A Maximum Entropy Model of Phonological Typology

Gasper Begus
Harvard University

Introduction. One of the most contested debates in phonology concerns identifying factors that affect typology. The Analytic Bias approach (AB) claims that biases in learning affect the typology, while the Channel Bias approach (CB) assumes phonetic precursors and transmission of language affect the typology (Moreton 2008). Empirical evidence in favor of both hypotheses exists: processes that are typologically rare have been shown to be underlearned; processes that are the result of phonologized phonetically motivated sound changes are also typologically frequent. An increasing body of work acknowledges both influences (Moreton 2008), but very few attempts have been made to model them together or to try to disambiguate the two. This paper aims to fill this gap and proposes a model that unifies the two influences as well as provides grounds for disambiguating AB and CB.

Probabilistic typology within CB. A quantified model of CB influences on typology is the first step towards a new typological model. Current proposals for deriving typology within CB are insufficient or do not produce implementable quantitative results (cf. Blevins 2004, Moreton 2008, Yu 2011, Cathcart 2015). We first argue that phonetically unnatural alternations can only arise through a combination of at least three sound changes (Minimal Sound Change Requirement, MSCR). This generalization is backed by a typological study of unnatural alternations as well as by a formal proof (in a given environment, a natural feature value cannot change into an unnatural one with only two sound changes). Then, we propose a new probabilistic model of CB typology and claim that for every synchronic alternation, we can calculate its HISTORICAL PROBABILITY (P_χ) based on the number of sound changes required for the alternation to arise and the probability of each sound change required. Calculation of P_χ is not a trivial task: we propose a new method of estimating historical probabilities from typological surveys called "bootstrapping sound changes" (BSC). Historical probabilities are bootstrapped (Efron 1979) from a sample of successes (languages in the sample with a sound change S_1) and failures (languages in the sample without S_1). If an alternation A_x requires $n > 1$ sound changes to arise, P_χ is bootstrapped from a product of probabilities based on the number of successes and failures (divided by $n!$ to account for the ordering of sound changes).

For example, we can estimate P_χ of natural and unnatural alternations (such as post-nasal voicing, PNV vs. post-nasal devoicing, PND). P_χ(PNV) is considerably greater (20.5%, BC_a CI = [15%, 26%]) than P_χ(PND) (0.047%, BC_a CI = [0.018%, 0.12%]) based on the number of sound changes the two alternations require and their probabilities and using the BSC method on the sample of sound changes in Kümmel (2007). P_χ of the other two unnatural alternations discussed here, final voicing (FV) and intervocalic devoicing (IVD), are also very low (0.0028%, BC_a CI = [0.00045%, 0.015%] for FV, and 0.0064%, BC_a CI = [0.0013%, 0.027%] for IVD). The BSC method has several implications: it allows us to (i) compare P_χ of different alternations with statistical inference (e.g. P_χ(FV) is significantly lower than P_χ(PND), with the BC_a CI of the difference being [0.02%, 0.11%]); (ii) identify historically equiprobable processes for artificial grammar learning experiments; (iii) predict the (un)attestedness of alternations in a given sample; and finally, (iv) BSC provides quantified means for encoding Channel Bias in a typological model.

Typology within AB and MaxEnt. Numerous studies experimentally confirm that some alternations are underlearned. The evidence for AB is strongest when testing featurally more vs. less complex alternations (structural bias, Moreton and Pater 2012). AB is 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 of different constraints to encode that some processes require more input data to be learnt. These priors are determined independently, from P-map related perceptual distance measures.

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 and Pater 2012). 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). This mismatch strongly suggests that a model that admits both AB and CB influences on phonology will perform better than the current split models.

A new typological model. The paper proposes a new model of typology that combines AB and CB and employs MaxEnt probability distribution over candidates. All constraints, both phonetically natural and unnatural, are admitted to CON in order to avoid undergeneration. In order to encode that some alternations are rare due to the required number of sound changes and their respective probabilities, historical probabilities estimated with BSC are used as an input to prior "historical" weights (w_χ). We can calculate differences in historical weights between Faithfulness and Markedness constraints from historical probabilities based on BSC (1).

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

To encode learnability differences, we adopt Wilson's (2006) approach of differentiating variance (σ^2) in the prior, which effectively means that some processes require more input data to be learnt.

When we model the learning of synchronic grammar in MaxEnt, different variances in the prior (σ^2) will encode that some processes are more difficult to learn (either based on structural complexity or P-map), but crucially speakers will have no access to historical weights (w_χ) or probabilities (P_χ). However, when we model the typology, both CB (w_χ) and AB (σ^2) will contribute to the final result: the rarity of PND compared to PNV, where both processes appear to be equally learnable (Do et al. 2016), is encoded by w_χ (CB), while the rarity of processes that are historically equiprobable but show different learnability is encoded by different variance (σ^2) in the prior (AB).

Disambiguation. The model also aims to provide quantitative grounds for disambiguating AB and CB: for every typological observation, we now *can* and *should* calculate the CB contribution (based on BSC) and AB contribution (based on learnability experiments). For most processes, there is overlap in AB and CB influences and it is difficult to disambiguate the two. In fact, it is possible that AB influences historical probabilities if we assume sound change primarily results from learning errors. However, this "duplication" problem is avoided in the case of unnatural processes: we argue that a combination of at least three sound changes is required for any unnatural alternation (MSCR). The fact that sound changes need to operate in combination means that CB has to independently influence historical probabilities of unnatural processes, even if we assume a radical AB hypothesis where probabilities of individual sound changes are primarily determined by learnability.

Example. We can calculate the differences in historical weights between PND and PNV based on their historical probabilities according to (1). Δw_χ between Faithfulness and Markedness constraints is 7.66 for PND and 1.36 for PNV (2). The two processes, however, appear to have equal variance (σ^2): the study that tested the learnability of the two found no significant difference (Do et al. 2016). Even if we use P-map (Steriade 2001) to encode σ^2, we would expect no difference between the two processes, at least under the symmetric P-map approach where Δ(T, D)/N__ and Δ(D, T)/N__ are equal. 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: out of a sample of 197 languages, 15 are reported to feature PNV as an alternation (Locke 1983), but only one language in a survey of all sources available to the author features PND.

(2)

/NT/	IDENT-IO $w_\chi = 10$	*NT $w_\chi = 8.64$	H_χ	P_χ	/ND/	IDENT-IO $w_\chi = 10$	*ND $w_\chi = 2.34$	H_χ	P_χ
a. [NT]		−1	−10	.795	a. [NT]	−1		−10	.99953
b. [ND]	−1		−8.64	.205	b. [ND]		−1	−2.34	.00047

Future directions. BSC allows us to identify historically equiprobable processes; experimental testing of such processes would directly confirm the effect of AB on typology. Additionally, dispreference against alternations is a consistent result in artificial language learning experiments. Discrepancies between historical prediction and attested typology suggest one of the prominent roles of AB is in reducing unnatural alternations, but future work should further test this hypothesis. Finally, the paper leaves open the question of what determines σ^2: it is possible that some constraints are innate and others acquired (cf. Tesar and Smolensky 1993, Hayes 1999) which would influence σ^2. Other possible inputs for σ^2 include P-map (Wilson 2006) or a structural complexity metric (Pater and Moreton 2012); the relationship between the two should also be explored further.