JOURNAL *of* **SLAVIC** LINGUISTICS

VOLUME 32 (2024) · ISSUE NO. FASL 32 ISSUE

Listener-Driven Sound Change in Exemplar Theory

ROSLYN BURNS
Yale university

ABSTRACT

This paper presents a formal model to derive both gradient and categorical behaviors in listener-driven sound change. As a case study, I model the development of opposing acoustic trajectories of high vowels before labials in SE Bulgarian, where **i* > *u*, and Polabian, where **u* > *i*. By encoding different listener expectations into my model of labial co-articulation, I am able to derive both change, like SE Bulgarian and Polabian, and stability. In order to simulate the observed behaviors, I embed a Harmonic Grammar (Flemming 2001, Katz 2010, Burns 2021) in an Exemplar Theory (Pierrehumbert 2001) framework. My model suggests that while systems where interlocutors are aligned in their exceptions can result in sub-phonemic innovation, systems where they are misaligned result in both sub-phonemic innovation and eventual re-categorization in the preexisting phoneme inventory.

KEYWORDS listener driven sound change · exemplar theory · harmonic grammar

1 INTRODUCTION

This paper presents a formal model for deriving categorical innovation in *listener-driven sound change*, a process wherein listener errors bring about novel forms (Ohala 1981, 1989, 2012). Although formal models of sound change often lack implementation of listener behavior, Burns (2022a) highlighted the importance of doing so for certain types of innovation. Using data from Slavic (South Slavic and West Slavic), I show that the same phonetic environment can trigger seemingly opposing innovations, a behavior characteristic of listener-driven sound change.

In order to formally account for the facts of Slavic, I modify the gradient Harmonic Grammar (HG) framework (Burns 2022a), a fully gradient constraint-based framework (Flemming 2001, Katz 2010, Burns 2021), to include listener sensitivity to speaker behavior. This modification allows me to derive different behaviors in the same context depending on listener expectations of speaker behavior. One of the issues not dealt with by my model, presented in Burns (2022a), is how categories are structured during innovation as gradient HGs do not reference categorical relationships.

In this paper, I expand my previous model (Burns 2022a) by integrating a formal categorization component of listener-driven sound change within the Exemplar Theory (ET) framework (Pierrehumbert 2001). ET is a usage-based approach to categorization designed to show the emergent structure of categories. In this theory, real world stimuli (e.g., acoustic properties) are binned into discrete categories which over time update and shift based on one's experiences. While ET has not previously been used to model listener-driven sound change, in this paper I propose that by integrating HG with the ET model to account for listener expectations, we can derive outcomes which are consistent with the Slavic data.

This paper is structured as follows. §1.1 discusses the three different listener modes in listener-driven sound change and the previous account of what formally underlies these behaviors proposed in Burns (2022a). §2 reviews the Slavic data which necessitate the use of a listener-based approach. These data have the same triggering environment but opposing outcomes in different branches of Slavic. §3 presents the formal models used to derive listener-driven sound change. §3.1 presents the HG developed in Burns (2022a). §3.2 outlines the ET model used to derive categorization and how the HG is integrated with ET. §3.3 presents the parameters of the control and experimental simulations. §4 presents the findings of the integrated model and finally §5 concludes with the implications and areas for future investigation.

1.1 LISTENER MODES

In conversations, listeners must decode signals sent by speakers in order for communication to occur. This requires knowing which parts of the signal are intentional and which parts are unintentional perturbations caused by other factors (e.g., naturally occurring co-articulation). There are three possible ways that listeners might respond to unintentional acts: correction, hypocorrection, or hypercorrection. In correction, which does not result in innovation, the listener successfully parses speech signals into intentional and unintentional parts. For example, anticipatory nasalization is a natural consequence of nasals being co-articulated with the preceding vowel as in the English word /dim/ 'deem' which has the surface representation (SR) [dĩm]. English listeners are aware that the SR vowel nasalization is not intentional and are able to reconstruct the unintended sequence of [dĩm] to the underlying representation (UR) /dim/.

The other two listener responses result in innovation or listener-driven sound change. In the case of *hypocorrection*, the listener fails to notice influence from the surrounding environment in the signal, leading to the failure to factor out the phonetic influence of the environment. Over time these once unintentional parts of the SR become encoded as necessary components of the UR. For example in French, /ɡʁɑn(d)/ 'big' was produced with anticipatory nasalization as in [ɡʁɑ̃n(d)]. Over time, listeners attributed the presence of vowel nasalization to an intentional act on behalf of the speaker, rather than unintentional co-articulation, thus restructuring the UR of [ɡʁɑ̃n(d)] to /ɡʁɑ̃n(d)/.[1]

In *hypercorrection*, listeners identify something in the SR which they believe to be an artifact of co-articulation, when in fact, it was something which the speaker intended to produce. Because the listener assumes that the SR includes an unintentional signal, they attempt to filter it out of the UR, thereby removing an intentional property of the speech signal. The loss of nasal consonants after the phonologization of vowel nasality may be one such form of hypercorrection.[2] If the listener assumes that there is one source of nasality in a word like [ɡʁɑ̃n(d)] (< speaker's /ɡʁɑ̃n(d)/), and if they assume that source is the vowel, then they will interpret the consonant's nasality as a derived artifact of the nasal vowel. In response, the listener would remove the presumed derived segment(s) from the UR, via deletion in French, resulting in a novel UR: /ɡʁɑ̃/ 'big'.[3] In sum, hypocorrection results in assimilation, but hypercorrection results in dissimilation.

In Burns (2022a), I propose that correction responses may actually be gradient in nature based on the findings of previous laboratory research. Following from this, I seek to account for the three outcomes (no change, assimilation, and dissimilation) in a gradient

---

[1] This reinterpretation eventually paved the way for the development of contrastive vowel nasalization when the coda was lost and /ɡʁɑ̃/ 'big' then contrasted with /ɡʁɑ/ 'fat'.

[2] I use Hyman's (1975) definition of phonologization which means encoding of an otherwise phonetic process as an expected phonological representation (either derived or underlying) in the language (1975: 171).

[3] This view is consistent with the findings of Lahiri & Marslen-Wilson's (1999) experiment on nasal gating. They show that listeners tend to attribute nasality to only one source in Bengali, a language with both derived and contrastive nasalization. When Bengali listeners heard the CṼ portion of CṼN, they assumed that the following consonant is oral and not nasal, i.e., the source of nasalization is clearly the vowel, therefore the consonant must be oral. If Bengali listeners heard the CV portion of non-nasal CVC, they assumed that the following consonant could be nasal (but not necessarily, as orality is not derived).

The Bengali ṼN consonant repair differs from French. In French, the following consonant is deleted whereas in Bengali it is denasalized. Burns (2022b) treats both of these outcomes as forms of dissimilation where removal of a feature is partial dissimilation but deletion is effectively total dissimilation.

In French, even though the N(C) is deleted in the UR, it appears in the SR when licensed due to other considerations like hiatus avoidance (i.e., liaison).

fashion grounded in differences in speaker/listener attentiveness to co-articulation. I suggest that constraint weighting, the HG equivalent of classical Optimality Theory (OT) constraint ranking, can be used to represent sensitivity to producing or perceiving a behavior. Weighting relationships between speakers and listeners can account for all three modes if we assume that interlocutors are not bound to the same sensitivities, a property which I call *weighting autonomy*. If interlocutors have the same expectation of the signal (i.e., speaker weight = listener weight), over time the signal remains stable like in correction. When interlocutors are misaligned, innovation occurs. If speakers are more sensitive to the production of a perturbation than listeners are to filtering it out (i.e., speaker weights > listener weights), the result will be failure to remove contextual effects and the incorporation of those effects into the community's speech (i.e., assimilation). If, however, listeners are more sensitive to filtering out the effects of a perturbation than speakers are to actually producing it (i.e., speaker weights < listener weights), the result will be the removal of properties associated with the context (i.e., dissimilation).

Critically, deriving the three outcomes requires an interaction between both speakers and listeners suggesting that the term "listener-based" sound change is a misnomer. Nonetheless, the rest of this paper will continue to use this term as it is consistent with previous literature that highlights the role of listeners in innovative speech patterns (as opposed to just focusing on the role of speakers).

2 SLAVIC DATA

This section presents the Slavic data in Burns (2022a) which highlight the need to account for the listener's role in sound change. These data concern post-vocalic effects of labials on high vowels. Cross-linguistically, post-vocalic segments have been observed to exert greater co-articulatory influence on the neighboring vowels than pre-vocalic segments, especially if the segment is tautosyllabic (Stanton 2018, Walker & Proctor 2019, Burns 2021).[4] The fact that post-vocalic positions often exhibit high degrees of co-articulation with the preceding vowel suggests that this position may be the ideal target of misinterpretation of the degree of co-articulation of the sort outlined in §1.1.

It is often difficult to disentangle the role of speakers vs. listeners in sound change, so Burns (2022a) specifically identifies innovations involving the effect of labials on vowels which are necessarily perceptual in nature. While the lips are involved in the production of labials, the tongue is independently involved in the production of vowels. Activation of either articulator can result in the same acoustic outcome, which is known as an *acoustically convergent* behavior; either lip compression or tongue retraction can lower F2. When listeners receive signals with low F2, they have to decide whether the lips, the tongue, or both are responsible for the signal when building the UR.

Bulgarian (South Slavic) retains reflexes of inherited *i and *u as in *l**i**stъ 'leaf' > [l**ì**s] and *s**u**xъ(jь) 'dry' > [s**ù**x].[5] In the southeast (SE Bulgarian), the production of /i/ with concomitant lip constriction results in backing of the vowel, but this is not the case in other regions of Bulgarian as shown in Table 1.[6]

---

[4] Languages differ as to whether or not heterosyllabic post-vocalic segments exert a high degree of co-articulation on neighboring vowels. Stanton's (2018) investigation of nasalization across South America found that while coda nasals exhibited the greatest co-articulation on vowels, languages differed according to whether onset co-articulation strength prioritized membership in the same syllable (i.e., tautosyllabic pre-vocalic nasals) or post-vocalic position (i.e., heterosyllabic post-vocalic nasals). Walker & Proctor (2019) investigate a range of vowel neutralizations in English which are driven by the pharyngeal constriction in English coda rhotics. Vowel neutralization does not occur if the rhotic is in an onset (either tauto- or heterosyllabic) even though the pharyngeal gesture is still present. Burns (2021) finds an effect in Polabian where coda rhotics exhibit the greatest co-articulatory effect on neighboring back vowels, followed by heterosyllabic onsets, which exhibit a weaker co-articulatory effect, and finally tautosyllabic onset rhotics, which exhibit no co-articulatory effect.

[5] The Bulgarian forms in this paper are based on transcriptions of native speakers of Bulgarian found in Alexander & Zhobov (2016).

[6] While the data represented in Table 1 are based on Alexander & Zhobov (2016), an anonymous reviewer rightly points out that other Bulgarian sources have reflexes of the Turkish loan *çivi* 'nail' as *čivija*. This reflex

SE Bulgarian	Other Bulgarian	Source
[nəlʲùəʒ] ← /na-lij-(a)va-eʃ/	[nəlʲìvəʃʲ]	*l**i**ti ‘pour’
[ʒʲuwòt]	[ʒivòt]	*ž**iv**otъ ‘life’
[pupèr]	[pipèr]	*Greek* π**ιπ**ερι ‘pepper’
[t͡ʃʲuvìjə] ‘peg’	N/A	*Turkish* ç**iv**i ‘nail’

Table 1: Bulgarian /iP/ reflexes (Burns 2022a), Data from Alexander & Zhobov (2016).

Because the co-articulation of lip compression with a front vowel results in a back vowel reflex, this means that the F2 lowering of the lips was not filtered out by the listener and instead was interpreted to be an underlying property of the tongue position (i.e., assimilation/hypocorrection). Listeners of other varieties of Bulgarian (henceforth “Standard Bulgarian”), however, attribute the source of the low F2 to lip compression, not the vowel, and the UR remains stable (i.e., correction).

Lechitic languages (West Slavic) retain inherited *i and *u as in *listъ ‘leaf’ > Polabian [laɪst], Polish [l**i**st] and *s**u**xъ(jь) ‘dry’ > Polabian [saʊçə], Polish [suxɨ].[7] Note that in Polabian, reflexes of *i and *u have diphthongized into [aɪ] and [aʊ] respectively.[8] In Polabian, the production of /u/ in the context of a labial results in fronting of the vowel (u > i > aɪ), but this is not the case in Polish as shown in Table 2.

Polabian	Polish	Source
[laɪpaɪkɐ] ‘knackerman’	[wupit͡ɕ]	*l**up**iti ‘to peel/skin’
[glaɪpə]	[gwupi]	*gl**up**ъ ‘stupid’
[lʲaɪby] ‘kindly/dearly’	[lubɨ]	*lʲ**ub**ъ ‘sweet/pleasant’
[traɪvojə] ‘marry 3sg’	N/A	*Middle Low German* tr**ūw**en ‘to marry’

Table 2: Lechitic /uP/ reflexes (Burns 2022a), Data from Polański & Sehnert (1967).

Low F2 is an intrinsic property of back vowels like /u/ and lip compression. If listeners assume that the consonant is the source of low F2, and that low vowel F2 is derived, reconstructing the UR would require filtering out the presumed derived lowering on the vowel, causing it to front (i.e., dissimilation/hypercorrection). Listeners of Polish, however, attribute the source of low F2 to both the consonant and the vowel and the UR remains stable (i.e., correction). The shared thread in the data presented above is that listener interpretation of the source of low F2 in a signal determines whether there is innovation or failure to innovate. When innovation does occur, the same trigger (a labial consonant) can yield opposing behaviors (i.e., fronting vs. backing), because listeners can misattribute the source of low F2 in different ways.

3 FORMAL MODELS OF LISTENER DRIVEN SOUND CHANGE

The previous section reviewed data highlighting the need to incorporate listeners into formal models of sound change. In this section, I present the models which I use to capture both the gradient and categorical behavior of innovation. §3.1 reviews the HG which I present in Burns (2022a) to derive gradient SRs and URs of acoustic signals (F2) based on interlocutor expectations of co-articulation. This model is able to derive cor-

---

fits the broader pattern wherein historical /i/ is conservative before labial consonants in dialects other than southeastern Bulgarian.

[7] An anonymous reviewer points out that while a commonly cited Modern Polish reflex of LCS *listъ is [liɕt͡ɕ], in fact [list] is much older in its attestation. The older reflex eventually specialized to mean ‘letter’ whereas the newer reflex, which exhibits a later spread of palatalization to coronals, is now the basic term for leaf.

[8] See Polański (1993), Timberlake (1995), and Burns (2021) for discussion of high vowel diphthongization in Polabian.

rection (no change), hypocorrection (assimilation), and hypercorrection (dissimilation) based on the weighting relationships outlined in §1.1.

§3.2 focuses on the ET model aimed at capturing the categorical aspect of innovation wherein reanalyzed forms are both assigned to specific categories in the language user's mind and produced as representative members of these categories. In order to model the three behaviors associated with listeners, and to derive the specific reflexes described in §2, I integrate the HG into the production and processing procedures of the ET model. Finally §3.3 presents the parameters of the integrated model used in the current simulations.

3.1 THE HARMONIC GRAMMAR OF LISTENER-DRIVEN SOUND CHANGE

Harmonic Grammar (HG) is a form of OT with gradient constraint weighting instead of fixed ranked constraints. Some forms of HG are fully gradient in nature (Flemming 2001, Katz 2010), operating along a single phonetic dimension with gradient constraint targets, candidates, and assessment of violations. In gradient HG, candidate violations are calculated according to the equation in (1). In this equation, let C = cost, x = a candidate value, T = the target value of a constraint, w = the weight of a constraint, and n = the total number of constraints.

(1) $C(x) = w_1(x–T_1)^2 + w_2(x–T_2)^2\ldots + w_n(x–T_n)^2$

According to **??**, first, the distance between a candidate and constraint's target is calculated. That difference is then squared and multiplied by the constraint's weight. The weighted values of all constraint violations are then summed to give any given candidate's overall cost. The candidate with the lowest overall cost wins. Tableau in Table 3 shows a hypothetical input on an arbitrary phonetic scale and a set of candidates in increments of 25.

Input: 300	Const1 Target: +50 Weight: 1	Const2 Target: +15 Weight: 3	Const3 Target: −75 Weight: 2	Total Cost
a. 225	15625 $=(225-350)^2$	8100 $=(225-315)^2$	0 $=(225-225)^2$	39925 =(15625*1) +(8100*3) +(0*2)
b. 250	10000	4225	625	23925
c. 275	5625	1600	2500	15425
☞ d. 300	2500	225	5625	14425
e. 325	625	100	10000	20925
f. 350	0	1225	22500	48675

Table 3: OT Tableau for Sample HG Implementation

In Tableau in Table 3, each constraint has a different candidate that comes closest to satisfying its ideal target value. Const1 is best satisfied by Candidate (f), Const2 is best satisfied by Candidate (e) and Const3 is best satisfied by candidate (a). Despite this, none of these candidates win because the winner in HG is always the candidate which serves as the best compromise between the ideal targets of all constraints based on their relative weight. For this reason, Candidate (d), which appears to be the original 300 value, wins. As I will show below, with the equation in (2), 300 is not the actual winner, but the closest to it in Tableau in Table 3.

In gradient HG, it is common to solve for the winning candidate rather than modeling a set of candidates in a tableau. The optimal candidate can be identified by the following weighted average equation in (2).

(2) $$optimal x = \frac{\sum(w_1T_1, w_2T_2, w_nT_n)}{\sum(w_1, w_2, w_n)}$$

According to this equation, the mock constraints and input presented in Tableau 1 would actually have a winning value of 290.83 (= ∑(1*350, 3*315, 2*225)/∑(1, 3, 2)), not 300. Tableaux are actually better suited to modeling discrete candidates rather than gradient ones because candidates are discrete representations. Depending on the increments used to represent gradient candidates, the true winning candidate might not be shown in a tableau's candidate set as in our example above. Due to this consideration and space considerations, in the rest of this paper, I use the equation in (2) to derive winning candidates instead of modeling tableaux.

Gradient HG has been previously used to model innovation in Slavic by using acoustic data from Russian (East Slavic; Purcell 1979) as a proxy for phonetic input to the model (Burns 2021). In Burns (2021), acoustic targets of co-articulation constraints are defined formant intrinsically as the difference between the average vowel formant value in the context of a particular consonant type (e.g., palatalized) and the average vowel formant value across all consonant types. For example, the average vowel F2 in the context of palatalized consonants is 1959.03 hz but the average vowel F2 across all consonant types is 1732.68 hz. This means that on average, palatalized consonants raise a vowel's F2 by 226.35 hz. For this reason, the co-articulatory constraint governing the realization of palatalized consonant co-articulation on F2 is defined as follows in (3).

(3) Palatalize(F2): A post-vocalic palatalized consonant must raise the preceding vowel's F2.
Target: F2 + 226.35 hz

All consonant co-articulation constraints will interact with constraints governing vowel faithfulness as defined in (4).

(4) Ident(V): A vowel should not change from the input value.
Target = 0 Hz

If we only had the constraints in (3) and (4) above, a hypothetical input /VC^{j}/ input, where V's F2 is 800 hz, would identify the vowel output in [VC^{j}] as 950.9 hz if Ident(V) had a weight of 1 and Palatalize(F2) had a weight of 2 (=∑(1*800, 2*1026.35)/∑(1, 2)).

The behavior of co-articulation that I defined in Burns (2021) is expanded in my model of listener-driven sound change (Burns 2022a). In this model, which uses the same set of Russian data as a proxy for modeling labial co-articulation across the Slavic family, I define a new constraint governing labial co-articulation. Vowel F2 is on average 195.13 hz lower in the context of a labial than it is across all consonant environments; thus the constraint governing labial-coarticulation has the following definition in (5).

(5) Labial(F2) A post-vocalic labial consonant must lower the preceding vowel's F2.
Target: −195.13 hz

In Burns (2022a), I stipulate two properties of constraints. First, because my model is bi-directional, meaning that interlocutors take turns producing and decoding each other's speech, all listener constraint targets are the additive inverse of the speaker's. (6) shows the relationship of a speaker uttering a labial-coarticulated vowel with an input F2 of 1200 hz and a listener interpreting this vowel with two constraints: Ident(V) weight: 1 and Labial(F2) weight: 2.

(6)	Speaker UR: 1200 hz	Speaker SR: 1069.91 hz (= ∑(1*1200, 2*(1200-**195.13**))/∑(1, 2)
	Listener SR: 1069.91 hz	Listener UR: 1200 hz (=∑(1*1069.91, 2*(1069.91+**195.13**))/∑(1, 2))

The second feature of my model is that listener weights and speaker-weights are not necessarily the same; that is, there is weighting autonomy (see §1.1). A speaker can have more sensitivity to the production of a constraint than the listener has to decoding that constraint or vice versa. In this respect, if we take the same hypothetical 1200 hz input signal and continue to use the same constraints as in (6) for the speaker, but either halve or double the weight for the listener's co-articulatory constraints (either 1 or 4), we get the listener outcomes in (7).

(7)
Listener SR: 1069.91 hz   Half-Weight Listener UR: 1167.47 hz
$(=\sum(1*1069.91,\ \mathbf{1*1265.04})/\sum(1,2))$
Listener SR: 1069.91 hz   Double-Weight Listener UR: 1226.01 hz
$(=\sum(1*1069.91,\ \mathbf{4*1265.04})/\sum(1,2))$

As shown in (7), when listeners are half as sensitive to co-articulation as speakers, they decode a signal which is further back than the original 1200 hz. If, however, listeners are twice as sensitive to co-articulation than speakers, they decode a signal which is further forward than the original 1200 hz. That is to say that co-articulatory effects of the labial consonant become encoded in the vowel as either as an assimilatory or dissimilatory effect depending on the listener's sensitivity.

Another feature of my model is that it is multi-generational; the output of one generation becomes the input of the next. When repeated over 10 generations, the UR and production remain stable if interlocutors have matched expectations about the weight of labial co-articulation. If the expectations are not matched, however, over 10 generations the vowel will back to an underlying form of 874.77 hz if the listener consistently hypocorrects (i.e., speaker weight > listener weight) or the vowel will front to an underlying form of 1460.17 hz if the listener consistently hypercorrects (i.e., speaker weight < listener weight).

Although I express optimism about the potential of the model, I acknowledge that it is incomplete. Importantly, the model fails to handle the behavior of categorization when competing categories are present. Without this component, it is not clear how the innovative vowel representations derived by the HG would would interact with other preexisting sound representations in the memory. For instance, Alexander & Zhobov (2016) do not propose that SE Bulgarian /i/ was phonemically anything other than /u/, a preexisting category, in the innovation i > u / ___ P. This is despite the possibility that there may have been subphonemic innovation along the way to the new category. The next section introduces an ET framework to address this problem.

3.2 AN EXEMPLAR THEORY MODEL WITH PHONOLOGICAL PROCESSING

At its core, ET is a theory about the dynamic nature of one's memory and how categories themselves are dynamic and emergent (Goldringer 1992, Pierrehumbert 2001, Johnson 1997, 2006, 2007, Bybee 2002, 2006). Individual experiences, or *exemplars*, are stored in one's memory and are called upon when engaging specific actions. ET has two sides: production and perception. Crucially both processes are influenced by the decay of memories which is defined in (8). In this equation, let t = the current time, T_i = the time at which an experience entered the memory, and τ = the amount of time that any given memory stays in the mind.

(8) $$\frac{t - T_i}{\tau}$$

The equation in (8) is used in both production and perception to weight the influence of individual memories at specific times. In production, a target is randomly selected from the active memory and used to create a new target based on similar memories within the same category. Although selection of the target is random, the equation in (8) is used to build a time-weighted probability which makes newer memories more likely to be selected than older memories. After this, time-weighted distances of all active exemplars sharing the category of the randomly selected target are calculated as shown in (9). In this equation, let x = randomly selected memory, L = a mental category matching the category of x, e = an exemplar (individual memory).

(9) $$distance_i = \left|target_x - e_i^L\right| exp\left(\frac{t - T_i}{\tau}\right)$$

The distance score for each exemplar is a product of the phonetic distance (the absolute value difference between the random target and any given memory) and the memory distance (the exponent of the positive memory weight).[9]

When building a new production target, it is important that more than one memory contribute to execution of the task, a property known as *entrenchment*.[10] The entrenched memories which are selected are those with the lowest distance scores from the original randomly selected target. The novel production target ($target_{x'}$) is then built as the time-weighted average of the entrenched memories as shown in (10).[11] In this equation, let *ntrench* = the number of entrenched memories.

(10) $$target_{x'} = \frac{\sum_{i=1 \ldots ntrench} \left(e_i \times exp\left(-\frac{t - T_i}{\tau}\right)\right)}{\sum exp\left(-\frac{t - T_i}{\tau}\right)}$$

After building the novel production target, random production noise and non-random processes which bias production are then added to $target_{x'}$ as the utterance is made. While Pierrehumbert's (2001) version of ET models non-random processes grounded in cross-linguistic behavior during production (i.e., lenition based on Lindblom 1983), I propose that non-random processes can also involve the application of language specific phonological processes. For this reason, I apply the speaker's harmonic grammar to $target_{x'}$ when a labial is present and then add the production noise.

In ET perception, listeners receive a signal from the real-world and classify it based on their previous experiences. (11) shows the equation for classification where x = a real-world signal and W = a square window of acceptable closeness.

(11) $$score(L, x) = \sum_{i=1 \ldots n} W\left(x - e_i^L\right) exp\left(-\frac{t - T_i}{\tau}\right)$$

In this equation, the score of how close any given incoming signal is to members of a particular category (score (L,x)) determines how the incoming signal is classified in the memory. Each category score is determined by summing the time-weighted phonetic distance of all exemplars within a particular category that fall within an acceptable range of the incoming signal (if the memories fall within an acceptable range, the W multiplier is 1; otherwise it is 0). The category with the highest score (meaning the most temporally relevant exemplars that are close in phonetic distance) is the winning classification assigned to the new memory.

There are two components of perception not included in Pierrehumbert (2001) which are in my model. First, I add random noise during perception to represent the listener correcting for random fluctuations that they expect to occur during production/trans-

---

[9] The exponent of the positive memory weight results in larger memory distance multipliers for older memories, thereby contributing to a greater overall distance score.

[10] Entrenchment allows ET models to exhibit the behavioral property where practice reinforces behavior (Pierrehumbert 2001). The entrenchment number is assigned by the researcher and is supposed to reflect that when neurons fire during production planning, a region of one's memory is activated rather than just a single memory.

[11] The exponent of the negative memory weight results in numbers ranging from 1 to 0, where older memories are closer to zero and therefore contribute less to the memory weighted average.

mission of the signal. Second, and more theoretically important, is the addition of context-dependent processing. I add a filtering phase to perception where the listener's HG applies to the incoming signal when a labial is present. This is an important departure from the previous literature on ET because the type of perceptual processing shown in (11) is *veridical processing* wherein a perceiver receives a signal from the real world and categorizes it directly without additional subjective measures about the potential categories involved. In order to integrate HG with ET, we have to assume that there are some parts of perception which can be *non-veridical* in nature; that is, the listener makes assumptions about the signal that they receive and pre-filters the signal before categorization occurs. I also add a parameter in perception in which categorization of any signal only involves comparison between the incoming signal and core members of a set; that is, exemplars which are not specified for the labial context.[12]

One of the advantages of using ET with random variation in both production and perception, alongside the HGs differing in speaker-listener sensitivity, is that we can get a sense of what, if anything, the different listener sensitivity models contribute to the development of categories over time. If controls, where speaker weight = listener weight, exhibit the same innovative behaviors as experimental trials, where speaker weights ≠ listener weight, then we can assume that differences in speaker-listener sensitivity do not contribute more to the integrated model than random variation in perception/production.

3.3 INTEGRATED MODEL PARAMETERS

I created code in R to test the behavior of the integrated model with the following parameters. Four conditions were built: two controls and two experimental. All conditions begin with three exemplars and two categories. The first two exemplars are non-HG members of the set, /i/ (3500 hz) and /u/ (800 hz). One additional exemplar specifying the labial environment (P) was added to each condition depending on the model. For the Standard Bulgarian and SE Bulgarian models, the starting P exemplar was a member of the /i/ category whereas for the Polish and Polabian models, the starting P exemplar was a member of the /u/ category. The starting frequency of P matched the other member of the set. While all four languages have labial contexts for both /i/ and /u/, different languages exhibit sensitivity to different VP pairings as outlined in §2. For this reason, I only model the VP pairings which are subject to change in each language group and assume that listeners have no difficulty correcting for the other VP context. All additional exemplars beside the initial three are built as the model runs.

The controls are designed to reflect the Standard Bulgarian and Polish behavior, where even though the speaker applies labialization whenever P is specified, the listener is able to correct for the effects of contextual labialization (speaker weight: 2 = listener weight: 2). In the experimental conditions, the speaker and listener weights do not align. In the SE Bulgarian example, listeners assume that labialization is much lighter than it is (speaker weight: 2 > listener weight: 1) and in the Polabian model, listeners assume that labialization is much heavier than it is (speaker weight: 2 < listener weight: 4).

Each condition was run for a total of 10,000 speaker-listener turns with an exemplar decay rate of 500 turns, an entrenchment value of 500, random noise from 0 to ±97.56 hz (half the target of the co-articulatory constraint LABIAL(F2)), and a perception window of 1350 hz (half the distance between the original category values).[13] All conditions alternated between the phoneme /i/ and /u/ for each successive speaker-listener turn. In

---

[12] Although ET is supposed to theoretically be an alternative to Prototype Theory (in which there are core and non-core members of a set), without stipulating that only contextless exemplars are used for category comparison, the model exhibits behavior in which the labial context exemplars of a set shift and eventually collapse when re-categorization occurs.

[13] Pierrehumbert (2001) uses an entrenchment value of 500 and a decay rate of 2,000. I use a decay rate of 500 which still allows me to demonstrate how classification changes over time without needing to increase the overall quantity of speaker-listener turns and overall compute time. Increasing the decay rate does not change the overall behavior of what happens during the process of classification; rather it only slows down the rate at which reclassification happens.

addition to assuming that both /i/ and /u/ have a equal frequency of use, I assume that labials occur post-vocalically with equal frequency as coronals, dorsals, and zero (i.e., word-final). In 1 out of every 4 turns, the program checks to see if a segment with P has been selected, and if so, applies the HG.

During production, if the listener selects a production target with P, they activate other memories of the same category and context to build a new production target to which the speaker HG applies. During perception, if the listener receives a signal with P, the listener HG first removes the contextual properties of P prior to classifying the modified signal using the structure outlined in (11). As stated in §3.2, classification involves comparing the incoming signal to only the core members of the set, not the P members of the set.

4 SIMULATION RESULTS

This section presents the results of the integrated HG/ET model. §4.1 analyzes the results of the two controls where interlocutor co-articulation weights are equal. §4.2 analyzes the results of the two experimental trials where the interlocutor co-articulation weights differ. Finally, §4.3 summarizes the findings.

Throughout this section figures represent the degree of reinforcement (y-axis) of specific instances of F2 production/perception (x-axis). In these figures, taller narrower density peaks represent more stable and reinforced behaviors within a specified period of time (i.e. linguistic stability). Shorter wider density peaks, on the other hand, represent less stable behavior over a specific period of time (i.e. linguistic change).

Additionally, I standardize representation of phonemes and allophones with line color and line type. Red lines represent members of the phoneme /i/ whereas blue lines represent /u/. Solid lines represent members of a phoneme category that were not subject to the operation of the HG (i.e., core members), whereas dashed lines represent members of the phoneme category subject to the HG (i.e., the P allophone).

4.1 CONTROL SIMULATION RESULTS

Figure 1 shows the behavior of Standard Bulgarian correction. Figure 1a to the left shows production and Figure 1b to the right shows perception.

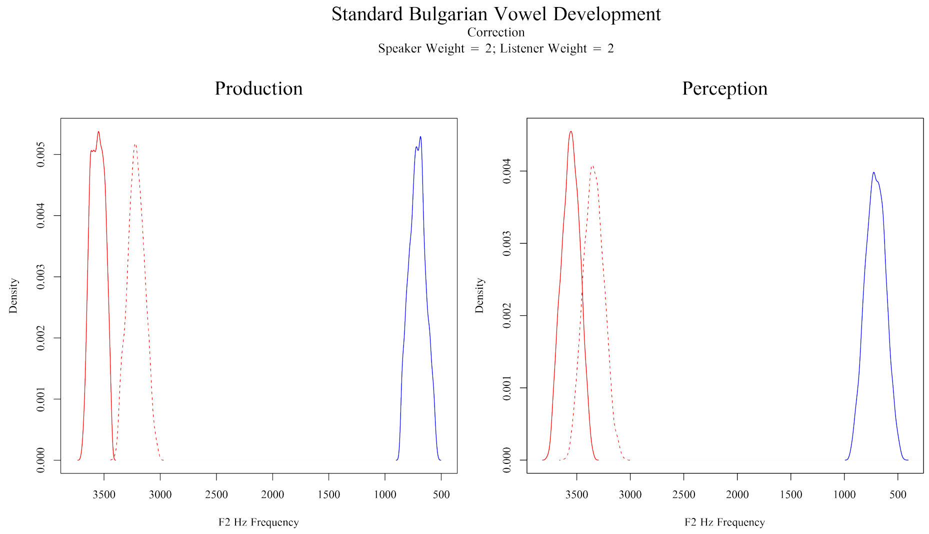

Figure 1: Standard Bulgarian: a. (left) Production; b. (right) Perception

As can be seen in Figure 1, Standard Bulgarian exhibits stable and reinforced categories of /i/ and /u/. Although the HG results in backing of i / ___ P, both production and perception of the P allophone remain close to the core members of /i/ set. Notably, perception of iP is more closely aligned to the core members of the set than production of iP (i.e., production is leading the change in the category).

Figure 2 shows the behavior of Polish correction. Figure 2a to the left shows production and Figure 2b on the right shows perception.

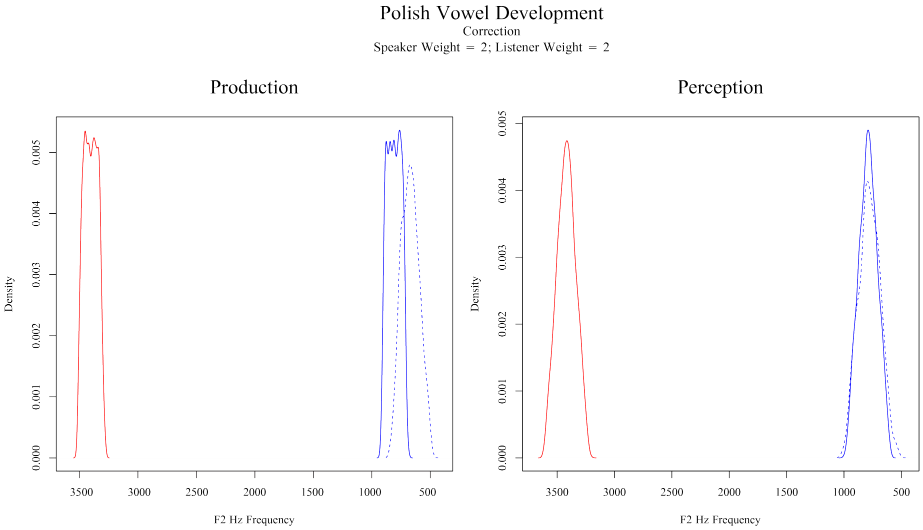

Figure 2: Polish: a. (left) Production; b. (right) Perception

As in Figure 1, Figure 2 exhibits stable and reinforced categories of /i/ and /u/. Although the HG results in the production of a backed allophone of u / ___ P, there is almost no change in the degree of perceptual overlap between the core members of the category and the uP allophones.

During different test runs of the correction model, it was most often the case that allophones developed in the acoustic direction of co-articulation. There were, however, instances where the inverse direction developed due to random variation, especially when the core members of the class retracted more than the allophone. In these instances, it was still the case that the allophones themselves showed a high amount of overlap with core members of the category in perception and production.

4.2 EXPERIMENTAL SIMULATION RESULTS

There were two experimental conditions, one where the listener was half as sensitive to co-articulation as the speaker (SE Bulgarian hypocorrection), and one where the listener was twice as sensitive to co-articulation than the speaker (Polabian hypercorrection). Figure 3 shows the behavior of SE Bulgarian hypocorrection where 3a on the left shows production and 3b on the right shows perception.

In Figure 3, while the core members of each allophone set exhibit stability and reinforcement, this is not the case for the P allophone. Hypocorrection leads to backing of the vowel and eventual re-categorization of i > u / ___ P. Unlike the Standard Bulgarian correction simulation (shown in Figure 1), where random variation also results in backing of the vowel, in the case where interlocutor expectations are misaligned, we find a rapid shift towards a new category.

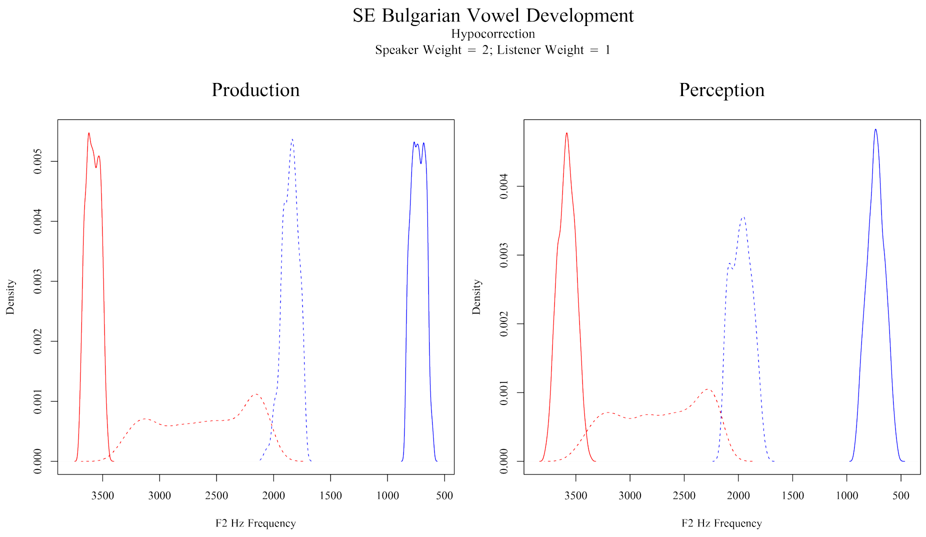

Figure 3: SE Bulgarian: a. (left) Production; b. (right) Perception

Figure 4 shows the behavior of Polabian hypercorrection where 4a on the left shows production and 4b on the right shows perception.

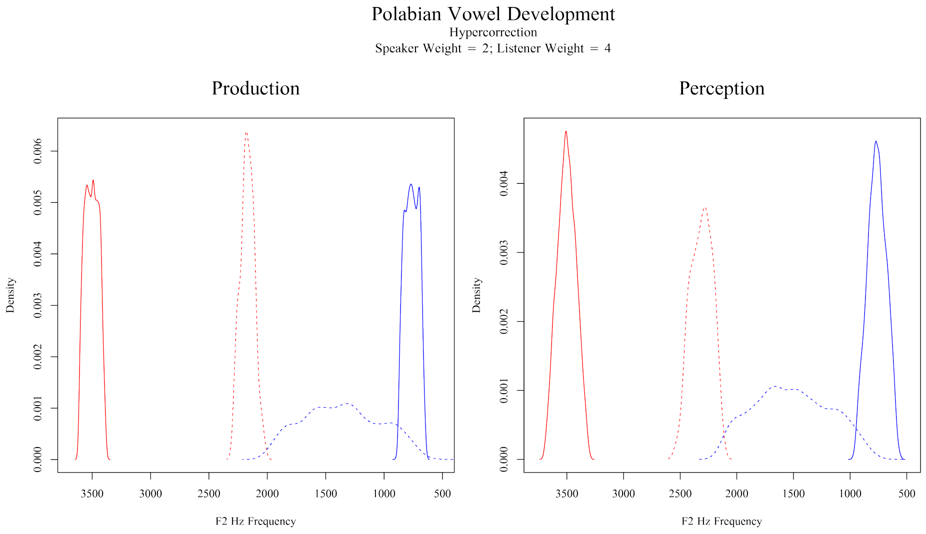

Figure 4: Polabian: a. (left) Production; b. (right) Perception

Similar to Figures 1–3, Figure 4 exhibits stable and reinforced core members of each phoneme. Hypercorrection, however, leads to fronting of the vowel and eventual recategorizaiton of u > i ___ P. Although the Polish correction model standards can also result in simulations with a fronted P, when interlocutor expectations are misaligned, there is a quick movement towards a new category which does not occur in correction.

Across the two experimental simulations, there are differences in how innovative production is adopted. In the hypocorrection model, changes in production precede changes in perception, but in the hypercorrection model, changes in perception precede changes in production. This differs from the correction controls simulations where changes in production precede changes in perception.

4.3 SUMMARY

As this section has shown, it is possible to have innovation in both systems where interlocutor expectations are matched and in systems where they are not. The innovations are, however, different in nature. In systems where expectations are matched, innovations are due to random variation in production and tend to stay close to the core members of a category. That is to say, unique allophones develop but they need a lot of time to become perceptually distinct from other members of the same set.

In systems where interlocutor expectations are not matched, allophones quickly diverge from core members of the set and ultimately converge towards other categories. In systems where speakers are more sensitive to co-articulation than listeners (speaker weights > listener weights), the direction of both allophonic innovation and eventual phonemic reassignment follow the acoustic direction of co-articulation which indicates that these systems are hypocorrective. In hypocorrective systems changes in production precede changes in perception. In systems where listeners are more sensitive to decoding co-articulation than speakers (speaker weights < listener weights), the direction of both allophonic innovation and eventual phonemic reassignment oppose the acoustic direction of co-articulation which indicates that these systems are hypercorrective. In these systems, changes in perception precede changes in production.

5 CLOSING

In this paper, I presented a model capable of capturing both gradient and categorical properties of listener-based sound change, a type of innovation which necessarily references both speaker and listener behavior. As shown in §1, one of the listener modes, correction, results in stable systems, but the other two modes lead to innovation. These innovations can be of the type where the same phonetic environment can trigger opposing acoustic development such as encoding assimilatory behavior into URs in the process of hypocorrection or the removal of a shared feature from the UR in hypercorrection. In §2, I laid out data from Slavic which suggests that correction, hypocorrection, and hypercorrection have all occurred in different Slavic languages with the shared phonetic environment of vowels before a labial consonant. In this particular set of innovations, it is necessary to discuss the role of the listener because two independent articulators, the tongue and the lips, are involved in the production of low F2. When innovation occurs, it is either because listeners fail to attribute the vowel's low F2 to the lips thereby encoding the low F2 into the vowel's UR (e.g., SE Bulgarian hypocorrection *i > u / ___ P) or over attribute the vowel's low F2 to the lips thereby removing the low F2 from the vowel's UR (e.g. Polabian hypercorrection *u > i / ___ P). In both cases, the position of the tongue is adjusted instead of the position of the lips.

In my formal model presented in §3, I integrate both gradient and categorical behaviors into processing. In the HG, both speakers and listeners apply gradient co-articulatory processes apply in certain phonological contexts. While in some cases listeners and speakers may be aligned in their expectations of co-articulation (resulting in correction), in other cases, their expectations may be misaligned, which results in either hypocorrection (if the speaker is more sensitive to co-articulation than a listener is to parsing it) or hypercorrection (if the listener is more sensitive to co-articulation than a speaker is to producing it). The gradient phonological system (HG) is integrated into a broader system of memory storage and recall (ET) which captures both gradient and categorical behaviors of sound systems. In the ET model, listeners store category-specific memories based on phonetic and temporal proximity to other memories and speakers produce category-specific utterances based on phonetic and temporal proximity to other recent memories.

Crucially, my model shows that when speaker-listener expectations are matched in the HG, the ET model can derive innovation, but it is always sub-phonemic in nature and lies close to the core members of the phoneme set. When speaker-listener expectations are not matched, however, there is innovation involving both sub-phonemic shift and eventual reclassification of allophones to a new phoneme. The direction of sub-phonemic shift and eventual reclassification depends on how the listener misinterprets the speaker's signal. If the listener underestimates the extent to which a speaker applied co-articulation, the acoustic change follows the acoustic direction of the process (i.e., hypocorrection). If, however, the listener overestimates the extent to which a speaker applied co-articulation, the acoustic change opposes the acoustic direction of the process (i.e., hypercorrection).

The findings of how speaker-listener expectations impact innovation in my model have broader implications beyond Slavic. One major implication of my model is that when we build models which are underspecified for listener behavior, we may potentially be committing ourselves to a particular listener mode, usually either correction (in the case of stable systems) or hypocorrection (frequently found in accounts of innovation). For example, while Pierrehumbert (2001) treats her model of innovative lenition as production-based in nature, the fact that the innovative production-based behavior becomes incorporated into the grammar means that listeners failed to correct for production, i.e., hypocorrection.

In terms of the potential range of phenomena that this type of model could explain, the integrated gradient and categorical system need not just apply to the effect that consonants have on vowels, but could also involve the effect that vowels have on the interpretation of consonants. While it is common to find hypocorrection of the sort n > ɲ / ___ i cross-linguistically (Bateman 2007), within southern Bantu, there are cases of hypercorrection leading to the same outcome with m > ɲ / ___ u (Doke 1954, Kotzé & Zerbian 2008, Braver & Bennett 2015). The current model can capture many of these behaviors, except that the HG constraints would need to reference how speakers and listeners interpret consonant place, which neither Flemming (2001) nor Burns (2021) address in their models.

There are still a variety of behaviors observed during innovation which remain unexplained by the model, but are also unexplained in terms of the experimental literature. For example, there is an asymmetry between the occurrence of hypocorrection and hypercorrection where the former is more common than the latter. In my model, it takes longer to achieve categorical reclassification in hypercorrection than in hypocorrection, but this behavior is predicted by the HG constraint weighting relationship alone. It is not clear if we should leave the constraint weights alone to explain this asymmetry or if there are other factors which need to be incorporated into the model. Second, it is known that misalignment tends to undergo atrophy over time (Harrington 2012). This suggests that there should be a mechanism that allows the weights themselves to be updated as time goes on, but identifying how the model should implement this relies in part on learning more about this type of feedback loop in language processing itself.

ACKNOWLEDGEMENTS

I would like to thank Darya Kavitskaya, Jason Shaw, Claire Bowen, Natalie Webber, Chelsea Sanker, and the Yale Phonologroup for their invaluable feedback on this work.

CONTACT

Roslyn Burns — *roslyn.burns@yale.edu*

ABBREVIATIONS

ET	Exemplar Theory	OT	Optimality Theory
HG	Harmonic Grammar	SR	Surface Representation
LCS	Late Common Slavic	UR	Underlying Representation

REFERENCES

Alexander, Ronelle & Vladimir Zhobov. 2016. Bulgarian Dialectology as Living Tradition. Website. http://www.bulgariandialectology.org.

Bateman, Nicoleta. 2007. *A crosslinguistic investigation of palatalization*. San Diego: University of California San Diego dissertation.

Braver, Aaron & William Bennett. 2015. The productivity of 'unnatural' labial palatalization in Xhosa. *Nordlyd* 42. 33–44.

Burns, Roslyn. 2021. Modeling gradient processes in Polabian vowel chain shifting and blocking. *Journal of historical linguistics* 11(1). 102–142.

Burns, Roslyn. 2022a. Deriving listener-driven sound change in a Harmonic Grammar. International Conference on Historical Linguistics 25, Oxford, UK.

Burns, Roslyn. 2022b. Persistent rules as reanalysis and extension. *Diachronica* 39(1). 39–87.

Bybee, Joan. 2002. Phonological evidence for exemplar storage of multiword sequences. *Studies in second language acquisition* 24. 215–221.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733.

Doke, C.M. 1954. *The southern Bantu languages*. Oxford: Oxford University Press.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18(1). 7–44.

Goldringer, Stephen. 1992. *Words and voices: Implicit and explicit memory for spoken words*. Bloomington: Indiana University dissertation.

Harrington, Jonathan. 2012. The coarticulatory basis of diachronic high back vowel fronting. In Maria-Josep Solé & Daniel Recasens (eds.), *The initiation of sound change: perception, production and social factors*, 103–122. Amsterdam: John Benjamins.

Hyman, Larry. 1975. *Phonological theory and analysis*. New York: Holt, Rinehart & Winston.

Johnson, Keith. 1997. The auditory/perceptual basis for speech segmentation. *OSU working papers in linguistics* 50. 101–113.

Johnson, Keith. 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of phonetics* 34(4). 485–499.

Johnson, Keith. 2007. Decisions and mechanisms in exemplar-based phonology. In Maria-Josep Sole, Patrice Speeter Beddor & Manjari Ohala (eds.), *Experimental approaches to phonology*, 25–40. Oxford: Oxford University Press.

Katz, Jonah. 2010. *Compression effects, perceptual asymmetries, and the grammar of timing*. Amherst: University of Massachusetts, Amherst dissertation.

Kotzé, Albert & Sabine Zerbian. 2008. On the trigger of palatalization in the Sotho languages. *Journal of African languages and linguistics* 29(1). 1–28.

Lahiri, Aditi & William Marslen-Wilson. 1999. The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38. 245–294.

Lindblom, Björn. 1983. Economy of speech gestures. In Peter MacNeilage (ed.), *The production of speech*, 217–245. New York: Springer.

Ohala, John J. 1981. The listener as a source of sound change. In *Proceedings of the Chicago Linguistics Society 17*, 178–203. Chicago, IL.

Ohala, John J. 1989. Sound change is drawn from a pool of synchronic variation. In Leiv Egil Breivik & Ernst Hakon Jahr (eds.), *Language change: Contributions to the study of its causes*, 173–198. Berlin: Mouton de Gruyter.

Ohala, John J. 2012. The listener as the source of sound change: An update. In Maria-Josep Solé & Daniel Recasens (eds.), *The initiation of sound change: Perception, production, and social factors*, 21–35. Amsterdam: John Benjamins.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (eds.), *Frequency effects and Emergent Grammar*, 137–157. Amsterdam: John Benjamins.

Polański, Kazimierz. 1993. Polabian. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic languages*, 795–824. London, New York: Routledge Press.

Polański, Kazimierz & James Allen Sehnert. 1967. *Polabian-English dictionary*. The Hauge: Mouton.

Purcell, Edward. 1979. Formant frequency patterns in Russian VCV utterances. *Journal of the Acoustical Society of America* 66(6). 1691–1702.

Stanton, Juliet. 2018. Environmental shielding is contrast preservation. *Phonology* 35(1). 39–78.

Timberlake, Alan. 1995. Mechanisms and relative chronology of Polabian sound changes. *Wiener Slawistischer Almanach* 35. 281–296.

Walker, Rachel & Michael Proctor. 2019. The organisation and structure of rhotics in American English rhymes. *Phonology* 36(3). 457–495.