/10.3389/neuro.10.004.2009.

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

A Alternatives to feature overwrite

The account of discrimination asymmetries in Section 2 can be given different implementations, all pointing to the same conclusion that short term memory mechanisms are involved. I chose an implementation that is simple and lends itself transparently to analytical treatment. It may be helpful here to give a sense of what other options exist for other accounts.

Within the field of memory research, one finds the expected (within field) debate on the specifics of the implicated mechanisms. For example, instead of feature overwrite (Nairne, 1990; Nairne, 2001; see also Neath and Surprenant, 2003), another mechanism referred to as feature overload may be used to account for the same asymmetries. The starting intuition here is that features contribute to discriminability only if they are sufficiently distinctive. Distinctiveness is related (in part) to how often a feature appears among the set of forms that are being discriminated. Features may become less distinctive when they are repeated ('overload'). This intuition finds a number of instantiations in the memory and perception literature under various names, e.g., adaptation, recalibration, and here feature overload. For memory models, as Nairne (1990: 255) writes, 'if the intact modality-dependent features are "overloaded" in the sense that they occur in all or a number of possible secondary memory traces …, then number of features, per se, will not map into a performance advantage'. Recall that in the 1vs2 condition, the stimuli are [k'ap'i-k'api], [kap'i-k'ap'i], [k'ap'i-kap'i], and [k'api-k'ap'i]. This 1vs2 condition is the complement of the 0vs1 condition: in 0vs1 only one segment is an ejective whereas in 1vs2 only one segment is not an ejective. Hence, the distances are the same for each pair in both conditions and thus the lower discriminability of the 1vs2 condition must derive from the repetition of ejectivization. Since in the 1vs2 condition, ejectivization ([+CG]) is present on most segments, the effect of this feature overload can be appreciated by setting the b_k parameter in the distance function $d(i,j) = \frac{\sum b_k M_k}{N}$ to a value lower than in the 0vs1 condition. Because b_k multiplies the featural mismatch counter M_k, lowering b_k means that, in the 1vs2 condition, any [+CG] difference between forms i and j contributes less to the distance $d(i,j)$ than in the 0vs1 condition. Since similarity is given by $s(i,j) = e^{-d(i,j)}$, the decrease in $d(i,j)$ for pairs

in the 1vs2 condition results in increased similarity compared to the 0vs1 condition. The same result as in Section 2 is obtained.

One last note is due with respect to how feature overwrite applies in the analysis given in the main text. In pair [k'ap'i-k'api], overwrite could, in principle, apply across the two nonce words so that the output would be [k'api-k'api], instead of the output used in the analysis of the main text [kap'i-k'api]. I assume, along with the short term memory literature on grouping, also referred to as 'chunking' (McLean and Gregg, 1967 et seq.), that overwrite applies only within groups. For example, Nairne (1990) writes: 'An encoded primary memory trace, B, will overwrite the features of trace A if and only if trace B is perceived as belonging to the same list segment as trace A. This means that how a subject chooses to group items, presumably on the basis of global list structure, importantly determines if overwriting occurs' (Nairne, 1990: 253). Grouping in the task analyzed in Section 2 is imposed directly by the experimental design. Specifically, the two stimuli waveforms in any pair, e.g., [k'ap'i-k'api], were presented as two separate 'words' with an interstimulus interval of 300 ms between the two.

B (Im)perfect memory

An overarching theme of the main text is that memory, both short term and long term memory, has not been considered in any systematic way as a source of selection forces in sound change, but it is argued in this paper to play out in crucial ways in accounting for long distance consonantal restrictions. To clarify the first part of this statement, long term memory does play a role in exemplar approaches to sound change: a so-called 'rich' memory, an all-encompassing storage of phonetic details, in concert with lexical frequency considerations, is argued to play out in the course of sound change (Wedel, 2006; Harrington et al., 2018; Todd et al., 2019; among others). Here, I mean not the rich but the fallible memory in the same way Ohala (1981) emphasized the fallible parsing of coarticulation by perception. This is the sense of memory that is involved in Section 2 of and argued to account for the discrimination asymmetries discussed therein. In Section 3, the principles of long term memory implicated in deriving the similarity requirement (in particular, recall likelihoods and the models of lexical access in which these statistics operate) are orthogonal to the 'richness' of 'rich' memory issue. It is in these two, so far neglected in the study of sound change, senses of memory where the novelty of the claims in the main text resides.

A related note is in order. That (some) languages develop identity imperatives, as in C^F-C^F, in their lexica is not inconsistent with short term memory interference effects (as in feature overwrite for certain features discussed in Section 2). Multiple forces and different time scales are involved: long term lexical storage versus percent correct in a same versus different discrimination task in short term memory where probabilistic interference can be registered.

For the experimental results in Section 2 (Gallagher, 2010), recall that the 1vs2 condition was more difficult (that is, responses to same versus different were less accurate) than the 0vs1 condition, which in turn was more difficult than the 0vs2 condition. Percent correct values per condition were around 65% (1vs2), 75% (0vs1), and 90% (0vs2), depending on the feature repeated and place of articulation. Thus, all conditions were well above chance, but reliably different from one another. It is these differences between conditions that the account in Section 2 derives.

Short term memory is a system for storing briefly presented words and manipulating these or taking decisions on these, as in responding to whether two just heard stimuli are the same or different, under time pressure. In those conditions, memory is limited (Baddeley, 1986). When these conditions do not apply (i.e., participants are not tasked with a discrimination decision and no time pressure is imposed), speakers can produce and listeners can perceive words with identical or similar segments.

C Decay with distance and time scales

There is evidence (Pierrehumbert, 1993; Frisch, 1996; Hayes and Londe, 2006; Wayment, 2009; Zymet, 2014) that the strength of phonotactic restrictions, including phonotactic restrictions of the long distance type, decays with distance in many cases. Different approaches to this property have been proposed (see, among others, Zymet, 2014; Kimper, 2011). Regardless of the specifics of any approach on capturing the decay property, a key question is: what is the basis of this property?

The answer from the perspective of the present paper is once again memory, specifically here, memory for syntagmatic intra-word relations. To my knowledge, a connection between decay in the strength of phonotactic restrictions and memory has not been drawn before. This is striking, given that the memory literature strongly indicates that traces decay (exponentially) with distance in space-time (among many others, see Wickelgren, 1970; Murdock, 1982, 1997; Mensink and Raaijmakers, 1988; Cowan, 1995; Cowan and AuBouchon, 2008; Zylberberg et al., 2009), but perhaps not entirely surprising given the time-honored implicit assumption that biases shaping phonological patterns derive mostly from production and perception (see the main text).

Now, whereas in the memory literature there is rather broad support for exponential decay (see references above), in the typological profile of long distance phonotactics the form of the decay, as argued below, appears to be more consistent with a power law rather than an exponential law. I will suggest in what follows that, assuming the power law trait is correct, this characteristic of the typological profile of long distance phonotactics may be related to the fact that what we observe at the level of the typology (not necessarily at the level of the individual) is shaped by contributions of forces at different time scales. At the most basic level, we can consider two time scales, a vertical and a horizontal. The vertical corresponds to the time scale at which a learner (say, a Harmonic Grammar learner) infers a set of principles and their prioritization from ambient input. The horizontal corresponds to the much slower time scale at which lexica (and other components of the language system) change. The reason why it is justified to assume that what happens in the horizontal time scale can be put aside for the purposes of inferring the form of the principles and the representations in individuals' grammars is not because lexica do not change but because change at that time scale is so slow, relative to the vertical time scale, that we can effectively treat it as a constant. But can we evade the different time scales when what is at issue is typology and in particular here the precise form of the decay in the typological profile of long distance phonotactics?

Let us embed the time scales acknowledgment in the context of long distance consonantal identity patterns. Recall the conspiracy of the three factors converging on the same outcome of identical [±anterior] or [±retroflex] consonants in a CVC(V) from Section 3.3: (factor 1) the propensity of the tip-blade to coarticulate (strictly locally) through vowels and neutralize the [±anterior] contrast between (pre-harmony stage) /sVʃ/-/ʃVʃ/ lexical pairs, (factor 2) the auditory saliency of repeated values of [±anterior] in sibilants (that is, the fact that the coarticulated output of /sVʃ/ → [ʃVʃ] is salient for listeners due to the repetition of the same value of [±anterior]), and perhaps (factor 3) the propensity of planning errors in such sequences of sibilants. The convergence of these factors may be seen to characterize the early stages of the development of long distance identity. At later stages, processes of extension of the short-range CVC(V) context must necessarily take effect, so that the pattern ultimately ends up holding also within larger spans, as in /sVpVʃ/ → [ʃVpVʃ], wherein the trigger and the target sites are separated by more than a single vowel. The factors implicated during that short span to longer span transition appear to draw on the auditory saliency of repeated sibilants. That is, sequences of repeated [s] versus repeated [ʃ] present the listener-learner with a salient dichotomy in spectral energy plateaux. The wider and somewhat more retracted channel of [ʃ] results in a turbulence of lower ('dull') frequencies compared to that of higher ('sharp') frequencies [s]. A division of words into two classes along the single dimension of spectral energy, 'dull' versus 'sharp' sibilants, suffices to capture the phonotactic. Furthermore and this becomes crucial now, the neutralization of the lexical contrast between (pre-harmony stage) lexical pairs /sVʃ/-/ʃVʃ/ to post-harmony stage /ʃVʃ/ (that is part of factor 1 in the preceding) has its own intrinsic time scale, which is different (much slower) from the time scales of the processes implicated in the listing of these three factors (coarticulation, perception, and planning).

Recognition of the different time scales that may contribute to the shaping of long distance phonotactics need not be seen as a nuisance or an aspect of the phenomenon that must be evaded. It can be turned into a hypothesis that makes specific predictions about the form of the decay.

A signature property of phenomena shaped by processes or mechanisms operating at different time scales is power laws. The power law put forth in what follows is a syntagmatic, intra-word law, just as with the (non-)identity imperatives in Section 2 and Section 3 of the main text, but stated in gradient terms so as to allow for expression of the decay property. Said in other words, the intra-word law expresses the strength of the relation between two word positions as a monotonically decreasing function of the distance between the

two positions. The monotonic decrease must be a decrease of a particular kind for the relation to be governed by a power law. Specifically, for long distance phonotactics, the strictness or the weight of the co-occurrence restriction should decay with distance x as in

$$w(x) = a\,x^{-\beta} \qquad (1)$$

where α and β are empirically determined constants. An equivalent formulation in which Eq. (2) has been transformed by taking the logarithm of both sides, is given by

$$logw = loga - \beta\, logx \qquad (2)$$

which is a straight line, in the strength by distance $logw$ - $logx$ plane, whose slope corresponds to the negative exponent β and whose intercept corresponds to $loga$. Power law behavior is scale independent. This means that every time distance increases by, say, a factor of 2, the weight of the co-occurrence restriction decreases by a factor of $2^{-\beta}$. Equivalently, $w(2x)/w(x)$ or $w(\varphi x)/w(x)$ for any factor φ is independent of x.

In our domain of long distance restrictions, if multiple time scales do contribute to the precise form of the phonotactic, then strong predictions ensue. The most encompassing of these is that proportional change in the strength of a long distance co-occurrence restriction could in principle apply at different distance granularities (segment, syllable, foot, morpheme and so on). This does not mean that the unit of distance is irrelevant for any given long distance restriction. It only means that power laws could in principle live at different granularities (where distance is defined over units as diverse as vowels, syllables, feet, morphemes and so on).

The requisite evidence for testing a power law conjecture in phonotactics is presently sparse. Relevant sources can be found in Martin (2015) and Zymet (2014) from segmental phonology and in Ryan (2016) from metrification. Martin (2015) studies the strictness of [±anterior] identity in Navajo coronals, verifying that within roots identity is categorical (McDonough, 1991), but when the coronals belong to different constituents in a compound the identity imperative weakens with the syllabic distance between the coronals. Thus. in compounds, when the sibilants are in adjacent syllables as in [**ts**^{h}é.**s**ǫ?] ‘glass’, identity is enforced at a rate of 66.4%, when the sibilants are separated by one syllable as in [ná**ʃ**.dóí.**ts**^{h}oh] ‘mountain lion’ enforcement rate is 24.2%, when separated by two syllables as in [**ts**i.dił.há.**ʃ**ii] ‘scorpion’ the rate is 8.5%, and when separated by three as in [**ts**^{h}ą́.ho.di.niih.**ts**^{h}oh] ‘typhoid fever’ the rate is 0.9%. However, the dataset consists in a total of 211 datapoints which is unfortunately not conducive to any meaningful statistics.

Zymet (2014) studies the effect of distance on rounding dissimilation in Malagasy, liquid dissimilation in Latin and English, and vowel harmony in Hungarian. Working in a Harmonic Grammar model (Smolensky and Legendre, 2006), Zymet (2014) proposes to scale the weight of the constraint expressing any given phonotactic restriction by a multiplicative factor $d(x) = 1/x^{k}$ (where x is distance between target and trigger sites) and variously refers to such scalers as ‘negative power’ or ‘(inverse) exponential’ functions (citing Kimper 2011 for the exponential property). Strictly speaking, these functionals are simplified power laws where a in Eq. (1) is set to unity (or the intercept in Eq. (2) is 0 and β is k). For concreteness, an exponential (decay) law would take the form

$$w(x) = \gamma\,\beta^{-x} \qquad (3)$$

where γ and β are constants or its equivalent, by taking the log of both sides,

$$logw = log\gamma - x\, log\beta \qquad (4)$$

representing a straight line in the $logw$ - x plane, a so-called semi-log plot (because only w is transformed logarithmically) as opposed to the log-log plot which would be appropriate for revealing the straight line corresponding to a power law as in Eq. (2). In an exponential law, Eq. (3), the base is a constant and the exponent is the variable x (distance between target and trigger). In a power law, Eq. (1), the exponent is a constant and the base is the variable.

In principle, laws of both forms can capture decay. Here, the task will be to compare the two law forms, exponential and power law, in their ability to fit long distance decay in the domain of phonotactics. The first comparison I will present derives from metrification data reported in Ryan (2016). This choice is based on two reasons. First, in the metrification case there is more data than in other domains. Second, Ryan (2016) cautiously sums up the observation, which appears implicit in present day phonological thinking perhaps due to the wide applicability of exponentials, that long distance phonotactics may obey an exponential pattern: ‘Distance-based decay (and perhaps all scalar mapping) in phonology, metrics, etc. seems generally to be exponential’ (Ryan, 2016: 3).

Meter in the Kalevala, the empirical ground of Ryan (2016), consists in lines of eight syllables

(in couplets) with alliteration within lines of the trochaic form: a line consists in four feet, each made out of a strong-weak syllable pair as in (S W) (S W) (S W) (S W). Primary stressed syllables should be heavy in the S and light in the W sites. At issue in Ryan (2016) is the strictness of the meter with respect to the two weight constraints. Ryan (2016) demonstrates, based on 17,485 script lines, that exceptions to the weight constraints are numerous at the beginnings of lines but quickly fall off with the distance from the start of the line. For example, for strong syllables, in the first position of the line there are 5,456 exceptions to the weight constraint out of 16,484 (33,1%), in the third position there are 95 exceptions out of 9,539 (1%), in the fifth position position there are 15 exceptions out of 8,777 (0,2%) and so on. There is thus a dramatic decrease (with distance) in the freedom exercised by the poet(s) in violating the weight constraints. The question now is whether this decay follows a law and if so what the form of that law may be. Ryan (2016) plots percent of violations on a log-y by position plane in an aim of assessing linearity as would be predicted by an exponential law (when the data are depicted on that scale) and finds reasonable conformity in that such a law captures the sharply decreasing shape of constraint violations. Here, I ask a more specific question by contrasting the performance of an exponential to that of a power law. This is done in Figure 3. The semi-log plot (top) redraws what is shown in Ryan (2016: 2) but augmented here with fits from both laws in the same plot. The log-log plot (bottom) is added to better appreciate the difference in fit between the two laws.

Recall that an exponential law is suggested by a straight line in the semi-log plane and a power law by a straight line in the log-log plane. Linear regressions are performed on the log-y values. Instead of the data plotted in Ryan (2016) (under item (4) in Ryan 2016: 2, which gives violation percentages), the values on the ordinate in Figure 3 derive from the original constraint violation counts for the weight constraints (as these offer better accuracy).[1] Regression errors (reduced chi-squared) are shown in the legend within each panel. These errors reveal that the power law outperforms the exponential law for both weight constraints (S: 0.74 error for the exponential law vs. 0.22 for the power law; W: 0.97 vs. 0.47).

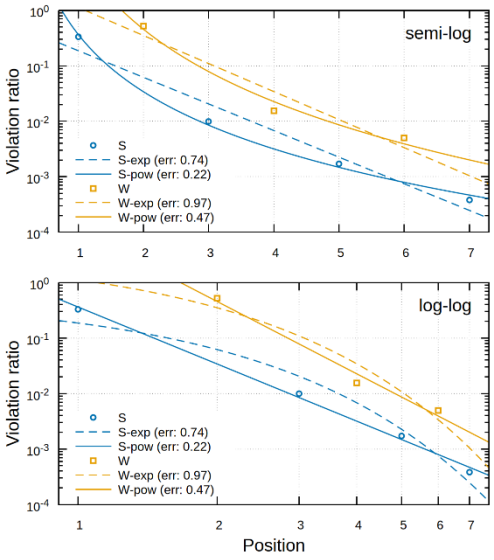

Figure 3. Exponential (dashed) and power law (solid) fits to the exceptions of the weight constraints in the Kalevala shown in a semi-log (top) versus a log-log (bottom) plot. Data points are shown as circles for strong 'S' syllables at positions 1, 3, 5, 7, and as rectangles for weak 'W' syllables at positions 2, 4, 6. 'S-exp' vs. 'S-pow' denote an exponential vs. a power law fit to the strong syllable weight constraint (and correspondingly, 'W-exp' vs. 'W-pow', for the weak syllable). Regression errors ('err') are in parentheses.

I present one more example, this time from English segmental phonology. The suffix /-əl/ as in 'distal' dissimilates to /-əɹ/ when attached to a liquid-final stem as in 'solar' and 'velar', but also at a distance as in 'lunar', 'lacunar' and so on. The rates at which dissimilation is enforced fall off rapidly with the number of syllables (0, 1, 2, 3) between the trigger in the root and target sites. From the four segmental processes studied in Zymet (2014), this case of dissimilation is the only one where the distance scale admits 4 distinctions, thus offering a better case-study for assessing the form of a potential underlying law than in the rest of the cases. Zymet (2014) used a maxent learner (Goldwater and Johnson, 2003; Hayes and Wilson, 2008) to estimate constraint weights for the markedness constraints banning two laterals in adjacent syllables, *[lat]-0-[lat], and separately for the constraints banning two laterals at a distance of 1, 2, and 3 syllables (*[lat]-1-[lat], *[lat]-2-[lat], *[lat]-3-[lat]). These weights, as given in Zymet (2014: 34), were used to fit an exponential and a power law in Figure 4.

[1] These counts are given in the table under (2) in Ryan (2016: 2). Thus, displayed on the ordinate in Figure 3 is a (log-transformed) constraint violation ratio, given by the count of exceptions in column 3 of the table, under (2) in Ryan (2016: 2), divided by the number of total stressed syllables given in column 4 of that table.

As in the metrification case, the power law seems to provide the better fit in this case from a long distance segmental phonotactic.

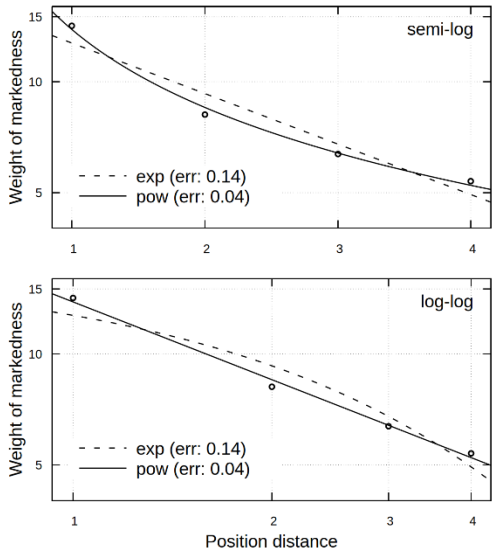

Figure 4. Exponential (dashed, 'exp') and power law (solid, 'pow') fits to English liquid dissimilation markedness weights. Data points are shown as circles. Regression errors ('err') are in parentheses.

In sum, there are indications that a power law captures the decay in the strength of long distance phonotactics better than an exponential law. In other cases where the data are available, putting aside phenomena that exhibit long distance effects but where there is reasonable support for local spreading (Section 3.2) and cases where the data are clearly insufficient, the fits show a general advantage for a power law. Based on the rest of the phenomena reported in Zymet (2014: 34), regression errors pattern as follows: 0.095 versus 0.015 for Malagasy dissimilation, 0.25 versus 0.14 for Latin liquid dissimilation, and 0.04 versus 0.074 for Hungarian vowel harmony, where the first value is the error for the exponential and the second is the error for the power law. That is, a power law provides the better fit with the exception of the Hungarian case. I leave further discussion of this pattern for future work.

To conclude, two points have been made. First, the decay property of memory seems to provide a basis for the generalization that long distance co-occurrence restrictions weaken with distance.

Second, at the level of typology, there are indications for the conjecture that the strength of long distance phonotactics may follow a power law with distance. This may be related to the different time scales that contribute to the formation of long distance phonotactic patterns.

Note that a power law conjecture, if validated, would not need to imply that the grammar within an individual should be responsible for the precise form of the decay in the typological profile of phonotactic strength. Rather, the point has been that the biases we can reasonably assume to be part of an individual's learning mechanism (exponential decay of memory traces) may not fully account for all aspects of the typological properties of the phonotactic (here, power law decay). Such a conclusion, if true, would be neither new nor surprising. What is new in the preceding is the realization that the power law conjecture for long distance phonotactics may help illuminate a way in which hypotheses about the tightness of the relation between the biases or the grammar in the individual and typological properties can be evaluated empirically if the requisite data can be made available.