

A Formal Model of Phonological Typology

Gašper Beguš

1. Introduction

One of the most contested debates in phonology concerns identifying factors that affect typology. Two lines of thought emerge in this discussion: Analytic Bias (AB) and Channel Bias (CB) approach (Moreton, 2008). The AB approach claims that cognitive biases in learning influence the typology, while the CB approach assumes phonetic precursors and transmission of language affect the typology (Moreton, 2008).

Both approaches are supported by empirical evidence. Phonological processes that are typologically rare have been shown to be more difficult to learn or their learning requires more input data (e.g. Wilson, 2006; White, 2014, and many others; for a survey, see Moreton & Pater, 2012a,b). On the other hand, processes that result from phonologization of phonetically motivated sound changes with clear phonetic precursors are also typologically more frequent (Blevins, 2004).

While an increasing body of work acknowledges that both AB and CB influence the typology (Moreton, 2008), very few attempts have been made to model the two approaches together: research either focuses on one or the other factor (Blevins, 2004; Kiparsky, 2006; Wilson, 2006; White, 2017, i.a.). This paper aims to fill this gap and proposes a model of phonological typology that models both AB and CB influences together. I argue that this unified model performs better in deriving the surface typology and has an advantage over the current “split” models in that it encodes not only that some processes are rare and others more frequent, but also why such typological differences arise.

The first step in the direction of the new typological model are quantitative models of the two sub-components: AB and CB. While several attempts have been made to quantify and model learning of phonology (AB; Wilson, 2006; White, 2017), quantitative approaches to CB have received less attention. The most recent attempt in Cathcart (2015) is computationally too demanding and fails to provide implementable quantitative results. In the first part of this paper, I present and expand on a new model of typology within the CB approach that introduces the concept of *Historical Probabilities of Alternations* (P_χ) (from Beguš, 2016). In the second part, I propose a model of typology that admits both CB and AB and employs the Maximum entropy probability distribution over candidates (Goldwater & Johnson, 2003) to derive typological probabilities. The new model of typology combines Historical Probabilities of Alternations introduced here with the learnability metric adapted from Wilson (2006) and unifies them in an MaxEnt-compatible framework.

2. Typology within CB

2.1. Problems

One of the most widely discussed objections to the CB approach is raised in Kiparsky (2006, 2008). Kiparsky (2006) invokes final voicing as a process that is never attested as part of a productive synchronic grammar (cf. Yu, 2004), yet he identifies at least five diachronic scenarios that would yield final voicing. He concludes that CB is not capable of deriving this systematic gap in typology, which means that AB has to be responsible for it.

It is true that current models of typology within CB are insufficient. The most common line of thought in deriving the typology within CB has been to assume that rare sound changes produce rare alternations (Blevins, 2004). Moreton (2008) attempts to quantify phonetic precursors with the goal of

* Gašper Beguš, Harvard University, begus@fas.harvard.edu. I would like to thank Kevin Ryan, Jay Jasanoff, Adam Albright, and the audiences at AMP 2017, phoNE 2017, and the 35th WCCFL. All mistakes are my own. This research is supported by Mind Brain Behavior Intefaculty Initiative at Harvard University.

reaching a more transparent phonetic metric for disambiguating sound changes, but this approach has problems, too (Yu, 2011). As will be shown in this paper, the results of sound changes operating in combination and phonetic precursors do not always align, which makes the quantification of phonetic precursors unsuitable for deriving the typology of all processes (including the unnatural ones). Finally, Cathcart (2015) attempts to quantify the CB influences on typology by identifying the number of combinations of sound changes that produce an unnatural alternation or phonotactic restriction such as final voicing. The number for each combination of sound changes is then compared to the number of all sound changes given a number permutation and a sample of sound changes. The problem with this approach is that it is computationally demanding and does not provide outputs that could be used for a typological model. The model in Cathcart (2015) also fails to distinguish alternations from phonotactic restrictions and does not establish the minimal number of sound changes required for an unnatural process to arise (see 2.2).

In the rest of this section, I present and expand on a new model of typology that estimates Historical Probabilities of Alternations (P_{χ}) based on a method that I call *Bootstrapping Sound Changes* (BSC) (Beguš, 2016). This method outputs results directly implementable in a MaxEnt model of phonological typology, proposed in section 4.

2.2. Minimal Sound Change Requirement

Before we turn to estimating Historical Probabilities of Alternations (P_{χ}), some key concepts, developed in Beguš (2016), need to be clarified.

First, Beguš (2016) proposes a new subdivision of natural and unnatural processes. I argue that what has traditionally been labeled as “unnatural” should be further divided into “unmotivated” and “unnatural”. In other words, natural processes are those that operate in line with Universal Phonetic Tendencies (UPT). UPTs are defined as universal phonetic pressures that are typologically common, phonetically motivated, and operate passively across languages (for a detailed definition, see Beguš, 2016; Beguš & Nazarov, 2017). *Unmotivated* processes are those that lack phonetic motivation. *Unnatural* processes not only lack phonetic motivation, but also operate directly against some UPT. For example, final devoicing, post-nasal voicing, and intervocalic voicing are UPTs; final voicing, post-nasal devoicing, and intervocalic devoicing are unnatural processes, because they operate against these UPTs.

Table 1 lists unnatural processes and the languages in which they appear which were identified by surveys of unnatural alternations and sound changes (Beguš, 2016) and unnatural phonotactic restrictions (Beguš & Nazarov, 2017) that target the feature [\pm voice].

Table 1: Unnatural processes and languages in which they appear.

Unnatural alternations & sound changes		
Post-nasal devoicing	Yagnobi	Xromov (1972, 1987)
	Tswana and Shekgalagari	Solé et al. (2010)
	Makhuwa and Bube	Janson (1991/1992)
		Janssens (1993)
	Konyagi	Merrill (2016a,b)
	Sicilian and Calabrian (south Italian dial.)	Rohlf (1949)
Murik, Buginese, and Land Dayak (Austronesian)	Blust (2013)	
Unnatural phonotactics		
Intervocalic devoicing	Berawan (and Kiput)	Blust (2005, 2013)
Post-obstruent stop voicing	Tarma Quechua	Adelaar (1977)

A diachronic treatment of post-nasal devoicing in all the languages identified in Table 1 reveals a common pattern: in all cases, post-nasal devoicing results from a combination of three natural sound changes. More precisely, all unnatural alternations undergo what I term a *Blurring Process* (in Beguš, 2016):

(1) *Blurring Process*

- a. A set of segments enters complementary distribution.
- b. A sound change occurs that operates on the changed/unchanged subset of those segments.
- c. Another sound change occurs that blurs the original complementary distribution.

Depending on whether the second sound change targets the changed or unchanged subset of segments, I identify two diachronic developments within the Blurring Process: *Blurring Cycle* (unchanged) and *Blurring Chain* (changed) (from Beguš, 2016). Let $A > B / X$ represent a natural sound change. The three sound changes of the Blurring Cycle and the Blurring Chain and the resulting unnatural sound change $B > A / X$ are schematized in (2).

(2) *Two subtypes of the Blurring Process*

<i>Blurring Cycle</i>	<i>Blurring Chain</i>
$B > C / -X$	$B > C / X$
$B > A$	$C > D$
$C > B$	$D > A$
<hr/>	<hr/>
$B > A / X$	$B > A / X$

As is argued in Beguš (2016) and Beguš & Nazarov (2017), all cases of post-nasal devoicing in Table 1 result from the Blurring Cycle; the Blurring Chain on the other hand, yields the phonotactic restriction against intervocalic voiced stops in Berawan and the restriction against clusters that agree in voicing in Tarma Quechua.¹

The Blurring Process further allows us to establish the Minimal Sound Change Requirement (MSCR; Beguš, 2016). Not only does the typological study of unnatural alternations show that they always arise through a combination of three natural sound changes, Beguš (2016) also provides a formal proof that unnatural alternations always require a minimum of three sound changes.

Let us define a sound change as a change in one feature in a given environment (Picard, 1994). We assume that a sound change can only operate in a phonetically natural direction (Garrett, 2014, *pace* Blust, 2005). This means that, if $A > B / X$ is a natural process (a UPT), $B > A / X$ cannot operate as a single sound change. We could imagine a scenario in which two sound changes would produce an unnatural result: $B > C / X$ and $C > A$. Note, however, that C and A in the second hypothetical sound change ($C > A$) differ in two features: ϕ_1 that distinguishes A and B and ϕ_2 that distinguishes B and C and $\phi_1 \neq \phi_2$. Because sound change is defined as a change in one feature in a given environment (Picard, 1994), we cannot get $C > A$ (where A and C differ in two features) with a single sound change; therefore, unnatural alternations can only arise through a combination of at least three sound changes. This requirement is summarized in ((3)).

(3) *Minimal Sound Change Requirement* (MSCR; Beguš, 2016)

Natural processes arise through a single sound change. A minimum of two sound changes have to operate in combination for an unmotivated process to arise. A minimum of three sound changes have to operate in combination for an unnatural process to arise.

The Blurring Process and MSCR provide grounds for deriving typology within the CB. I propose that for each synchronic alternation (A), we can calculate its Historical Probability (P_x) based on the number of sound changes (S) required for the alternation A to arise and their respective probabilities (Beguš, 2016).

The MSCR already predicts that natural alternations are more frequent than unmotivated ones, which in turn are more frequent than unnatural alternations by virtue of the number of sound changes they require: the probability of a single sound change occurring is higher than the probability of three sound changes occurring in combination, all else being equal.

This distribution of probabilities of alternations according to the MSCR, however, does not yet

¹ That *telescoping* (Wang, 1968) — or, in other words, combinations of sound changes — can produce unmotivated results has long been known. In this paper, I propose that for unnatural alternations we need a special type of combination of sound changes: the Blurring Process.

provide quantitative means for our typological model. Historical Probabilities of Alternations are estimated not only based on the number of sound changes, but also based on their respective probabilities. Estimating probabilities of sound changes, however, is not a trivial task.

Beguš (2016) outlines a method for estimating probabilities of alternations using the statistical technique of bootstrapping (Efron, 1979) and surveys of sound changes. We can estimate probabilities of individual sound changes by comparing the number of languages with a sound change S_1 and the number of all languages in a given survey.

(4)

$$P_{\chi}(S_1) = \frac{\text{number of languages with sound change } S_1}{\text{number of languages surveyed}}$$

The probability $P_{\chi}(S_1)$ equals the historical probability of an alternation A_1 when A_1 requires only S_1 to arise. If, on the other hand, alternation A_1 requires more than one sound change, we estimate the joint probabilities of each sound change required, dividing by $n!$ if the ordering of sound changes matters.

(5)

$$P(A_1) = \frac{P(S_1)P(S_2) \cdot \dots \cdot P(S_n)}{n!}$$

This method of estimating Historical Probabilities is called *Bootstrapping Sound Changes* (BSC) and requires some crucial assumptions. First, samples of sound changes (surveys) are ideally well-balanced and representative. Second, we assume that sound changes in combination operate independently of each other. This is not problematic: there is no reason to believe that the operation of one sound change affects the probability of the operation of another sound change. More problematic is the additional assumption that sound changes are independent of the synchronic phonemic inventories on which they operate. The dependency of sound change and phonemic inventories is at least to some degree captured by the fact that we always estimate probabilities of sound changes in a given environment. For practical purposes, we can currently disregard dependency of sound change and phonemic environment until more comprehensive surveys of sound changes are available.²

There are several advantages of BSC that, to my knowledge, are not available under any other quantitative approach to typology within CB. We can now (i) estimate the Historical Probability of any synchronic alternation (even unattested alternations), (ii) compare the Historical Probabilities of two or more alternations and perform inferential statistics on the comparison, (iii) predict whether an alternation is expected to be attested in a given sample, and (iv) identify historically equiprobable alternations. Finally, estimated Historical Probabilities can be directly implemented in a typological model within the MaxEnt framework.

2.3. An example

Let us take as an example the natural process of post-nasal voicing (PNV) and its unnatural counterpart post-nasal devoicing (PND). In a survey of consonantal sound changes in Kümmel (2007), there are approximately 41 languages (out of approximately 200 languages surveyed in total) in which PNV operated as a sound change. On the other hand, there are 56 languages with intervocalic fricativization of voiced stops, 11 with intervocalic devoicing of voiced stops, and 37 languages with intervocalic occlusion of voiceless fricatives to stops (the three sound changes of the Blurring Cycle that are required for PND to arise).

We estimate $P_{\chi}(\text{PNV})$ and $P_{\chi}(\text{PND})$ using the *boot* package (Canty & Ripley, 2016; Davison & Hinkley, 1997) in R statistical software (R Core Team, 2016) with 10,000 bootstrap replicates. BSC yields the following estimates with 95% adjusted bootstrap percentile (BC_a) intervals:

² If more comprehensive surveys of sound changes were available, an appropriate process for estimating probabilities based on rates of sound changes that includes the temporal dimension would be the Poisson Stochastic Process (see Beguš, 2016). For practical purposes, we can disregard the temporal dimension and estimate Historical Probabilities with BSC.

- (6) a. $P_{\chi}(\text{PNV}) = 20.5\%$, $\text{BC}_a \text{ CI} = [15\%, 26\%]$
 b. $P_{\chi}(\text{PND}) = 0.047\%$, $\text{BC}_a \text{ CI} = [0.018\%, 0.12\%]$

The Historical Probabilities in (6) show that CB is capable of deriving the typology: it predicts PND is less frequent compared to its natural counterpart PNV. Moreover, BSC predicts that PND will be attested once in a given sample: we can compare $P_{\chi}(\text{PND})$ with the probability of a process being attested once in a given sample $P(\frac{1}{200})$. The difference between the two probabilities is not significant (the 95% $\text{BC}_a \text{ CI}$ of the difference being $[-0.35\%, 2.10\%]$), which means that we expect PND to be attested. PND is in fact attested once in the survey in Kümmel (2007). BSC also allows comparison between PND and other unnatural alternations, such as final voicing (FV) or intervocalic devoicing (IVD), two processes that are believed to be unattested. The difference between the two is significant: $P_{\chi}(\text{PND})$ is significantly higher than $P_{\chi}(\text{FV})$ (95% $\text{BC}_a \text{ CI}$ of the difference being $[0.01\%, 0.09\%]$) and $P_{\chi}(\text{IVD})$ (95% $\text{BC}_a \text{ CI}$ of the difference being $[-0.10\%, -0.02\%]$). Figure 1 shows distributions of the three unnatural processes, PND, IVD, and FV.

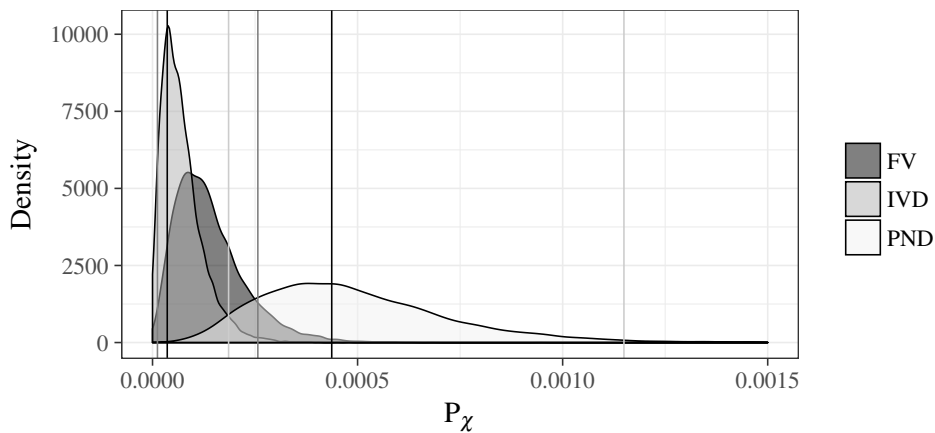


Figure 1: Bootstrap distribution of Historical Probabilities of final voicing (FV), intervocalic devoicing (IVD), and post-nasal devoicing (PND) with 95% BC_a confidence intervals (with 21 bootstraps removed).

These and other applications of BSC are, however, beyond the scope of this paper. The following sections focus on the implementation of Historical Probabilities in a MaxEnt model of typology.

3. Typology within AB

Numerous studies experimentally confirm that some alternations are underlearned (for a survey, see Moreton & Pater, 2012a,b). Learnability differences are encoded in MaxEnt models of phonological learning in two similar ways: Wilson (2006) differentiates variance (σ^2), while White (2017) differentiates weights (μ) in the regularization term (prior) of different constraints to encode that some processes require more input data to be learnt. These prior variances or weights are determined independently from P-map related perceptual distance measures in both Wilson (2006) and White (2017).

The evidence for AB is strongest when testing featurally more vs. less complex alternations (structural bias, Moreton & Pater, 2012a,b). While structurally complex alternations are consistently underlearned, much less robust results are obtained when testing alternations that target a single feature value where one direction is phonetically natural and typologically common and the other is unnatural and rare (substantive bias; Moreton & Pater, 2012a,b). In fact, two studies specifically tested the learnability of PND and IVD compared to their natural counterparts (PNV and intervocalic voicing) and found no significant difference between the natural/unnatural pairs (Seidl et al., 2007; Do et al., 2016).³

³ Recently, a study testing the generalization of final vs. initial voicing contrast also yielded negative results: there

These natural/unnatural pairs of alternations pose a problem for the AB approach to phonology. Typologically, unnatural alternations are considerably rarer (7.6% vs. approximately 0.4% for PNV vs. PND, based on Locke, 1983, reported in Hayes & Stivers, 2000, and our survey), yet artificial grammar learning experiments seem to yield negative results: unnatural processes do not seem to be underlearned compared to their natural counterparts. In modeling terms, this means that we cannot assume different σ^2 in the prior of the natural (e.g. *NT) and unnatural (e.g.*ND) constraints: both should have equal σ^2 . Even if we assume σ^2 is determined by P-map (Steriade, 2001) related metrics (as is assumed in Wilson, 2006 and White, 2017), we would not expect differences in prior variance between *NT and *ND, at least not under the symmetric P-map approach where $\Delta(T, D)/N_{_} = \Delta(D, T)/N_{_}$.

The discrepancy between artificial grammar learning experiments that test the learnability of natural vs. unnatural processes on the one hand and the typology on the other hand suggests that a model that combines AB and CB influences on typology will perform better than the current “split” models. The next section outlines a new model that combines the MaxEnt approach to modeling learning (and consequently AB) as proposed in Wilson (2006) with the new model of typology within the CB, proposed in section 2 above and in Beguš (2016).

4. A new model of typology

The ability to derive typological predictions has long been a strength of constraint-based Optimality Theory (OT) and OT-related theories (Prince & Smolensky, 1993/2004). Typological predictions are primarily achieved by restraining constraint inventory (CON): most versions of OT disallow phonetically unnatural constraints in CON. Under this approach, the natural constraint *NT is part of the CON, while the unnatural constraint *ND is not (cf. Hyman, 2001). This means that for the input /ND/, the output candidate [NT] is harmonically bounded because it violates both the markedness *NT constraint as well as the faithfulness IDENT-IO(voice) constraint. In other words, the mapping /ND/ → [NT] (PND) is predicted to be impossible. This, however, is an undesired prediction. PND has been confirmed as a productive synchronic alternation in Tswana and Shekgalagari (Coetzee & Pretorius, 2010). Thus, OT with CON restricted to natural constraints undergenerates.

Restricting CON poses problems for the derivation of not only categorical alternations, but also of gradient processes. While classical OT is suitable for deriving categorical processes, gradient phenomena are usually modeled with Harmonic Grammar or related frameworks that operate with weighted constraints (Legendre et al., 2006; Pater, 2008, 2009; Coetzee & Pater, 2008). Beguš & Nazarov (2017) identify a typological prediction of Harmonic Grammar and related frameworks that has so far gone largely unnoticed: “HG with restricted CON predicts that the probability of the natural feature value in a given environment is always equal or greater than the probability of the unnatural value in a given environment” (Beguš & Nazarov, 2017). This generalization is called the “Natural Gradient Bias” (NGB). Beguš & Nazarov (2017) present two synchronic gradient phonotactic restrictions that violate NGB: phonotactic restrictions that operate in the unnatural direction where the unnatural feature value is significantly more frequent than the natural one in a given environment.

The cases of unnatural alternations or phonotactic restrictions from Table 1 point to the fact that both the Classical OT as well as HG and related frameworks undergenerate under the restricted CON approach. However, to simply relax the CON and admit all constraints is not a viable solution either: this would result in the loss of all predictive power of the OT and HG family of frameworks. Instead, a model of typology should derive all attested processes, but at the same time encode that some processes are rare and others more frequent.

The new model of typology proposed here admits all constraints, natural, unmotivated, and unnatural, in the CON in order to avoid undergeneration. The model outputs typological probabilities over candidates for given inputs: a MaxEnt-compatible framework. To encode that some processes are rare due to learning biases, we adopt Wilson’s (2006) model of differentiating variance (σ^2) in the prior of different constraints. To encode that some processes are rare due to the number of sound changes they require and their respective probabilities (CB), we introduce “Historical Weights” (w_χ) of different

were no differences between the more natural word-initial voicing contrast and the unnatural word-final voicing contrast (Glewwe, 2017).

constraints. Differences in Historical Weights between different faithfulness and markedness constraints can be directly calculated from estimated Historical Probabilities using the BSC, as shown in ((7)).

$$(7) \quad \Delta w_\chi = -\log\left(\frac{P_\chi}{1-P_\chi}\right)$$

Let us take for example the case of PNV and PND. Typologically, there exists a considerable difference between the two: 15 languages of 197 surveyed in Locke (1983) (reported in Hayes & Stivers, 2000) feature PNV as a synchronic alternation. On the other hand, only two related languages known to me, Tswana and Shekgalagari, feature PND as a productive synchronic alternation.⁴

Artificial grammar learning experiments suggest that the two processes, PNV and PND, are equally learnable (Do et al., 2016). Further experiments are warranted, but currently we have to assume that prior σ^2 for both *NT and *ND constraints should be equal. Equal prior σ^2 for the two constraints should also be assumed under the symmetric P-map assumption.

The difference in Historical Weights between faithfulness IDENT-IO(voice) and markedness *NT and *ND constraints can be calculated using the Historical Probabilities in (6) and the formula in (7).

$$(8) \quad \begin{array}{l} \text{a. } \Delta w_\chi(*\text{NT}, \text{IDENT-IO(voi)}) = 1.36 \\ \text{b. } \Delta w_\chi(*\text{ND}, \text{IDENT-IO(voi)}) = 7.66 \end{array}$$

Because the prior variance of the two markedness constraints has to be equal, we can disregard the influence of AB on the typology and use Historical Weights to encode that the mapping /NT/ \rightarrow [ND] is typologically much more frequent than the mapping /ND/ \rightarrow [NT].

Table 2: Tableaux for inputs /NT/ and /ND/ that show the MaxEnt probability distribution over candidates in a typological model.

/NT/	IDENT-IO $w_\chi = 10$	*NT $w_\chi = 8.64$	H_χ	P_χ	Typol.
a. [NT]		-1	-10	.795	.924
b. [ND]	-1		-8.64	.205	.076
/ND/	IDENT-IO $w_\chi = 10$	*ND $w_\chi = 2.34$	H_χ	P_χ	Typol.
a. [ND]	-1		-10	.99953	$\approx .996$
b. [NT]		-1	-2.34	.00047	$\approx .004$

To be sure, learners have no access to Historical Weights: when we model phonological learning, Historical Weights should be disregarded completely; only prior variance should determine differences in learning of different processes (in Wilson’s 2006 terms). However, when we model typology, both prior variance and Historical Weights should affect the outcome.

The combined typological model thus encodes that PND and PNV are equally learnable, but one has a higher Historical Probability. Both are derivable, but PND is correctly predicted to be much less frequent. Table 2 illustrates how Historical Weights affect the Historical Probabilities over candidates for given inputs in a MaxEnt-compatible model of typology. The table also shows that Historical Probabilities closely match the observed typology (based on the survey in Locke, 1983 and my estimates of the typological frequency of PND).

For every synchronic alternation and its typological distribution, we now *can* and *should* calculate its Historical Probability (CB part) based on the proposed BSC, and its σ^2 (AB part) which should ideally be calculated on the basis of learnability experiments. As already mentioned in section 1, there is a good amount of overlap between the two: processes that are more complex are underlearned in artificial grammar learning experiments (e.g. saltation, White, 2014), but they also have lower Historical Probabilities, because complex alternations require more sound changes to arise. In fact, one could argue that AB and CB are not independent and that frequencies of sound changes are primarily influenced by

⁴ It is difficult to establish how many languages were included in my survey of post-nasal devoicing, which is why I estimate it at approximately 500 languages.

learnability (AB) (cf. Kiparsky, 1995, 2008). Unnatural alternations, as defined in Beguš (2016) and in section 2.2 above, however, offer a crucial basis for disambiguating AB and CB. Even under the radical assumption that frequencies of sound changes are primarily influenced by learnability, the Channel Bias crucially has to influence frequencies of unnatural alternations (as defined above), precisely because they arise through a combination of sound changes — the Blurring Process. In other words, even if individual sound changes are influenced by learnability, the combination of sound changes comprising the Blurring Process itself is not, which means that unnatural alternations provide a valuable resource for disambiguating AB and CB (as is shown in the new typological model in Table 2 above).

5. Future directions

The proposed model of typology within CB that operates with Historical Probabilities and the BSC is most accurate with large and representative samples of sound changes. The first next step in improving the model is to expand surveys of sound changes.

In addition to solving the problem of undergeneration while retaining typological predictions, the model of typology proposed in this paper bears the potential to solve the Too Many / Too Few Solutions (TMTFS) problem (Steriade, 2001). For some markedness constraints, there exist several different repair strategies, whereas other markedness constraints seem to allow only one repair strategy (or rely on one strategy considerably more frequently than others). The new model of typology is able to encode such distributions: greater differences in Historical Weights between markedness constraints and different faithfulness constraints that correspond to these markedness constraints result in a strong preference for a single repair strategy, while at the same time allowing for other repair strategies (with much lower frequencies). Smaller differences in Historical Weights, on the other hand, mean that more repair strategies will be available and that they will tend to be more equiprobable. Perhaps the two most famous cases illustrating the Too Many / Too Few Solution problem are repairs of the markedness constraints *NT and *D#. Pater (1999) shows that languages exhibit a variety of strategies for repairing the markedness constraint *NT, including deletion, nasalization, and voicing. On the other hand, in the majority of languages that show an alternation in word-final voiced stops, the repair strategy is that of devoicing. Steriade (2001) proposes a P-map solution, claiming that perceptual distance between voiced and voiceless stops is smallest in word-final position (smaller than the perceptual distance between voiced stops and any other segment). Recently, however, a synchronic system that repairs *D# with nasalization has been reported in Noon (Merrill, 2015). Final nasalization in Noon arises through a combination of sound changes. Beguš & Nazarov (2017) also suggest that other reports of final nasalization (in Blust, 2016) arise as by-products of the Blurring Chain. Initial investigation into the TMTFS problem thus shows that Historical Weights (or CB) affect the typology of repair strategies (TMTFS) as well. The relationship between Historical Weights (CB) and prior variance (AB) when deriving the typology of repair strategies should be explored further.

Finally, the paper leaves open the question of what determines the prior variance of different constraints and how exactly it affects the typology (cf. Staubs, 2014): it is possible that some constraints are innate and others acquired (cf. Tesar & Smolensky, 1993; Hayes, 1999), which would influence σ^2 . Other possible inputs for σ^2 include the P-map (Wilson, 2006) or a structural complexity metric (Pater & Moreton, 2012); the relationship between the two should also be explored further.

References

- Adelaar, Willem F. H. (1977). *Tarma Quechua: Grammar, texts, dictionary*. The Peter de Ridder Press, Lisse.
- Beguš, Gašper (2016). Post-nasal devoicing and a probabilistic model of phonological typology. Ms., Harvard University.
- Beguš, Gašper & Aleksei Nazarov (2017). Lexicon against naturalness: Unnatural gradient phonotactic restrictions and their origins. Ms., Harvard University.
- Blevins, Juliette (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge University Press, Cambridge.

- Blust, Robert (2005). Must sound change be linguistically motivated? *Diachronica* 22:2, 219–69.
- Blust, Robert (2013). *The Austronesian Languages*. Asia-Pacific Linguistics, Canberra.
- Blust, Robert (2016). Austronesian against the world: where the P-map ends. Paper presented at 42nd Annual Meeting of the Berkeley Linguistic Society, University of California, Berkeley, February 5-6, 2016.
- Canty, Angelo & Brian Ripley (2016). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-18.
- Cathcart, Chundra (2015). A probabilistic model of evolutionary phonology. Buiand, Thuy & Deniz Özyıldız (eds.), *Proceedings of the Forty-Fifth Annual Meeting of the North East Linguistic Society (NELS 45)*, GLSA, Amherst, MA, vol. 1, 145–50.
- Coetzee, Andries W. & Joe Pater (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26:2, 289–337.
- Coetzee, Andries W. & Rigardt Pretorius (2010). Phonetically grounded phonology and sound change: The case of Tswana labial plosives. *Journal of Phonetics* 38, 404–21.
- Davison, A. C. & D. V. Hinkley (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge.
- Do, Youngah, Elizabeth Zsiga & Jonathan Havenhill (2016). Naturalness and frequency in implicit phonological learning. Talk presented at the 90th Annual Meeting of the Linguistic Society of America, Washington, DC, January 7-10, 2016.
- Efron, Bradley (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 7:1, 1–26.
- Garrett, Andrew (2014). Sound change. Bowerman, Claire & Bethwyn Evans (eds.), *The Routledge Handbook of Historical Linguistics*, Routledge, New York, NY, 227–48.
- Glewwe, Eleanor (2017). Substantive bias in phonotactic learning: Positional extension of an obstruent voicing contrast. Talk presented at the 53rd meeting of Chicago Linguistic Society, Chicago, IL, May 25-27, 2017.
- Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a maximum entropy model. Spenader, Jennifer, Anders Eriksson & Östen Dahl (eds.), *Proceedings of the Workshop on Variation within Optimality Theory*, Stockholm University, Stockholm, 111–20.
- Hayes, Bruce (1999). Phonetically-driven phonology: The role of Optimality Theory and inductive grounding. Darnell, Michael, Edith Moravcsik, Michael Noonan, Frederick Newmeyer & Kathleen Wheatly (eds.), *Functionalism and Formalism in Linguistics*, John Benjamins, Amsterdam, vol. 1, 243–85.
- Hayes, Bruce & Tanya Stivers (2000). Postnasal voicing. Ms.
- Hyman, Larry (2001). The limits of phonetic determinism in phonology: *NC revisited. Hume, Elizabeth & Keith Johnson (eds.), *The Role of Speech Perception in Phonology*, Academic Press, San Diego, CA, 141–86.
- Janson, Tore (1991/1992). Southern Bantu and Makua. *Sprache und Geschichte in Afrika* 12/13, 1–44.
- Janssens, Baudouin (1993). *Doubles réflexes consonantiques: quatre études sur le bantou de zone A (bubi, nen, bafia, ewondo)*. Ph.D. thesis, Université libre de Bruxelles, Faculté de Philosophie et Lettres.
- Kiparsky, Paul (1995). The phonological basis of sound change. Goldsmith, John A. (ed.), *The Handbook of Phonological Theory*, Blackwell, Oxford, 640–70.
- Kiparsky, Paul (2006). Amphichronic program vs. Evolutionary Phonology. *Theoretical Linguistics* 32:2, 217–36.
- Kiparsky, Paul (2008). Universals constrain change, change results in typological generalizations. Good, Jeff (ed.), *Linguistic universals and language change*, Oxford University Press, Oxford, 23–53.
- Kümmel, Martin (2007). *Konsonantenwandel*. Reichert, Wiesbaden.
- Legendre, Géraldine, Antonella Sorace & Paul Smolensky (2006). The Optimality Theory—Harmonic Grammar connection. Smolensky, Paul & Géraldine Legendre (eds.), *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*, MIT Press, Cambridge, MA, 339–402.
- Locke, John (1983). *Phonological Acquisition and Change*. Academic Press, New York, NY.
- Merrill, John (2015). Nasalization as a repair for voiced obstruent codas in Noon. *LSA Annual Meeting Extended Abstracts* 6:14, 1–5.
- Merrill, John (2016a). A historical account of the Fula and Sereer consonant mutation and noun class systems. Ms., University of California, Berkeley.
- Merrill, John (2016b). Konyagi post-nasal devoicing? Ms., University of California, Berkeley.
- Moreton, Elliott (2008). Analytic bias and phonological typology. *Phonology* 25:1, 83–127.
- Moreton, Elliott & Joe Pater (2012a). Structure and substance in artificial-phonology learning. Part I. *Language and Linguistic Compass* 6:11, 686–701.
- Moreton, Elliott & Joe Pater (2012b). Structure and substance in artificial-phonology learning. Part II. *Language and Linguistic Compass* 6:11, 702–18.
- Pater, Joe (1999). Austronesian nasal substitution and other NC effects. Kager, R., H. Van der Hulst & W. Zonneveld (eds.), *The Prosody Morphology Interface*, Cambridge University Press, Cambridge, 310–43.
- Pater, Joe (2008). Gradient phonotactics in Harmonic Grammar and Optimality Theory. Ms., University of Massachusetts, Amherst.
- Pater, Joe (2009). Weighted constraints in generative linguistics. *Cognitive Science* 33, 999–1035.

- Pater, Joe & Elliott Moreton (2012). Structurally biased phonology: Complexity in learning and typology. *The EFL Journal* 3:2, 1–44.
- Picard, Marc (1994). *Principles and Methods in Historical Linguistics*. McGill University Press, Montreal.
- Prince, Alan & Paul Smolensky (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell, Malden, MA. First published in 1993, Tech. Rep. 2, Rutgers University Center for Cognitive Science.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Rohlf, Gerhard (1949). *Lautlehre*, vol. 1 of *Historische Grammatik der Italienischen Sprache und ihrer Mundarten*. Francke, Bern.
- Seidl, Amanda, Eugene Buckley & Alejandrina Cristià (2007). Complexity trumps naturalness. Talk presented at the 81st Annual Meeting of the Linguistic Society of America, Anaheim, CA, January 4-7, 2007.
- Solé, Maria-Josep, Larry M. Hyman & Kemmonye C. Monaka (2010). More on post-nasal devoicing: The case of Shekgalagari. *Journal of Phonetics* 38:4, 299–319.
- Staubs, Robert D. (2014). *Computational Modeling of Learning Biases in Stress Typology*. Ph.D. thesis, University of Massachusetts, Amherst.
- Steriade, Donca (2001). The phonology of perceptibility effects: The P-map and its consequences for constraint organization. Ms., University of California, Los Angeles.
- Tesar, Bruce & Paul Smolensky (1993). The learnability of Optimality Theory: An algorithm and some basic complexity results. Ms., University of Colorado at Boulder. ROA 2.
- Wang, William S-Y. (1968). Vowel features, paired variables, and the english vowel shift. *Language* 44:4, 695–708.
- White, James (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130:1, 96–115.
- White, James (2017). Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a P-map bias. *Language* 93:1, 1–36.
- Wilson, Colin (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30:5, 945–82.
- Xromov, Al'bert (1972). *Jagnobskij jazyk*. Nauka, Moscow.
- Xromov, Al'bert (1987). Jagnobskij jazyk. *Osnovy iranskogo jazykoznanija. Novoiranskie jazyki: vostočnaja grupa*. Nauka, Moscow, 644–701.
- Yu, Alan C. L. (2004). Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language* 80:1, 73–97.
- Yu, Alan C. L. (2011). On measuring phonetic precursor robustness: a response to Moreton. *Phonology* 28:3, 491–518.