

Emergent learning bias and the underattestation of simple patterns

Charlie O’Hara

Accepted with minor revisions at Natural Language and Linguistic Theory DRAFT: June 18, 2022

Abstract This paper investigates the typology of word-initial and word-final place of articulation contrasts in stops, revealing two typological skews. First, languages tend to make restrictions based on syllable position alone, rather than banning particular places of articulation in word-final position. Second, restrictions based on place of articulation alone are underrepresented compared to restrictions that are based on position. This paper argues that this typological skew is the result of an emergent bias found in agent-based models of generational learning using the Perceptron learning algorithm and MaxEnt grammar using independently motivated constraints. Previous work on agent-based learning with MaxEnt has found a simplicity bias (Pater and Moreton 2012) which predicts the first typological skew, but fails to predict the second skew. This paper analyzes the way that the set of constraints in the grammar affects the relative learnability of different patterns, creating learning biases more elaborate than a simplicity bias, and capturing the observed typology.

1 Introduction

A major goal of phonological theory is to develop a model that can generate the range of possible phonological systems. Particularly in constraint-based frameworks—Optimality Theory (Prince and Smolensky 1993/2004; McCarthy and Prince 1995), Harmonic Grammar (Legendre et al. 1990, 2006; Pater 2016), etc.—a large part of this labor has historically been delegated to THE GRAMMAR, the phonological component that explains the mapping between underlying and surface representations.

The set of possible phonological patterns is deeply connected to the set of patterns actually observed in the world’s languages, allowing typological results to be used as evidence in phonological theory. Traditionally, constraint-

based models of phonology are concerned only with whether or not predicted patterns are attested, but we may also use the relative rate of attestation of phonological patterns, which I call the *SOFT TYPOLOGY* of the world's languages, as evidence for our phonological theories.¹ The attestation rate of a pattern is likely based on a large variety of factors, ranging from perceptual and articulatory influences (Blevins 2004) to sociolinguistic and sociological factors. Recent work has focused on the impact of learnability to shape typology (Pater and Moreton (2012); Staubs (2014); Stanton (2016); Hughto (2019), and others). Much of this work assumes a Maximum Entropy Harmonic Grammar (MaxEnt) (Goldwater and Johnson 2003; Hayes and Wilson 2008) as the model of grammar, and the perceptron algorithm as the learning algorithm (Rosenblatt 1958; Boersma and Pater 2016), and finds that skews in typological frequency are correlated to which patterns are easy or hard to learn. Broadly, easy to learn patterns are predicted to be better attested because they are more likely to be accurately learned across many generations.

But what makes a pattern easy or hard to learn? Is learnability more the result of grammatical assumptions, such as the set of constraints used, or the result of the learning algorithm? One possibility is that differences in learnability simply replicate already existing patterns present from grammatical assumptions. In Optimality Theory and its descendants, there can be more than one ranking (or weighting) of constraints that results in the same phonological pattern, and some patterns can be generated with more possible rankings than others. The proportion of possible rankings (or weightings) that can generate a particular pattern is that pattern's *R-VOLUME* (Bane and Riggle 2008). Patterns that can be generated with more rankings of constraints have been argued to be more frequent in synchronic variation and cross-linguistic typology (Anttila 1997; Coetzee 2002; Bane and Riggle 2008; Anttila 2008). Are certain patterns easier to learn simply because there are more ways to learn them? Indeed, some findings about the relative learnability of particular patterns appears to parallel findings about those patterns' r-volumes. Staubs (2014) shows that frequent stress patterns are easier to learn using a MaxEnt model with the Perceptron learning algorithm, and Bane and Riggle (2008) show that attested stress patterns have larger r-volumes than unattested patterns using a similar set of constraints to the ones used in Staubs' simulations. Further, while Hughto et al. (2014); Hughto and Pater (2017); Hughto (2019) show that patterns that require gang effects to model in a weighted constraint grammar are harder to learn than those that do not, Bane and Riggle (2009); Carroll (2010) note that such patterns tend to have relatively small r-volumes. From these studies, it is not clear whether the learning bias is just a result of the skews in r-volume or not. One extreme hypothesis therefore is that the soft typology of the world's patterns is correlated to the r-volume of those patterns

¹ Some have argued that the computational structure imposed by the grammar on the search space may create imbalances in the likelihood of some patterns. As an example, see notions of r-volume (Anttila 1997; Coetzee 2002; Bane and Riggle 2008; Anttila 2008). See section 4.2 for further discussion of this approach to capturing rates of cross-linguistic attestation.

with a particular set of constraints (and that the positions with low r-volumes will across the board be harder to learn than those with high r-volumes).

At the other extreme, another approach has investigated the relative learnability of different pattern in MaxEnt using a minimally restrictive constraint set. Pater and Moreton (2012); Moreton et al. (2017) show that MaxEnt learners with simple, symmetrical, and unrestrictive sets of constraints (GMECCS or Gradual Maximum Entropy with a Conjunctive Constraint Schema) are biased towards featurally simpler patterns, resembling typological generalizations of feature economy (Clements 2003). Where the r-volume of patterns using some particular set of constraints puts the responsibility of soft typology in the hands of the grammar rather than the learner, the GMECCS model assumes a minimal grammar leading the dynamics of the learning algorithm to be more responsible for the learning results. In this paper, I examine typological skews that are best explained through the interaction of the grammar and the learner, rather than prioritizing one of the two components like the r-volume or GMECCS models. I show that the learnability of patterns in a MaxEnt model with independently motivated set of constraints makes closer matches to typological skews than the r-volume of patterns with that set of constraints, or the learnability of those patterns with a minimal set of constraints. The interaction of learner and the grammar interact in ways that are not apparent from focusing on either one in particular.

The interactive model of learning and grammar allows for types of learning bias to be predicted that cannot be predicted by the GMECCS model with a simple symmetrical constraint set. The GMECCS model predicts that learning biases are favor structurally simple patterns over more complex patterns, but leaves phonetically natural substantive typological skews to be explained by external factors.

In the most basic case, this simplicity bias favors patterns that can be defined based on a single feature over those that require multiple features.

(1) Simplicity Scale:

A restriction based on one feature is simpler than one based on two.

$*[+F]$ is simpler than $*[+F,+G]$ and $*[+F]\vee[+G]$

For example, Pater and Moreton (2012) consider the typology of obstruent stop inventories. Languages that have three voiceless stops [p], [t], and [k] may allow all (2), none (3), or a subset of the voiced stops [b d g] (4). The pattern that bans all voiced stops can be defined using just [+voice], making it simpler than the pattern that bans only [g], which requires $*[+voice, dorsal]$. Pater and Moreton (2012) present a typological survey showing that voiced stops [b] and [g] pattern together across most of the 451 languages in the UPSID-92 database (Maddieson and Precoda 1992). They find that 54% of these languages allow both, 34% allow neither, but only 9.5% allow [b] and ban [g], and 2.7% allow [g] but ban [b]. The patterns that treat [b] and [g] differently are less common, and also involve more complex restrictions.

- (2) **Simple:** All stops are available
 p t k
 b d g
- (3) **Simple:** Voiced stops are banned, $*[+voice]$
 p t k
b d g
- (4) **Complex:** Voiced dorsal stops are banned, $*[+voice]\&[dorsal]$
 p t k
 b d **g**

Pater and Moreton (2012) show that this typological skew matches a simplicity bias present in MaxEnt learners. Structural simplicity bias has been long observed in non-linguistic learning tasks (Shepard et al. 1961; Hunt and Hovland 1960; Bruner et al. 1956), and have recently been observed in artificial grammar learning experiments (Moreton and Pater 2012a; Pertsova 2012; Glewwe 2019) as well.

Typological skews may also be substantive, where the phonological substance of the features involved in a pattern affects its rate of attestation (Wilson 2006). Substantive typological skews are those where patterns defined based on some feature value, or combination of feature values are more well-attested typologically than other feature values, even though the patterns are structurally similar. For example, languages tend to ban voiced stops word-finally more often than they ban voiceless stops word-finally. One pattern bans the $[+voice]$ feature in word-final stops, whereas the other bans the $[-voice]$ feature in this position, making these patterns structurally indistinguishable. Only the phonetic substance, or the identity and values of the features themselves, can distinguish these patterns.

Typically, the substantive biases (and asymmetries) focused on in the literature are those that favor phonetically natural patterns. Phonetically natural typological skews can often be easily understood as an effect of channel bias (Moreton 2008), so it is unclear whether they are necessarily the result of a learning bias. Evidence for substantive bias in artificial grammar learning experiments has been mixed (see Moreton and Pater 2012b; Glewwe 2019), though substantive biases have been observed (Wilson 2006; White 2014; Finley 2012; Kimper 2016; Lin 2016; Martin and Peperkamp 2020). These substantive biases have also been observed more strongly in natural language experiments (Hayes and White 2013; Prickett 2018).

However, substantive biases do not need to be based on phonetic naturalness. Any learning bias that favors one feature or feature value over another can be classified as a substantive bias. When phonetically natural patterns are better attested than structurally similar phonetically unnatural patterns, the resulting skew in typology may be able to be explained solely as the result of some channel bias. However, not all typological skews are either structural or phonetically natural: in some cases two equally simple patterns differ greatly in attestation, but neither is clearly more phonetically natural than the other.

While phonetically natural skews in typology may be able to result from channel biases without being encoded directly in the phonological grammar, it is not clear that all substantive skews in typology can be easily explained through channel bias. On the other hand, when MaxEnt learners are supplied with constraint sets that are not simple and unbiased, those learners may have a substantive bias towards some patterns over others. This paper investigates how encoding substantive information, such as markedness hierarchies, in a set of constraints affects the learnability of phonological patterns, creating substantive biases that are not clearly based in phonetic naturalness.

Specifically, this paper focuses on the interaction of place of articulation and syllable position in stop inventories. If both place of articulation and syllable position are treated as features by a learning model, a simplicity bias would predict that patterns that make restrictions based on only syllable position or only place of articulation would be more easily learned than patterns that require an interaction of both of these dimensions, and therefore the simple patterns would be better attested. Section 2 presents a typological survey of place and position in stop inventories. This survey is not entirely consistent with the simplicity hypothesis: languages make restrictions based on syllable position much more often than they make restrictions based on the interaction of position and place of articulation, but this effect is not symmetric: languages are much more likely to make restrictions on position alone than on place of articulation alone.

Section 3 presents a generational model of MaxEnt learning, also called an iterated learning model. Learning simulations using an independently motivated set of constraints produce a close match to the typological survey. Typologically underattested patterns are learned slightly slower and less accurately than common patterns are. Across many generations, the small difference in learning speed leads these underattested patterns to be learned less stably. In section 4, two alternative approaches to capturing these typological generalizations are explored: the GMECCS model, a learning-based model without substantive bias in the constraints themselves, and the r-volume model, which attempts to capture typological tendencies through grammatical bias without learning ((Bane and Riggle 2008; Anttila 1997; Coetzee 2002)). It is demonstrated that each of these alternatives is missing half of the picture, and cannot capture the observed typology as well as the substantively biased model.

This paper demonstrates new typological generalizations based on a typological survey of word-initial vs. word-final stop inventories. Further, these generalizations are shown to be predicted based on the interaction of the Perceptron learning algorithm and independently motivated sets of constraints.

2 Soft Typology of word initial and word final stop inventories

This section reports the soft typology of place of articulation restrictions of word-initial and word-final stops. The typological survey presented here shows two major skews in rates of attestation, summarized in (5). The All-or-Nothing

Skew can be understood as a purely structural skew in the typology, but the Positional Priority Skew is inherently substantive. This typology involves two dimensions: word-position and place of articulation, both have been argued to involve markedness hierarchies, but languages are far more likely to make restrictions based on position rather than place. This is a substantive skew because it is an asymmetry between the substance of the dimensions themselves.

- (5) a. **All-or-Nothing Skew:** Languages that allow [t], [p], and [k] word initially tend to allow either no places of articulation for stops word-finally, or all of the places of articulation available for stops in word-initial position. Languages are relatively unlikely to allow a proper subset of places of articulation available word-initially in word-final position
- b. **Positional Priority Skew:** Languages are more likely to make restrictions based on word-position (i.e. ban all final stops), than to make restrictions based on place of articulation (i.e. ban all dorsal stops).

Sections 2.1 and 2.2 will discuss some of the preliminary assumptions and methodologies used in the typological survey, and section 2.3 will present the results and demonstrate evidence for the all-or-nothing and positional priority skews.

2.1 Positional Stop Inventories

The simplicity bias hypothesis can be evaluated by finding the attestation of patterns that make restrictions using interactions of two different dimensions. Here I focus on major place of articulation and syllable position, because information about these properties is usually easy to find in language grammars, and they are uncontroversial enough to be reliably observed. Both of these dimensions are thought to involve cross-linguistic markedness hierarchies, but we will see that subtle differences exist in how these hierarchies are usually thought about.

Here I will assume the place of articulation hierarchy as defined by de Lacy (2006), expressed in (6). de Lacy (2006)'s implementation of this hierarchy builds on a large body of literature on place of articulation markedness (Kean 1975; Paradis and Prunet 1991; Jun 1995; Lombardi 1998). de Lacy's major contribution relevant to this paper is the relationship between labials and dorsals. In this hierarchy, coronal segments are less marked than labial segments, which in turn are less marked than dorsal segments.²

- (6) *Place of Articulation Markedness Hierarchy*
Dorsal \prec Labial \prec Coronal

² \prec here denotes the relation "is less harmonic (or more marked) than".

Evidence for the unmarkedness of coronals comes from a variety of typological evidence, such as patterns of default epenthesis and outputs of neutralization processes (see de Lacy (2006) for a review).³ Throughout this paper, I will assume the existence of the markedness hierarchy proposed by de Lacy (2006), but it is possible that the observed typological skews could also be modeled using slightly different hierarchies. The typological survey in this paper is too small to directly test this hierarchy, however see (REDACTED §2.5) for further discussion of the evidence of this hierarchy in a larger typological survey.

Turning to the other dimension, languages tend to exhibit more contrasts in onset position than coda position (Kingston 1985; Itô 1986; Goldsmith 1990). Further, there are strong typological implications with regard to syllable structure. All languages have CV syllables—V syllables imply CV syllables, and CVC syllables imply CV syllables—suggesting that onsetful syllables are less marked than onsetless syllables, whereas codaless syllables are less marked than codafull syllables (Jakobson and Halle 1956; Blevins 1995).⁴ Thus, onset stops can be considered less marked than coda stops.

There are several differences between these two scales worth calling attention to. First, there appears to be a difference between the relative markedness of members on these two scales: onsets are less marked than onsetless syllables, so consonant deletion can sometimes be markedness increasing on the position dimension. There is no place of articulation that is less marked than the absence of a consonant, so deletion of a consonant will always result in markedness reduction on the place of articulation hierarchy. Second, pressures exist favoring preservation of marked places of articulations over less marked places, but similar pressures favor preservation of unmarked positions (i.e. onsets) over more marked positions. These distinctions are encoded in the set of constraints assumed for the learning simulations in section 3, and will be shown to be responsible for the learning bias against simple patterns using only place of articulation. Another difference between these scales is that there are more members on the place of articulation scale than the syllable position scale. This difference will be returned to in section 4.1.

2.2 Typological Survey

To find the positional asymmetries in place of articulation contrasts, a survey was undertaken examining word-initial and word-final consonant inventories. Many previous typological databases (i.e. WALS (Dryer and Haspelmath 2013), PHOIBLE (Moran and McCloy 2019), etc.) have information on the syllable structure of a language and have information about the phonemic

³ Though cf. Morley (2015) who argues that consonant epenthesis may be less attested than de Lacy (2006) argued. Also see Rice (2008), who notes several languages that seem to show neutralization to labial or dorsal place, and Hume (2003) who argues against universal markedness hierarchies, and reviews cases of labial unmarkedness.

⁴ Cf. Breen and Pensalfini (1999) who argue that Arrente has obligatory codas and no onsets, though see Topintzi and Nevins (2017) for a recent reanalysis of Arrente.

inventory of a language. However no publicly available database that I was aware of at the time of writing explicitly examines the interaction of the place and position inventories—the inventory of consonants relative to different positions.

Rather than directly investigate onset and coda positions, I investigated consonant inventories possible in word-initial and word-final positions. While onset vs. coda position is the most common way of looking at a position scale, discovering the forms that are legal in onset and coda in a language is more challenging. Firstly, syllabification is not always apparent in word-medial positions. Word medial coda consonants must occur in a VCCV context, but distinguishing VC.CV syllabification from V.CCV can be difficult. If a [pr] cluster appears intervocalically it is ambiguous from this form (especially a transcribed rather than phonetic form) whether the cluster is a complex onset [V.prV] or if the [p] is in coda, [Vp.rV]. Such a determination must be made based on a larger language-specific picture, based on whether [pr] is a licit complex onset, whether [VpCV] clusters exist with less sonorous Cs, whether stress-weight interactions show that the syllable would be heavy, etc. Word-final consonants are unambiguously in word-final position, and are usually interpreted as codas.

Secondly, word-medial codas are vulnerable to assimilation processes in ways that word-final codas are not. For example, if a language has nasal place assimilation, the surface inventory of nasal stops may be larger in word-medial positions than word-finally. Word-final segments do not risk this confound in the same way, because a word-final segment sometimes surfaces before any possible onset in the language. As a result, the documentation usually reflects a neutral option rather than all possible forms.

Languages are coded according to what stops are available in word-final and word-initial position, rather than available in root initial or root final position, (or any other metric). I chose to focus on word-level restrictions over restrictions at the morpheme level because word-boundary positions are subject to similar phonological pressures in most languages, but roots are subject to a variety of different pressures based on the morphology of the language. For example, a language where classes of roots never appear without a vowel-initial suffix or theme vowel may be more likely to have consonant final roots than languages where roots tend to appear freely without suffixes, or languages with consonant-initial suffixes. Root-initial or final positions are not as likely to be linked to onset or coda positions as word-initial and final positions, and typological surveys of the phonotactics in those positions would require significant controlling for morphology, beyond the scope of this current paper. Further, at least some stages of early acquisition that are modeled by the learning simulations in this paper occur before children have morphologically parsed their lexicon fully (Hayes 2004), meaning that learners would not be aware of the difference.

2.2.1 Methodology

In this paper, I present the results from a genetically balanced sample based on the 100-language sample created by the editors of the World Atlas of Language Structure (WALS, Dryer and Haspelmath 2013). The original WALS sample is available online.⁵ The original sample was meant to maximize genealogical and areal diversity, however the editors note that there are five genera represented more than once in the sample. Before performing the survey, I randomly selected one language from each of these overrepresented genera so that the survey included 94 languages from 94 distinct genera.

For each language in the sample, I attempted to find some documentation of some variety⁶ of that language that listed which consonants were available in word-initial and word-final position.⁷ I was unable to find sufficient documentation for seven languages in the WALS sample, however for six of these languages, I was able to find documentation on a different language in the same genus, based on the classifications in either WALS or Ethnologue (Lewis et al. 2013).⁸ I was unable to find sufficient documentation on Rama (or any closely related Chibchan languages), so it was excluded from the sample.

In this paper, I will discuss only those languages that have no supralaryngeal places of articulation beyond [t], [p], and [k]. There are 48 languages in the sample that allow no more places of articulation than these three, and 46 that allow all three. Languages with just three place contrasts overwhelmingly contrast coronal, labial, and dorsal sounds. Languages with at least four places of articulation show significantly more variation. These additional places of articulation include uvular sounds [q], labiovelar sounds [kp] or [k^w], or contrastive minor coronal places of articulation, etc. I exclude the languages with more than three supralaryngeal place contrasts here because no particular inventory of more than three contrasts is well-attested enough to make reliable generalizations about asymmetries between word-initial and word-final position. Different places of articulation are subject to different phonetic pressures, and no particular four place inventory pattern is well enough attested to control for

⁵ Available at <https://wals.info/languoid/samples/100> as of August 14, 2019.

⁶ Different dialects or varieties of a language might have different restrictions on final consonants. Here, I used the first dialect for which I was able to find sufficient description. Sampling different dialects or descriptions for each language could result in slightly different typological counts, though it is highly unlikely that such resampling would affect the strong statistical skews noted below. Descriptions of languages may not match an individual speaker's grammar, but this type of noise is unlikely to skew the results here either.

⁷ In some cases, I was unable to find a description that explicitly listed the consonants available in each position, but I was able to infer positional stop inventories through sets of minimal pairs. This was done for Alambalak (Edmiston and Edmiston 2003), Arapesh (Mountain) (SIL 2011a), and Daga (SIL 2011b).

⁸ The six replacement languages were: Abun instead of Maybrat, Ma'anyan instead of Malagasy, Humburi Senni instead of Koyraboro Senni, Chontal Mayan instead of Jakaltek, Oromo (Mecha) [West-Central] instead of Oromo (Harar), and Sierra Popoluca instead of Zoque (Copainalá).

such pressures.⁹ If the documentation noted a minor place specification, but it was not contrastive (i.e. all coronal stops are dental in the language), the sound was classified as coronal labial or dorsal. Sources varied as to what the symbols ⟨c⟩ and ⟨tʃ⟩ represented. I classified the sounds as stops or affricates according to how they were described by the documentation.¹⁰

The presence or absence of the glottal stop [ʔ] is ignored in the analysis in this paper because a) glottal stops are more susceptible to specific positional effects (i.e. being banned from just onset or just coda) and b) the markedness relationship between glottal stops and [p], [t], and [k] is complex, without a clear typological implication (de Lacy 2006). As such, a language that allows [p], [t], [k], and [ʔ] in a position is treated the same as a language that allows [p], [t], and [k] but not glottal stop in that position in this survey, rather than being excluded from the survey like a language that allows [p], [t], [k], and [q].

Languages were coded according to their initial and final plain stop inventories.¹¹ If a segment was noted to be restricted to ideophones or loanwords, the language was treated as if that segment was absent. Further, if a stop was thought to exist on an abstract underlying level, but never surfaced it was not included. For example, in Finnish, former word-final /k/ is a ghost consonant that is apparent in compensatory lengthening, but never surfaces (Itkonen 1964; Keyser and Kiparsky 1984). For languages that had multiple series of stops with different laryngeal features, the series with the largest inventory of licit places of articulations was selected, in order to find the unmarked laryngeal series, avoiding potential mistranscription of aspiration vs. voicing contrasts (see Honeybone 2005). However, if a language banned [p] but allowed [b], it was treated as allowing labials, because such a gap is due to an interaction between place and voicing (Hayes and Steriade 2004; Flack 2007), rather than place markedness.

A list of all languages included in this study along with their sources and stop inventories is available in Appendix A.

⁹ Languages with more than three contrasting places of articulation for stops still appear to be largely subject to the all-or-nothing skew, allowing some but not all of the word-initial stops in word-final position at non-significantly higher rates than languages with only [p t k] word-finally. See REDACTED §2.4.3 for more details.

¹⁰ A language including an affricate [tʃ] was ignored for the purpose of our study, but containing a palatal stop [c] resulted in the language being excluded from the survey because it had more than three places of articulation for stops. Thorough investigation of whether sounds were stops or affricates, or whether palatals were allophonic or phonemic (see controversies around Greek palatals Householder 1964; Arvinito 2007), were beyond the scope of this survey. Instead I chose to trust the first description of a language I found. As a result, some languages' classifications may be disputable. However, debates about the nature of postalveolar segments and whether sounds are allophonic or phonemic are common, and I see no reason that the grammars I accessed would be biased to fall more on one side of that dispute than another.

¹¹ de Lacy (2006) (p. 21) notes that phonotactic generalizations, like those investigated here, are less clear evidence of actual phonological generalizations than synchronic alternations. Finding descriptions of synchronic alternations for a large genetically balanced sample of languages is significantly more difficult, so is beyond the scope of this paper.

2.3 Results

There are two crucial results from the study. First, in section 2.3.1, I show that the languages with [p], [t], and [k] word initially exhibit an all-or-nothing skew in word-final inventories. The simple patterns that can be defined using only syllable position are better attested than the more complex patterns that require interactions of position and place. Second, section 2.3.2 presents evidence that this simplicity bias does not extend symmetrically to the place of articulation hierarchy. Languages that only ban a place of articulation without making syllable position specific restrictions are underattested relative to those languages that ban positions without restricting place of articulation.

2.3.1 All-or-nothing skew in word-final position

Considering languages with [p], [t], and [k] word initially, there are four word-final inventories that are “harmonically complete” as defined by Prince and Smolensky (1993/2004); de Lacy (2006). An inventory of word-final stops is harmonically complete if any particular segment is allowed word-finally, so must every less marked stop. Table 1 (1-4) presents an example of a language with each pattern, along with (near) minimal triplets for each position. English (Table 1 1) allows all three of [p], [t], and [k] word finally, so I call it an ALL-FINAL language. Kiowa (Table 1 2) bans [k] word finally, but allows [p] and [t], so I call it a [PT]-FINAL language, because [p] and [t] are licit finally. Finnish (Table 1 3) allows just [t], so is called a [T]-FINAL language. Spanish does not allow any of these three stops word-finally, so is called a NO-FINAL language. There are also four possible non-harmonically complete inventories: those that ban one place of articulation word finally, but allow some more marked place of articulation. For instance, in Lavukaleve (Table 1 5) labial stops, such as [p^h] are banned word finally, but the more marked dorsal stop [k^h] is available word finally. Two of these patterns (the [tk]-Final pattern and the [pk]-Final pattern) are attested in the survey and are also included in Table 1.

The relative rates of attestation of these patterns are presented in Table 2, represented graphically in Figure 1. Languages with [p], [t], and [k] word-initially overwhelmingly prefer to allow all three places of articulation word-finally or none of them. The most common patterns are the All-Final and the No-Final patterns. Because the vast majority of languages in the sample exhibit either the All-Final or the No-Final pattern, there is not enough statistical power to be confident of any skews amongst the other patterns, whether if there is a skew favoring patterns with two places of articulation over one, or favoring the absence of particular places of articulation over others. However, some trends among these patterns may be observed. In this sample, there are more languages that allow two stops word-finally than just one. Further, the place of articulation markedness hierarchy appears to be reflected in attestation rate: the two harmonically complete patterns, the [pt]-Final pattern and the [t]-Final pattern, are better attested than inventories of the same size that are non-harmonically complete. It is important to note that patterns

Table 1: Example languages of each pattern

1. *Language with [p], [t], and [k] finally: English (All-Final)*

[pɑp]	pop	[tɑp]	top	[kɑp]	cop
[pɑp̚]	pop	[pɑt̚]	pot	[pɑk̚]	pock
2. *Language with [p] and [t] finally: Kiowa ([pt]-Final) (Watkins 1980)*

[pó:]	'eat'	[tó:]	'tepee'	[kí:]	'meat'
[sép]	'descend'	[kút]	'mark'	*[kúk]	
3. *Language with [t] finally: Finnish ([t]-Final)*

[pelata]	'to play'	[telata]	'to paint with a roller'	[kelata]	'to wind'
*[olup]		[olut]	'beer'	*[oluk]	
4. *Language with no stops word-finally: Spanish (No-Final)*

[popa]	'stern'	[topa]	'bump into'	[kopa]	'glass'
*[kasap]		*[kasat]		*[kasak]	
5. *Language with [t] and [k] word-finally: Lavukaleve ([tk]-Final) (Terrill 1999)*

[p ^h ara]	'larrikin'	[t ^h at ^h a]	'spider'	[k ^h ariala]	'easily'
*[k ^h eup ^h]		[k ^h eut ^h]	'skin'	[fa'luk ^h]	'cabbage'
6. *Language with [p] and [k] word-finally: Lakota ([pk]-Final) (Rood and Taylor 1996)*

[p ^h i]	'liver'	[t ^h i]	'to live'	[k ^h i]	'to reach home there'
[t ^h ob]	'four'	*[t ^h ot]		[pa't ^h ag]	'stopping'

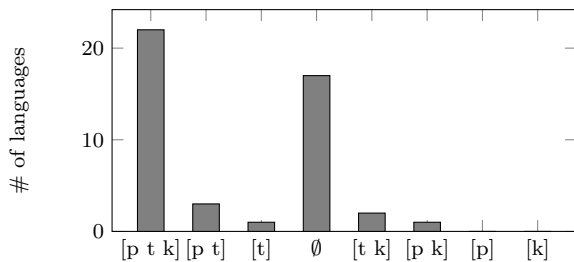
unattested in this sample are not necessarily unattested in any human language. However, it would be highly unlikely for the unattested patterns to be more common than either the All-Final pattern or the No-Final pattern as we consider more languages of the world. Of the 46 languages considered, 39 lan-

Table 2: Typological Survey: Languages with [p], [t], and [k] word-initially

Word-Final	Number	Languages
[p t k]	22	Abun, Alambak, Asmat, Canela-Krahô, Chamorro, Cree (Plains), Chontal Mayan, Daga, English, Georgian, Karok, Koasati, Korean, Kutenai, Lango, Ma'anyan, Meithei, Persian, Sierra Popoluca, Tagalog, Turkish, Yaqui
[p t]	3	Indonesian, Kiowa, Oromo (Mecha)
[t]	1	Finnish
[∅]	17	Apurinã, Arapesh (Mountain), Barasano, Fijian, Greek (Modern), Hixkaryana, Japanese, Kewa, Mandarin, Otomí (Mezquital), Pirahã, Quechua (Imbabura), Sanuma, Spanish, Supyire, Tukang Besi, Yagua
[t k]	2	Imonda, Lavukaleve
[p k]	1	Lakota
[p]	0	
[k]	0	
Total	46	

guages, or 84.8%, exhibit either the all-final or no-final patterns. If we assumed that each of these patterns was equally likely to be observed, we would expect

Fig. 1: Word-final stop inventory given [p t k] onsets



only 25% of the languages to exhibit either the all-final or no-final patterns, because there are six possible subset inventories and only two all-or-nothing inventories. Even if we assume subset inventories and all-or-nothing inventories are equally likely, languages that allow [p], [t], and [k] word-initially are significantly more likely to exhibit all-or-nothing inventories than any subset inventories ($\frac{39}{46}, p < .001$ by two-sided binomial test). Going further, I tested whether the presence or absence of each stop finally in the languages of the sample was correlated with the presence or absence of each other stop. The presence of [k] in the word-final inventory is not independent from the presence of [t] word-finally ($p < .001^{12}$). The same is true for [p] and [t] ($p < .001$), and [k] and [p] ($p < .001$). In other words, a language that lacks one of these stops word-finally is highly likely to lack the other two as well, and a language that allows one of the stops word-finally is highly likely to allow the other two. This sample therefore presents evidence for a structural typological skew: languages tend to make one-feature restrictions based on final position, rather than two-feature restrictions based on the interaction of position and place of articulation.

2.3.2 Positional priority skew

The magnitude of the all-or-nothing skew is similar to other skews observed in the distribution of place of articulation—Pater (2012) shows that 88% of the languages in UPSID have both or neither of [b] and [g]. Phonologists have recognized these type of skews for decades, arguing that languages tend to maximize feature economy (Martinet 1955; Clements 2003), symmetry (Hockett 1955), or structural complexity (Pater and Moreton 2012; Moreton et al. 2017; Seinhorst 2017). These different approaches differ subtly, but share the major prediction that languages tend to make restrictions based on one feature, rather than the interaction of features. The structural complexity approach directly ties into domain general cognitive biases favoring simple categories, observed in visual concept learning (Feldman 2006). In this section, I will show that the No-Finals pattern is equally complex as the Only-[pt] pattern, and

¹² All p-values testing pairwise independence of stops are from Fisher’s Exact Test for Count Data

show that despite this similarity in complexity, the typology exhibits a positional priority skew, favoring the No-Finals pattern over the Only-[pt] pattern.

For the purposes of this paper, the structural complexity of a pattern is based on how many parameters must be attended to in order to separate the forms into those that are licit and those that are banned.¹³ For example, the No-Finals pattern bans [Vp], [Vt], and [Vk], which can be reduced to a ban on [Final] stops. The place of articulation of the stop in a form is irrelevant to whether it is banned or not in the No-Finals pattern. On the other hand, in the [t]-Final pattern, both the place of articulation and the position of the stop in a form are necessary in order to determine whether or not it would be banned. The [t]-Final pattern bans [Vp] and [Vk], which at its simplest bans stops that are [Final] and [Dorsal] or [Final] and [Labial]. The No-Finals pattern can be assigned a complexity of 1, because it attends to the value on just one parameter. The [t]-Final pattern receives a complexity of 2, because both place and position are critical for distinguishing banned and licit forms.

Complexity is successful in predicting the All-or-Nothing skew observed above, as shown in Figure 2. The All-Final pattern has no restrictions and is maximally simple. Both the [pt]-Final and [t]-Final patterns require both place and position, and are significantly underrepresented relative to the No-Final and All-Final patterns.

Fig. 2: Complexity Example

(a) All-Final Simple <i>Complexity</i> = 0 $\frac{22}{46} = 47.8\%$	(b) [pt]-Final Complex <i>Complexity</i> = 2 $\frac{3}{46} = 6.5\%$	(c) [t]-Final Complex <i>Complexity</i> = 2 $\frac{1}{46} = 2.2\%$	(d) No-Final Simple <i>Complexity</i> = 1 $\frac{17}{46} = 40\%$
pV tV kV Vp Vt Vk	pV tV kV Vp Vt Vk	pV tV kV Vp Vt Vk	pV tV kV Vp Vt Vk

Structural complexity is blind to the substance of the parameters considered. A pattern that can be discriminated on the basis of position alone is equally complex as a pattern that can be discriminated by just place of articulation. Structural complexity therefore predicts that patterns that make use of only place of articulation should be similarly well attested as the No-Finals pattern. As a result, we would expect the Only-[pt] pattern presented in (7), which bans dorsal stops in all positions, but allows labial and coronal stops in all positions to be relatively well attested. However, this pattern is quite rare.

(7) *Only-[pt]: Simple. Dorsal stops are banned. Complexity=1.*

pV	tV	kV
Vp	Vt	Vk

¹³ This metric is a simplification of the algebraic complexity metric proposed by Feldman (2006). Both metrics perform the same on all of the complexity comparisons discussed here, but the more precise algebraic complexity metric can make finer-grained distinctions between patterns not compared here. For more details, see REDACTED.

Instead of showing that equally simple patterns that ban final stops and ban dorsal stops are equally frequent, the typological survey shows that languages are significantly more likely to ban all final stops than all dorsal stops. In other words, languages are more likely to make restrictions using word-position rather than place of articulation, which I call the positional priority skew. The WALS sample discussed above included 93 languages, none of which explicitly exhibit the Only-[pt] pattern. The simplicity metric cannot distinguish the Only-[pt] pattern from the pattern that allow only coronals and dorsals (Only-[tk]), or the pattern that allows only labials and dorsals (Only-[pk]). Neither of these patterns are attested in the WALS sample either. In fact, there are no languages in the WALS sample that clearly allow only two contrasting supralaryngeal places of articulation. If structural complexity was directly linked to the rate of attestation, we would expect that the Only-[pt] pattern (or similar patterns) would be attested at a similar rate to the No-Finals pattern. The difference in attestation between the relatively common No-Finals pattern (attested 17 times in WALS) and the unattested Only-[pt] pattern is enough to reject the hypothesis that they are similarly likely based on their similar complexities, based on the results of a two-sided binomial test ($\frac{17}{17}, p < .001$).

Two languages in the WALS sample exhibit a pattern similar to the Only-[tk] pattern. Oneida (Iroquoian: Julian 2010) and Wichita (Caddoan: Rood 1975) lack one of the major places of articulation in both initial and final positions. Both of these languages lack simple labial stops (i.e. [p] and [b]), but they both are argued to have a contrast between velar [k] and labiovelar [k^w] stops. Thus, both languages do still have three contrastive supralaryngeal places of articulation for stops, [t], [k], and [k^w], available word-initially and word-finally. The presence of the third contrastive stop makes these languages more similar to the All-Final pattern than the Only-[pt] pattern. Even if further analysis found that either or both of these languages treat [kw] as a cluster rather than a segment, there would only be two languages in the WALS sample that distinguish between licit and banned forms using only place of articulation, whereas there are seventeen that use only position. This difference is still significant by a two-sided binomial test ($\frac{17}{19}, p < 0.001$).¹⁴

Readers may be familiar with other languages not included in the WALS survey that do not allow one of the three supralaryngeal places of articulations for stops. Hawaiian, which does not have coronal stops, is perhaps the best known example. However, Hawaiian and languages like it are not examples of the Only-[pk] pattern, because stops are not allowed in final position in Hawaiian. Such languages offer no evidence for or against the positional priority skew examined here, because the pattern exhibited in them restricts both place of articulation and position. For more examples of such languages and further discussion, see REDACTED §2.6.2.

¹⁴ In fact, the difference between these categories remains significant with up to seven languages of the Only-[pt] type.

2.4 Summary of Results

In summary, two major results arise from the typological survey. First, we observe the presence of an all-or-nothing skew: there is a typological asymmetry within the languages with [p], [t], and [k] word-initially to allow all or none of those stops word-finally, resembling the predictions of a simplicity bias. Secondly, there is evidence for a positional priority skew: the Only-[pt] pattern, which allows no dorsals in either word-initial or word-final position, but allows word-final [t] or [p], is significantly underrepresented relative to the similarly simple No-Finals pattern. One would expect that the Only-[pt] pattern would be attested at a similar rate to the No-Finals pattern if there was a bias favoring simple patterns. The following section shows how a bias exists in the Perceptron with independently motivated constraints that could lead to both of these typological asymmetries.

3 Learning + Grammar Simulations

In this section I will demonstrate how agent-based generational models of language learning can make predictions about soft typology. In this case, I will show that a model that uses independently motivated constraints based on the markedness hierarchies on the place of articulation and position hierarchies interact with the Perceptron learning algorithm in order to capture two learning biases: one bias consistent with the All-or-Nothing skew, and one bias consistent with the Positional Priority skew. The model assumed is explicated in 3.1. The results of the generational simulations are presented in 3.2, and shown to capture the typological skews observed in section 2.

3.1 Modeling Learning Bias

The phonological learning problem is difficult. The phonological grammar is a mapping between underlying forms to surface forms, including which forms are licit and which ones are not. However, the learner is only directly exposed to licit surface forms. There is no direct information about illicit forms, and no direct information as to what the underlying forms are. The immense number of possible grammars further contributes to the learning challenge. Thus, efficient algorithms for this problem tend to have inherent biases favoring some grammars over others. Sometimes, these are directly (and intentionally) encoded by the designer of the algorithm (see the substantively biased learner of Wilson 2006). Often there are *emergent* biases present in the learning algorithm that are less apparent (and less intentional) in the design of the learning algorithms. Recent work has shown that common learning algorithms for phonological learning exhibit a variety of emergent biases ranging from favoring systemic simplicity (Pater 2012) to favoring deterministic grammars over variable grammars (Hughto 2018) to avoiding accidental gaps in the lexicon

(O’Hara 2017). Most any learning algorithm will have some emergent biases, and artificial language learning studies have shown some promise to test the cognitive reality of such biases (see Moreton and Pertsova 2014). Here, I will focus on the biases present in the commonly used MaxEnt framework with the Perceptron algorithm, showing that the emergent biases in this learning algorithm match the typological skew present in the typological survey in section 2. This paper uses a Generational Stability Model (GSM)¹⁵ which simulates the stability of different patterns across transmission over many generations, enhancing small differences in learnability of these patterns.

In the interest of concreteness, I will begin the discussion of learning biases by introducing the specific grammatical framework and learning algorithm used in this paper, though many of the principles discussed in this section hold for a large class of frameworks and learning algorithms.

3.1.1 MaxEnt Harmonic Grammar

The MaxEnt framework is a probabilistic implementation of a constraint-based grammar like Optimality Theory or Harmonic Grammar, in the sense that it assigns a probability distribution to the candidates in a tableau based on those candidates violations of a set of constraints. The probability of each candidate y for a given input form x , $P(y|x)$, is proportional to the exponential of its wellformedness (8). Wellformedness is measured via the Harmony score, the weighted sum of constraint violations the candidate incurs ($H(x, y)$ 9), where each constraint C is a function from an input-output pair (x, y) to a number of violations that that candidate incurs on the constraint (here all constraints will assign a non-positive integer, though this assumption is not necessary.)

(8) *Probability of a candidate y given input x in MaxEnt*

$$P(y|x) = \frac{e^{H(x,y)}}{\sum_{z \in \text{Cand}(x)} e^{H(x,z)}}$$

(9) *Harmony Score*

$$H(x, y) = \sum_{c \in \text{CON}} w_c * C(x, y)$$

3.1.2 Perceptron Algorithm

The learning algorithm implemented here is a truncated modification of the Perceptron algorithm (Rosenblatt 1958; Jäger 2007) proposed by Magri (2015). This algorithm is ERROR-DRIVEN in that the learner is exposed to one form at a time and attempts to match that form. If it makes an error, the learner updates its weighting of constraints in order to make the observed form more likely to be accurately matched when produced in the future.

Specifically, in the simulations in this paper, the learner is attempting to match the behavior of a teacher agent, which itself has a grammar mapping each input form to some probability distribution across output forms. At each

¹⁵ Also known as an Iterative Learning Model, (Kirby and Huford 2002; Hugto 2018).

step in the learning procedure, an input form x was sampled uniformly from the set of possible inputs $\{/pV/, /tV/, /kV/, /Vp/, /Vt/, /Vk/\}$, regardless of what the target grammar is.

Both the learner and the teacher select output forms for the input form x based on the probability distributions in the learner’s current grammar and the teacher’s grammar. Here, I denote the learner’s selected form as y_L and the teacher’s form as y_T . Then, the learning agent’s constraint weights were updated proportionally to the difference in violations between the learner’s form (y_L) and the teacher’s form (y_T) as seen in (10). When a constraint is violated more by the teacher than the learner, the weight of that constraint is increased, because violations are negative integers. When both teacher and learner select the same output, the difference in violations on each constraint is 0, so the weights do not change.

(10) *Perceptron Update Rule*

η is a learning rate constant between 0 and 1.
 $w_C(t + 1) = w_C(t) + \eta * (C(x, y_L) - C(x, y_T))$

There are two sources of stochasticity in this algorithm, the input form chosen at each iteration of learning is randomly selected, and the output form each agent selects for that input is randomly selected according to the probability distribution set by that agent’s MaxEnt grammar. These sources of randomness prevent the algorithm from being deterministic—any two learners have seen a slightly different distribution of input forms, and both the teacher and learner may have chosen different output forms for each input. As the number of learning iterations increases, it becomes vanishingly unlikely that any two learners would have identical weights. However, Fischer (2005) shows that if the learning rate (η) is sufficiently small, this algorithm is (weakly) convergent, meaning that the expected grammar of a learner (the average across many learners) will approach the target grammar as the number of learning iterations increases. However, if learners only receive a finite amount of training data from their target grammar, they may learn slightly different grammars than their target grammar.

3.1.3 Agent-Based Model

To model how small learning biases can skew typology, I make use of an agent-based model of iterated learning to simulate generational transmission following (Kirby and Huford 2002; Staubs 2014; Hughto 2018). Across generations, hard-to-learn patterns end up being more likely to change, and are less likely to be changed into, resulting in those patterns becoming less well attested typologically (Bell 1971; Greenberg 1978). In this model, each learner loosely represents a generation in a language community. A learner is initialized at the same starting grammar, and then exposed to a set number of randomly sampled forms from the target grammar of the teacher. After the learner has seen the set number of forms, it *matures*, a new learner is initialized, and is

trained with training data from whatever grammar the newly mature agent settled on. This method is repeated for a fixed number of generations, after which I analyzed the grammar learned by the final agent.

As an approximation of a language’s likelihood to change across generations, I measure the stability of each pattern across many generations. The stability of a pattern, and what it is likely to change into when it is unstable can be determined by performing a large number of runs of this generational model trained off of each pattern. A run qualifies as stable if the pattern learned by the final generation is the same as the original target pattern. For the purpose of this paper, each final grammar is binned according to the closest categorical pattern for several reasons. First, this allows better comparison to the typological survey, because very few of the grammars surveyed discussed variation, let alone reported quantitative data to allow me to confidently determine the observed rates of variation. Second, as noted by Hughto (2018), generational models show an emergent bias towards categorical patterns as the number of learning iterations increases. Investigation of simulation results further shows that if one generation learns a grammar closer to a different categorical pattern than its target pattern, it is very rare for a future generation to reverse this change, but it is quite common for one generation to become slightly less categorical and then the next generation to learn a more categorical pattern. Thus, binning each run based on the closest categorical pattern is the best way to compare the simulation results to the typological survey. This binning was performed by selecting the output candidate that has the highest harmony score for each input. This candidate receives more probability than any of its competitors. This binning procedure is the same as the Harmonic Grammar evaluation process. In other words, the closest categorical pattern for a set of weights is the same pattern that would be selected with those weights in standard Harmonic Grammar.

3.1.4 *The constraints*

The last aspect of the model is the structure of the grammar itself. This includes the set of competing candidates considered by the grammar, and the set of constraints used. As an abstraction, in the simulations performed here, each input form had three output candidates, a faithful candidate, a deletion candidate, and a debuccalization candidate. A larger, perhaps infinite, set of candidates would drastically increase computation time and make the internal dynamics of learning less apparent. These three candidates include the major common repairs for marked stops, for further discussion of other types of repairs such as epenthesis, see REDACTED §3.3.5.

In order to encode the place of articulation hierarchy, I will make use of stringently related constraints, following Prince (1999); de Lacy (2002); Kiparsky (1994). The three markedness constraints I included are included in (11, based on de Lacy (2002)’s place of articulation markedness constraints.

- (11) *Constraints encoding the Place of Articulation Hierarchy*

- a. *K—Incur a violation for each dorsal segment in the output.
- b. *KP—Incur a violation for each dorsal or bilabial segment in the output.
- c. *KPT—Incur a violation for each dorsal, bilabial, or coronal segment in the output.

The place of articulation hierarchy is crossed with a syllable position hierarchy between onsets and codas. However, as discussed in section 2.1, there seems to be a difference in the markedness of items on the position hierarchy compared to the place of articulation hierarchy. While the places of articulation differ in relative markedness, all are marked compared to the absence of any segment. Deletion even reduces the markedness of coronals. However, in the syllable position hierarchy, the presence of onsets is preferred to their absence, motivating the constraint ONSET, which marks onsetless syllables. In the small typology considered here, ONSET and NOCODA are somewhat redundant for purposes of factorial typology: ONSET causes segments to be protected in onset when they would delete in codas, and NOCODA drives codas to delete where they do not in onsets. This redundancy is important to capture the larger factorial typology, and is responsible for the asymmetries in how learning bias treats the place and positional features.

(12) *Constraints along the Syllable Position Hierarchy*

- a. NOCODA—Assign a violation mark for any syllable that has a consonant in coda position
- b. ONSET—Assign a violation mark for any syllable that lacks an onset consonant.

For the simulations presented here, I will consider two unfaithful candidates for each input, a deletion candidate and a debuccalization candidate. In the deletion candidate the input stop is deleted (i.e. /Vt/→[V]), and in the debuccalization candidate, it is debuccalized to a glottal stop (i.e. /Vt/→[Vʔ]).¹⁶ To prevent stops from deleting, I use the faithfulness constraint MAX (13).

- (13) MAX—Assign a violation mark for any segment in the input without an output correspondent.

Using just deletion, MAX and the markedness constraints discussed above, all of the patterns consistent with the markedness hierarchies can be modeled in Harmonic Grammar (and even ranked OT). This is because deletion resolves violations of markedness constraints based on position (like NOCODA) and place of articulation (like *K).

To pressure against debuccalization, I include a general IDENT[PLACE] constraint (14). However, debuccalization does not resolve violations of NOCODA. This means that with only a general faithfulness constraint IDENT[PLACE], it

¹⁶ While I use [ʔ] here, this debuccalization candidate is meant to represent a large class of repairs that do not delete the marked consonant, including debuccalization to [h], gemination with following consonants, or potentially changes in manner.

is impossible to capture languages where debuccalization occurs only in coda. Thus, I also include a *specific* faithfulness constraint, IDENT/ONSET, which is a positional faithfulness constraint following Beckman (1998) (15).

- (14) IDENT[PLACE] —Assign a violation mark for any segment in the input that has an output correspondent that does not have the same place of articulation as the input.
- (15) IDENT/ONSET[PLACE]—Assign a violation mark for any segment in the input that has an output correspondent in the onset of a syllable that does not have the same place of articulation as the segment in the input.

I also included specific faithfulness constraints along the place of articulation dimension. Unlike position where marked (or unprivileged) positions are less faithful than unmarked (or privileged) positions, marked places of articulation appear to be more faithful than unmarked places of articulation. In order to capture asymmetries in place assimilation patterns and to capture gapped inventories that do not directly follow the markedness hierarchy, de Lacy (2002, 2006) introduces “marked faithfulness” constraints. Marked faithfulness constraints protect marked forms from changing and form a family of stringency related constraints parallel to those used to encode the place hierarchy. In effect, this adds two additional specific faithfulness constraints, IDENTK and IDENTKP, which are respectively violated when input dorsals change place of articulation, and when input dorsals or labials change place of articulation. The third constraint in the family (16b), IDENTKPT, is identical to the general IDENT constraint.

- (16) *Marked Faithfulness constraints for Place of Articulation Scale*
- a. IDENT{X} Assign a violation for a segment on the input that has a place of articulation within the set *X* that does not have that same place of articulation in the output.
- b. IDENT{K}, IDENT{KP}, IDENT{KPT}

Together, the constraints reviewed in this section (and repeated in (17)) are sufficient to generate all the patterns observed in the typological survey in section 2.

- (17) **Constraints used:**
- a. *Markedness Constraints:* *K, *KP, *KPT, ONSET, NoCODA.
- b. *Faithfulness Constraints:* MAX, IDENTKPT, IDENTKP, IDENTK, IDENT/ONSET

This set of constraints predicts 108 distinct categorical grammars in Harmonic Grammar according to OT-Help (Staub et al. 2010). Many of these grammars differ only in whether unfaithful forms delete or debuccalize. Since this paper is focused on distributional patterns rather than input-output mappings, these 108 grammars can be binned into 27 distinct distributional patterns. Most of these 27 patterns are unattested—some for obvious reasons (a

Table 3: (Categorical) Patterns Predicted by Factorial Typology. Patterns highlighted in gray are simulated in this paper.

Name	Licit Forms						Attested?	Example
	pV	tV	kV	Vp	Vt	Vk		
No-Stops	✗	✗	✗	✗	✗	✗	✗	
[t]-Initial	✗	✓	✗	✗	✗	✗	✗	
[pt]-Initial	✓	✓	✗	✗	✗	✗	✓	Tahitian
No-Final	✓	✓	✓	✗	✗	✗	✓	Barasano
[t]-Final	✓	✓	✓	✗	✓	✗	✓	Finnish
[pt]-Final	✓	✓	✓	✓	✓	✗	✓	Kiowa
All-Final	✓	✓	✓	✓	✓	✓	✓	Tagalog
Only-[t]	✗	✓	✗	✗	✓	✗	✗	
[pt]-Initial+[t]-Final	✓	✓	✗	✗	✓	✗	✗	
Only-[pt]	✓	✓	✗	✓	✓	✗	✗	

language with no place contrasts for stops at all is extremely impractical for communication), but some (like the Only-[pt] pattern discussed in section 2) are more surprising. Ten of these patterns fully respect the place of articulation markedness hierarchy, that is, if some place of articulation is banned in a position, all more marked places of articulation are also banned in that position. These ten patterns are presented in Table 3. The remaining 17 patterns are “gapped patterns” where a more marked place is present in a position where a less marked place is banned, e.g. the pattern where final [p] is banned but final [k] is preserved (observed in Imonda and Lavukaleve). Further study of these gapped inventories is beyond the scope of this paper.

For the simulations presented here, I will focus on five patterns. These five patterns are lightly shaded in Table 3. Patterns that allow all three of [p], [t], and [k] word initially are named according to their final stop inventory, i.e. [pt]-Final. Patterns that allow only a subset of [p], [t], and [k], but allow all initially available stops word-finally are named Only-[x], i.e. Only-[pt]. Patterns that allow only a subset of initial stops are named according to their initial inventory, i.e. [pt]-Initial. Each of these patterns can be generated given the constraints discussed above, but the constraints alone say little about the prevalence of these patterns.¹⁷ To see that the all-or-nothing skew discussed in section 2.3.1 emerges from the generational simulations, I examine the four hierarchy-respecting patterns that have three places of articulation available word-initially: All-Final, [pt]-Final, [t]-Final, and No-Final. I will also run simulations on the Only-[pt] pattern to test the positional priority skew favoring simple patterns defined using position rather than place of articulation

3.2 Simulation Results

The Generational Stability Model was run using each of the highlighted patterns in Table 3 as categorical target patterns for the first generation. Each

¹⁷ Though see the predictions of r-volume approaches in Section 4.2.

target grammar was used to train a learner for 3600 learning iterations, with a learning rate of .05, after which that learner’s current grammar was used to train the next generation for 3600 learning iterations, and so on and so forth for 25 generations. After 25 generations, the final learner’s grammar was classified according to the closest categorical pattern, by finding the Harmonic Grammar winners for each input. One hundred runs of the simulation were performed with each target grammar to acquire transitional probabilities from each categorical pattern to the patterns it might change into over time.

In this paper, following a common assumption in the phonological learning literature, all learners are initialized with the same starting grammar. Each learner begins with the markedness constraints weighted high (at 50), and the faithfulness constraints weighted low (at 1). This initial weighting serves to promote learning of restrictive patterns, and best simulates intermediate stages of acquisition (Gnanadesikan 2004; Tesar and Smolensky 2000; Jesney and Tessier 2011).

The setting of these parameters affect the relative stability of all patterns: increasing the number of learning iterations per generation makes all patterns more stable, as does decreasing the number of generations. The relative stability between patterns remains the same across a variety of parameter settings. These parameter settings were chosen relatively arbitrarily with four goals in mind: learners should (usually) be closer to the categorical target pattern than any other pattern after the first generation, learning rate is very small relative to initial markedness constraint weights to ensure weak convergence, simulation time was relatively short (<8 hours), and patterns are not all categorically stable or unstable.

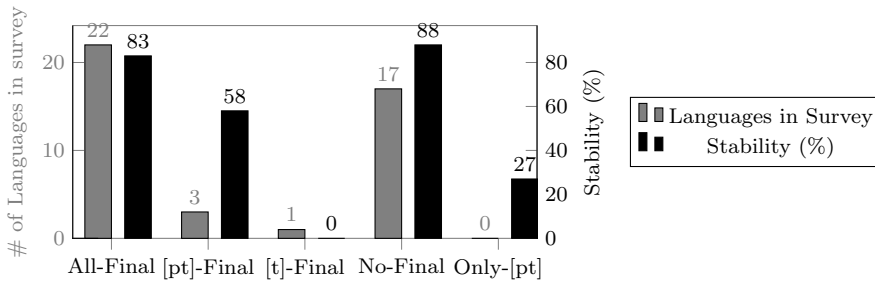
With markedness constraints initialized high and faithfulness initialized low, learners start with a near categorical grammar that deletes codas and debuccalizes onsets. At each new generation, the new learner starts with this grammar. Learners are initialized with this starting grammar regardless of the target pattern they are being trained on.

(18) *The grammar of an initialized learner*

Input	/pV/	/tV/	/kV/	/Vp/	/Vt/	/Vk/
Output	[?V]	[?V]	[?V]	[V]	[V]	[V]

The first generation of each run was trained off of a categorical grammar that exemplified the target pattern. If a form was banned in the target pattern, the training grammar deleted it if it was in final position, and debuccalized it if it was in onset position, reflecting the initial state of a learner. This prevents learners from disagreeing with the target grammar on the unfaithful mapping. The simulations in this paper abstract away from the learning of underlying representations by providing the teacher’s intended input form to the learner at each stage. A more realistic learner would need to induce the teacher’s intended input through algorithms like lexicon optimization (Tesar and Smolensky 2000; Inkelas 1995) is beyond the scope of this paper. For patterns where some forms are not licit, such as the [pt]-Final pattern which bans final [Vk], the target grammar mapped the illicit form to its deletion

Fig. 3: Well attested patterns are stable in the simulations, but poorly attested ones are not.



candidate if it were a final stop, and its debuccalization candidate for initial stops. As a result, when the initial teacher produces an unfaithful mapping it matches the learner’s initial unfaithful mapping, making learners unlikely to make updates where both teacher and learner select different unfaithful mappings for the same input. This is to replicate the fact that learners seeing the teacher produce an unfaithful output form (say /Vk/→[V?]) would be unable to reconstruct the underlying form /Vk/ without sufficient evidence from alternations.

The Python code utilized to run these simulations is available at [REDACTED].

3.2.1 Pattern Stability

First I will examine the stability of the patterns in question. A run is classified as stable if the weights learned by the last grammar would generate the original target pattern in Harmonic Grammar. Figure 3 shows how the stability of a pattern is correlated to how well attested that pattern was in the typological survey. The two patterns that were most common in the typological survey, No-Final (88% stable) and All-Final (83% stable) were both stable more than 80% of the time. The inventories that require interactions of place and position ([pt]-Final (58% stable) and [t]-Final (0% stable)) were both found to be less stable, with the more common [pt]-Final pattern being more stable than the very rare [t]-Final pattern. Further, we can see that even though the Only-[pt] pattern is as structurally simple as the No-Final pattern, it is far less stable (27% stable compared to 88% stable). Pairwise comparisons via Fisher’s Exact Test between each of these rates show significant differences between the stability of any two patterns (with $p < 0.0001$) except the All-Final and No-Final pattern ($p = 0.422$). Typologically rare patterns are learned significantly less stably than common patterns.

3.2.2 Likely patterns of change

The stability of a pattern contributes to its attestation, but the flip side of stability is also important for typology: what do unstable patterns change

into? In these simulations, the direction of change is very consistent. In these simulations, change occurs when the learner does not fully reach the target pattern. With markedness high and faithfulness low, learners start with all forms being banned. Incomplete learning involves learning only a subset of the target patterns licit forms.

Table 4 shows the probability of a run with a specific target pattern ending up with each pattern after 25 generations. No run ever innovates new forms. Instead, unstable runs lose forms, starting with a licit form that is maximally marked on some dimension.¹⁸ For the All-Final, No-Final, and Only-[pt] patterns, there is a single form that has the most marked place of articulation licit in the pattern and the most marked position. Unstable runs trained on the All-Final pattern lose final [Vk], which is the most marked place of articulation and position available. Unstable runs trained on the No-Final pattern lose initial [kV], which is the most marked place of articulation and the most marked position available in the target grammar. Runs trained on Only-[pt] lose final [Vp].

However, for the [t]-Final and [pt]-Final patterns, there is not a single form that is maximally marked on each hierarchy. In these cases, position takes priority over place of articulation. For example, in the [pt]-Final pattern, [Vp] is more marked than [kV] on the position dimension, but less marked on the place dimension. They even violate the same number of markedness constraints: [Vp] violates *KP, *KPT and NoCODA, whereas [kV] violates *KPT, *KP, and *K. Loss of [kV] would change the [pt]-Final pattern into the simple Only-[pt] pattern, whereas loss of [Vp] changes it into the equally complex [t]-Final pattern. Yet, unstable runs of the [pt]-Final pattern consistently change first into the [t]-Final pattern—and then eventually into the No-Final pattern, due to the extreme instability of the [t]-Final pattern.

This prioritizing of position over place of articulation is consistent with the underattestation of the Only-[pt] pattern. The [pt]-Final pattern is the only pattern that could lose one form to turn into the Only-[pt] pattern, but it consistently changes into the [t]-Final pattern instead. The underattestation of the Only-[pt] pattern is predicted because it is not only relatively unstable, but also unlikely to be innovated through change of another pattern, even though it is a structurally simple pattern.

3.2.3 Generational dynamics reflect how the first generation is learned

These simulations reflect the two major observed typological skews discussed in Section 2. First, the All-Final and No-Final patterns are stabler than the two patterns with a subset of final stops ([t]-Final and [pt]-Final), reflecting the typological skew favoring patterns that were defined only using syllable position and not place of articulation. Second, the Only-[pt] pattern is not only unstable, but lacks any paths where other patterns would be mislearned

¹⁸ Noticeably, no runs involved a learner changing into a gapped inventory, such as one where [p] is banned word finally but [k] is not, even though that pattern could be generated by the set of constraints.

Table 4: Transitional Probabilities from Generational Simulations with debuccalization.

Initial Pattern	Final Pattern	[t]-Initial	[pt]-Initial	No-Final	[t]-Final	[pt]-Final	All-Final	[pt]-Initial+[t]-Final	Only-[pt]
No-Final		.01	.11	.88	0	0	0	0	0
[t]-Final		.01	.08	.91	0	0	0	0	0
[pt]-Final		0	0	.23	.19	.58	0	0	0
All-Final		0	0	0	.01	.16	.83	0	0
Only-[pt]		.04	.58	0	0	0	0	.11	.27

as one of them. Thus, learners are unlikely to innovate a simple pattern that uses only place of articulation to define its restriction.

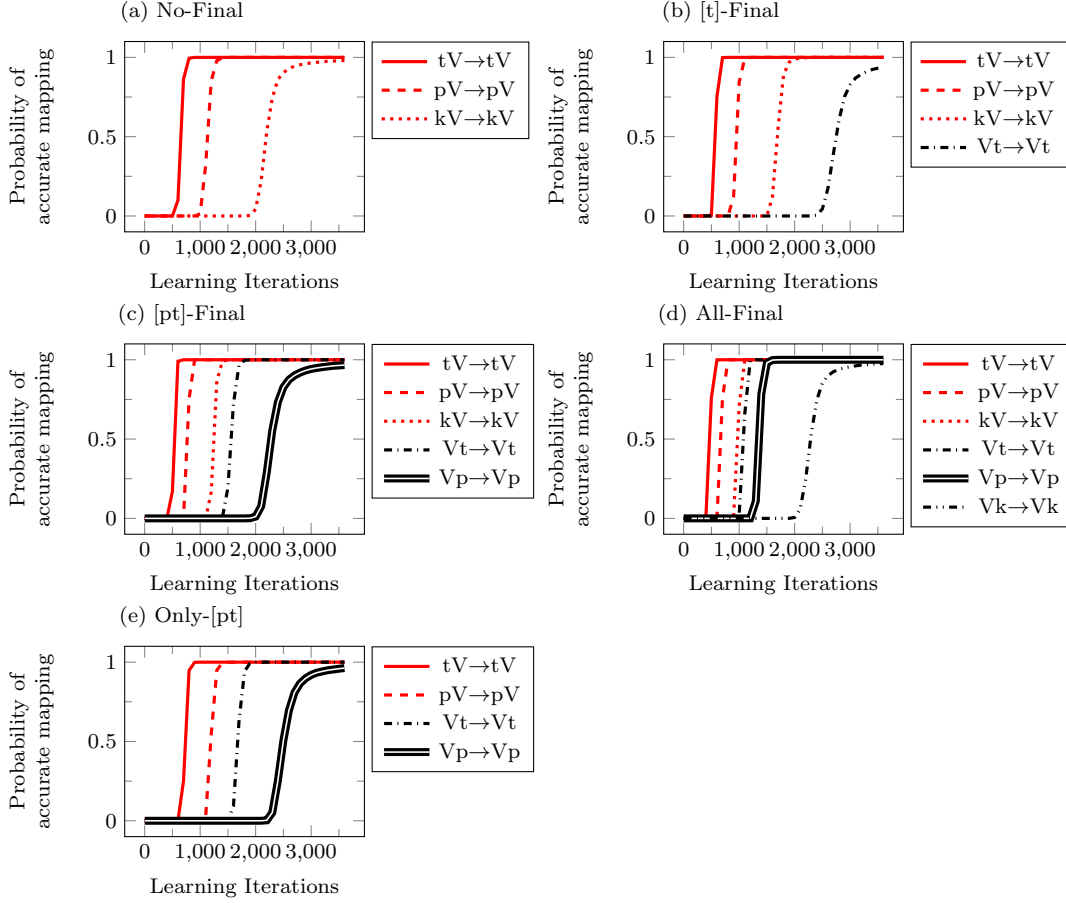
The generational behavior of the simulations can be separated into two fundamental components: the stability of a pattern, and what that pattern changes into when it is unstable. Both of these properties are predictable from the behavior of the average learner in the first generation. When patterns change, they lose the form that is on average learned last. The more quickly learners learn the last form of a pattern, the more stable that pattern is.

Figure 4 shows the learning path of the first generation learner training off of the target pattern, averaged over 100 runs. In learners trained on each pattern, all licit stops in initial position (colored in red) are learned before any stops in final position. In each position, stops are learned in order of place of articulation. This learning order is consistent across runs. To see that these orders are very unlikely to change across runs, I looked at the /kV/→[kV] mapping and the /Vt/→[Vt] mapping when trained on the All-Final pattern. Because the All-Final pattern has more final stops speeding the learning of final [Vt] than any other pattern under consideration, the expected gap in learning iterations between when [kV] is learned and [Vt] is learned is the smallest of the patterns investigated here. Thus, if noise between runs ever subverted the learning order, we would expect to see it in the All-Final pattern.

I found the number of learning iterations it took for each form to pass the threshold of .9 probability in the All-Final pattern. There was a significant difference via pairwise t-test in the number of learning iterations needed to learn /kV/ (M=1071, SD=47.8) and the number of learning iterations needed to learn /Vt/ (M=1173, SD=44.6) conditions; $t(99) = -13.8, p < .001$. Thus, we can see that learning order is quite consistent across runs. As a result, runs are highly unlikely to differ in direction of change across generations.

The last form learned by a learner is also the least categorically learned form. As a result, the last learned form is the first form to be lost across generations. Because the learning order is so consistent in these simulations,

Fig. 4: Learning paths for five different patterns (averaged over 100 runs). The legend of each plot shows the mappings in the order that they are learned. (Unfaithful mappings that remain at one throughout the learning process are omitted for ease of reading.)



it is completely predictable which forms will be lost across generations if the target pattern is unstable.

How likely a pattern is to be stable is linked to how categorically that last learned form is learned. In turn, the accuracy of the last learned form is associated with how many learning iterations it typically takes for that form to be learned to some degree of certainty. By comparing how many learning iterations are needed to learn the last learned form of each pattern, it can be predicted how stable the pattern is relative to other patterns.

Table 5 presents the average number of learning iterations needed to learn each pattern. For each run, I checked the probability of each form on every 100th learning iteration. I then found the first learning iteration after which the

Table 5: The average number of learning iterations required to learn all forms in the target grammar with more than .9 probability.

Pattern	No-Final	All-Final	[pt]-Final	Only-[pt]	[t]-Final
Mean Iterations Needed	2518	2609	2716	2929	3237
Stand Dev	142	131	133	134	185
Stability %	88	83	58	27	0

learner assigned all forms in the target grammar a probability of at least .9. In effect, this picks out the point at which the last learned form in each grammar receives .9 probability. The No-Final pattern is learned fastest, followed by the All-Final pattern, then the [pt]-Final pattern, the Only-[pt] pattern and the [t]-Final pattern. Each comparison between patterns is highly significant via an independent samples t-test ($p < .001$).

3.3 Discussion

The learning simulations presented in this section show a learning bias favoring patterns that ban all or no final stops (resembling the all-or-nothing skew observed in the typology), and a bias favoring patterns that ban all places of articulation in one position (No-Finals) over those patterns that ban one place of articulation in all positions (Only-[pt]), also replicating the skew observed in the typology. It is worth noting that with this relatively simple model, all patterns are expected to change across generations, and change only moves one way—towards loss of licit stops. This suggests that the process of learning stop inventories creates a pressure towards loss of sounds, but there are a number of possible factors not included in this model here that could lead to acquisition of new sounds across generations. If learners were able to come into contact with other agents who had lost a form (whether speakers of the same “language” or not), they could regain stops that had been lost in previous generations. A further elaborated system could model other phonological processes such as final vowel deletion which could create evidence for new final stops that had been lost. However, adding the possibility of final vowel deletion to this system is unlikely to hurt the fit of this model—the possibility of final vowel deletion would cause more patterns to change into the All-Final pattern, creating a sort of circuit between the highly stable No-Final and All-Final patterns. This model also has no overt pressure towards patterns that are communicative—after a high number of generations, the consistent loss of sounds would result in a completely impractical language. However, this sort of slippery slope towards a maximally simple language is quite typical in this sort of generational agent-based learning model that focuses on cross-generation learning and not communication between agents. Elaborated models that involve interaction between agents during learning appear to show emergent homophony avoidance and pressures towards expressiveness (see Kirby (2017) for a general overview, Pater and Moreton (2012) section 2.2 for an example

from phonology and Hughto (2019) on the differences between interactive and iterated MaxEnt models of phonology).

It is further worth noting that these learning simulations show that these skews *could* be the result of a learning bias, rather than that they *must* be the result of a learning bias. Learning bias is not the only factor that could be responsible for skews in the typology: phonetic factors can favor certain patterns of misperception between speaker and hearer (i.e. channel bias Moreton 2008; Blevins 1993), the patterns of language contact, migration, and death could cause certain patterns (particularly those found in languages spoken by major imperialistic forces) to be overrepresented, and other factors. Teasing apart the typological impact of each of these factors on typology is beyond the scope of this paper, but will require an understanding of each of the extent of the capability of each of these biases. This paper intends to show how far learning bias alone can get us.






4 Alternatives

In order to ensure that the learning bias captured by the interactive learning and grammar model in the previous section is indeed caused by an interaction of both the learning algorithm and the structure of the grammar projected from a set of substantively biased constraints, these results must be compared to a model that focuses on learning and a model that focuses on grammar. In this section, the predictions of a model similar to the GMECCS model, which does not have substantively biased constraints, and thus only uses featural simplicity (Moreton and Pater 2012b), and the predictions of the r-volume model, a model that does not use learning, will be considered. The first alternative isolates the effect of learning on typology, whereas the second isolates the effects of the structure of the grammar on typology. The failure of both of these alternatives shows the importance of the interaction of these factors to capture the typological skews under investigation. These alternatives further elucidate the interaction of initial distance and speed to shape typology. The substance-free model only can distinguish between patterns based on update speed, and all forms have the same initial markedness difference. On the other hand, the results of the r-volume approach are inherently correlated to the initial markedness difference values with the set of constraints considered in this paper.

4.1 Learning prioritized over grammar: Structural Only

Pater and Moreton (2012) discover a structural bias in MaxEnt learners. If the learner is given a CON of constraints including general and specific ones, but lacking any hierarchical structure like the markedness hierarchies used in this paper, the patterns that are consistent with general constraints are preferred. Such a model should be able to capture the all-or-nothing skew, because it is a

Table 6: Stability of Patterns Given Substantively Unbiased Constraints

	Initial			Final			Stability (%)	
No-Final	pV	tV	kV	✗	✗	✗	76	
[t]-Final	pV	tV	kV	✗	Vt	✗	52	
[pt]-Final	pV	tV	kV	Vp	Vt	✗	35	
All-Final	pV	tV	kV	Vp	Vt	Vk	100	
Only-[pt]	pV	tV	✗	Vp	Vt	✗	89	

purely structural skew, but will have a harder time capturing the substantive positional priority skew.

In order to show the predictions of this model, I ran generational stability simulations as before on the same set of five patterns, but with an unbiased set of constraints (19).¹⁹ These constraints are defined as to create no substantive bias in the grammar. Therefore, for any constraint NOCODA or *T, there is an inverse constraint CODA or +T, that reward the behavior the other constraint marked. There are also no markedness hierarchies encoded in this constraint set—dorsal, coronal, and labial are simply different labels for three features that are computationally the same as far as the learner and the grammar are concerned. 100 runs were performed with each pattern. Learners were initialized with MAX and IDENT at zero, but all other constraints at 50. Each generation was exposed to 2500 forms²⁰, and each run went for 20 generations.

- (19) *T, *P, *K, *KT, *KP, *PT, *KPT, +T, +P, +K, +KT, +KP, +PT, +KPT, ONSET, NOONSET, CODA, NOCODA, MAX, IDENT

The structurally simple patterns are learned more stably because the general constraints are more likely to be promoted than specific constraints on any error, and each time a general constraint is promoted it affects the harmony of more forms in the grammar than a specific constraint would. With this set of constraints, the all-or-nothing bias with regard to syllable codas is captured. Given [p], [t], and [k] in onset, the bias prefers having all or none of those stops in coda. This makes sense because these patterns are more structurally simple than the patterns that have a non-empty subset of [p], [t], and [k] in coda.

¹⁹ The model explored here differs in some major ways from the GMECCS model used in Pater and Moreton (2012); Moreton et al. (2017), in order to closer parallel the models used elsewhere in this paper. First, the GMECCS model uses a phonotactic MaxEnt grammar that defines the relative probability of all word forms, where this model maps inputs to outputs, thus no faithfulness would be used in such a model. The GMECCS algorithm also allows for both negative and positive constraint weights. The combination of these factors is responsible for the need to have both positive and negative constraints for each form.

²⁰ Because in this model all forms initially have the same harmony score, learners initially assign equal probability to all possible output forms. This means that the learners start closer to the target grammar than they do in the simulations from the previous section, where learners initially produce the least marked candidate for each input near categorically. As a result, each form can be learned in relatively fewer learning iterations compared to the model with substantively biased constraints. Thus, fewer learning iterations are used here for both computational speed and to increase the probability of any mistransmission.

However, under this model, the simple Only-[pt] pattern is also found to be stable. Because it is simpler, it is preferred to the pattern that only bans dorsals in coda [pt]-Final. This is the opposite of the result predicted by the model with substantively biased constraints presented in the previous section, which finds the Only-[pt] pattern that bans dorsals in all positions to be maximally unstable, while the language that only bans final [Vk] is intermediately stable. The typological data is better matched by the model with substantively biased constraints than this model with unbiased constraints.

While not shown directly here, this model also has issues capturing implicational relations between places of articulation or syllable positions. Since there is no substantive bias in the grammar, all places of articulation are in effect equivalent, and interchangeable, as are onset and coda. A language that bans all onsets but allows all codas is as stable as a language that bans all codas. Moreton and Pater (2012b) suggest that the difference in attestedness between these two patterns could best be explained through another medium, say channel bias. Future work will explore if there is a way for channel bias to replace the substantively biased constraints used in section 3.

4.2 Grammar without learning: R-volume

Another alternative makes predictions about the soft typology of a grammar while focusing only on the space of possible hypotheses, without making assumptions about the direct effects of the learning algorithm on soft typology.

In constraint based grammars, a single phonological pattern can be described by multiple discrete rankings of constraints. As a result, certain phonological patterns can be derived from *more* rankings of constraints than others. The number of rankings (or weightings) that model a given pattern is proportional to the volume of the total grammatical hypothesis space that results in such a pattern. Thus, the proportion of rankings that model some pattern is that pattern's *r-volume*. It has long been hypothesized that the relative number of rankings that results in a particular pattern may be able to explain some probabilistic aspects of phonology. The relative number of rankings that result in a one output candidate vs another has been used to model variation between those candidates in synchronic grammars (Kiparsky 1993; Anttila 1997), and the number of rankings that result in a particular pattern over another has been argued to be responsible for relative rates of attestation in the soft typology of such patterns (Coetzee 2002; Bane and Riggle 2008). Because weighted constraint systems have infinite possible grammars, the r-volume can be approximated by randomly sampling across the possible weights (Bane and Riggle 2009; Carroll 2010).

Results from this approach have ended up mirroring those observed through learning based models of soft typology. Bane and Riggle (2008) show that the r-volume for a given stress pattern is correlated to its typological frequency, which Staubs (2014) shows emerges from a MaxEnt model similar to the one structurally-based model used in this paper. Hughto et al. (2014); Hughto and

Table 7: Predicted attestation rates based on R-volume

	Initial			Final			Percentage
No-Final	pV	tV	kV	✗	✗	✗	14.1
[t]-Final	pV	tV	kV	✗	Vt	✗	5.1
[pt]-Final	pV	tV	kV	Vp	Vt	✗	2.1
All-Final	pV	tV	kV	Vp	Vt	Vk	.4
Only-[pt]	pV	tV	✗	Vp	Vt	✗	.4

Pater (2017) show that an iterated learning MaxEnt model disprefers patterns that require gang effects to be modeled, though Bane and Riggle (2009) and Carroll (2010) note that patterns that require gang effects tend to have relative small r-volumes. Is it possible that both the stability factors discovered in the simulations and the cross-typology attestation rates are simply reflecting the fact that certain patterns have greater r-volumes than others?

To approximate the volume of weightings that model each pattern, 1000 random constraint weight vectors were sampled from a uniform distribution between 0 and 50, using the `runif` function in R (R Core Team 2014). In MaxEnt, the probabilistic nature of the grammar makes it highly unlikely that any two sets of weights predict the exact same grammar. Like in the other simulations in this paper, weight vectors were classified based on which form received the plurality of the probability, matching the Harmonic Grammar choice function.²¹ The constraints used were those used in the simulations in section 3, repeated in (20). Again here, I am including both debuccalization and deletion as possible repairs for marked structure, but simulations with just deletion performed similarly.

(20) *Constraints Used in R-Volume Calculation*

*K, *KP, *KPT, MAX, ONSET, NOCODA, IDENTKPT, ID/ONS, IDENTK, IDENTKP

I present the percentage of the r-volume that can be attributed to each of the five patterns of focus in this paper in Table 7.²² From these patterns we can see that the all-or-nothing skew cannot be explained via r-volume alone because the All-Final pattern has the smallest r-volume of the five patterns, lower than the [t]-Final and the [pt]-Final patterns. If r-volume with this set of constraints modeled the all-or-nothing skew, we would expect the All-Final pattern to have a larger r-volume than those two patterns, and have a r-volume more similar to the No-Final pattern. On the other hand, the Only-[pt]

²¹ Using the Harmonic Grammar choice function, the space of possible constraint weightings is separated into different patterns by planes that pass through the origin, because if all constraints are weighted at zero, all candidates have the same harmony score. Thus, regardless of what the upper bound is, the relative volume of each region bounded by these planes is the same, so the choice of the upper bound on constraint weights does not effect the relative r-volume of any of these patterns.

²² A full table including the r-volume of each pattern that was sampled in the r-volume simulations can be found in the appendix B. Most of the remaining r-volume consists of the patterns that lack all stops word-finally and at least one of the three stops word-initially.

pattern has a much smaller r-volume than the No-Final pattern. This result is consistent with the positional priority skew observed in the typology.

One reason the r-volume of the All-Final pattern is so small is because with this set of constraints, there are far more markedness constraints that can be satisfied by deletion than faithfulness constraints militating against deletion. In order to get the All-Final pattern, the constraint weights must be such that /Vk/ maps to [Vk] more than [V]. This mapping requires that MAX is higher weighted than the sum of NOCODA, *K, *KP, and *KPT. This condition is only met when MAX is weighted quite high, and each of those constraints are weighted quite low. Imagine the weighting of each constraint is a die roll. In order for [Vk] to occur, one die needs to roll better than four dice added together. In order for this to happen, MAX must be weighted very high, and/or all the markedness constraints must be weighted very low.

The positional priority skew is observed in the r-volumes because there are more ways to capture the No-Final pattern than to capture the Only-[pt] pattern. Simplifying matters, in the Only-[pt] pattern *K must be particularly high and active. The other markedness constraints must be relatively low so that [p] and [t] are produced in onsets and codas, but *K must be high weighted enough so that (when summed with the weights of *KP and *KPT) it surpasses faithfulness constraints. On the other hand, there are multiple ways to derive the No-Finals pattern: it could be caused by NOCODA being weighted high, ONSET being weighted high, or IDENT/ONSET being weighted high.

One might wonder if it would be possible to modify the constraint set in order to capture the all-or-nothing skew. The lack of the all-or-nothing skew was particularly dependent on the fact that there is just one MAX constraint. If we duplicated MAX a few times, the r-volume results could change substantially, without affecting the factorial typology of generable patterns. If we added more MAX constraints, it would be easier for the faithfulness constraints to outweigh the sum of markedness constraints that militate against final [Vk]. If more constraints were adding together to protect [Vk], [Vk] would surface more. However, it seems likely that this would only serve to shift the skew in the r-volume from small inventories to big inventories, rather than being able to capture the all-or-nothing bias. As we increase the number of constraints militating against deletion, patterns with a good amount of deletion, such as No-Final, will become less likely.

With this set of constraints, r-volume fails to capture the all-or-nothing skew, even though the all-or-nothing skew was emergent in generational stability simulations using this set of constraints in section 3. This shows that the learnability of a pattern in this framework is not only determined by its r-volume, but some interaction of the constraint set and the learning algorithm determine which patterns are easy or hard to learn. The r-volume may be partially able to explain the Positional Priority Skew, because the Only-[pt] pattern has a much smaller r-volume than the equally similar No-Final pattern. However, the Only-[pt] pattern, which was unattested in our typological sample and relatively hard to learn in the learning simulations in section 3 has a similar r-volume to the well-attested and easy to learn All-Final pattern.

Without understanding the interaction of learning and grammar, it is unclear why the All-Final pattern’s learnability surpasses its r-volume so significantly, but the Only-[pt] pattern’s does not so much.

4.3 Summary

To summarize, I compare the predictions of each of the models of soft typology investigated here and compare them to the real typology. The positional priority skew observed in the typological data is predicted by the interactive model in section 3 and by the r-volume of those constraints, but not by the Structural Bias model, which tests the learnability of patterns with an unbiased set of constraints. On the other hand, the all-or-nothing skew observed in the typological data is observed in the interactive model and the Structural Bias model, but not the r-volume model. Further, I tested the correlations between each of these models and the observed typological counts via linear regressions. The Interactive model had the closest fit with a R^2 of 0.72, but the other two models fit the typology far worse, with the structural bias model having an R^2 of 0.33, and the r-volume model having an R^2 of 0.09. The evi-

Table 8: Comparison of the three models to the typology

Model	Positional Priority skew	All-or-nothing skew	R^2
Interactive Model (§3)	✓	✓	0.72
Learning: Structural Bias (§4.1)	✗	✓	0.33
Grammar: r-volume (§4.2)	✓	✗	0.09
Typology (§2)	✓	✓	

dence suggests that more languages of the world favor the position hierarchy above the place hierarchy—i.e. languages that ban dorsals initially, lack final stops. This is evidence favoring the Generational Stability Model to the simplicity bias hypothesis. There is evidence that languages that allow all or no places in coda are preferred to the non-empty proper subsets; this is evidence favoring the Generational Stability Model to the R-Volume hypothesis.

5 Conclusion

This paper has shown two typological skews present in the soft typology of positional stop inventories, the all-or-nothing skew and the positional priority skew. Languages tend to show either all or none of the stops available in word-initial position in word-final position, and languages tend to make restrictions with position rather than with place of articulation. The first of these can be

cast as a structural skew, but the second one is a substantive skew. Both of these skews were shown to be predicted by learning biases in MaxEnt learners that were given independently motivated constraints to encode markedness hierarchies on each of these dimensions, but they were not both predicted by the r-volume of that same set of constraints, nor by the learning biases in MaxEnt learners with an unbiased constraint set. This finding suggests that this skew might be best explained through an integrated model that includes inductive biases based on the learner, the grammar and their interaction, rather than a model that focuses on deriving inductive biases from either grammar or learning.

More generally, this finding shows that changes in the set of constraints given to a MaxEnt learner can greatly impact the relative learnability of different patterns, potentially affecting soft typology. Further, this finding highlights how important considering learnability is when attempting to describe typology with constraint-based grammars. The factorial typology (and the r-volume) are not sufficient to predict which phonological patterns are likely to be attested. When a phonologist seeks to add or modify a constraint a theory of universal CON, they should also consider how different theories of CON may predict different patterns are more or less learnable. To cast the finding in a different light, this paper argues that the limitations of a synchronic grammar can impact the types of diachronic change that occurs (Kiparsky 2008).

This paper also shows a case where two orthogonal sets of substantively-grounded constraint hierarchies interact to make a substantive learning bias favoring one feature over another, without encoding that bias directly in the constraints. This result raises multiple questions. Can the positional priority skew be cast in a way that is phonetically natural? Is it likely to emerge from channel biases? Beyond this particular skew however, this paper raises additional questions regarding substantive biases that are not clearly phonetically grounded. Here we saw that the interaction of substantively-grounded constraints and learning could create new substantive biases. Some approaches (Pater and Moreton 2012) would argue that the typological skews markedness hierarchies encoded in the substantively grounded constraints are the result of channel biases rather than innate constraint hierarchies. Models of channel bias can be integrated into the types of generational learning models used here to explain typological skews with less needing to be directly encoded in the grammar itself. Can the same set of non-phonetically natural substantive biases be captured by the interaction of substantive channel biases as the interaction of substantively grounded constraints?

A Information about Language Survey

Table 9: Languages in typological survey

Language	Family	Genus	Source	Initial	Final
Abun	West Papuan	North-Central Bird's Head	Berry and Berry (1999)	p t k	p t k
Alamblak	Sepik	Sepik Hill	Edmiston and Edmiston (2003)	p t k	p t k
Apurina	Arawakan	Purus	Facundes (2000)	p t k	0
Arapesh (Mountain)	Torricelli	Kombio-Arapesh	SIL (2011a)	p t k	0
Asmat	Trans-New Guinea	Asmat-Kamoro	Voorhoeve (1965)	p t k	p t k
Barasano	Tucanoan	Tucanoan	Jones and Jones (1991)	p t k	0
Canela-Krahô	Macro-Ge	Ge-Kaingang	Fopjes and Fopjes (1986)	p t k	p t k
Chamorro	Austronesian	Chamorro	Seiden (1960)	p t k	p t k
Chontal Mayan	Mayan	Mayan	Knowles (1984)	p t k ?	p t k ?
Cree (Plains)	Algic	Algonquian	Wolfart (1996)	p t k	p t k
Daga	Dagan	Dagan	SIL (2011b)	p t k	p t k
English	Indo-European	Germanic	Personal Judgement	p t k	p t k
Fijian	Austronesian	Oceanic	Dixon (1988)	p t k	p t k
Finnish	Uralic	Finnic	Harms (1964)	p t k	t
Georgian	Kartvelian	Kartvelian	Butskhrikidze (2002)	p t k	p t k
Greek (Modern)	Indo-European	Greek	Romeo (1964)	p t k	0
Hixkaryana	Cariban	Cariban	Maddieson et al. (2014-2019)	p t k	0
Imonda	Border	Border	Seiler (1985)	p t k	t k
Indonesian	Austronesian	Malayo-Sumbawan	Lapoliwa (1981)	p t k	p t
Japanese	Japanese	Japanese	Maddieson et al. (2014-2019)	p t k	0
Karok	Karok	Karok	Bright (1957)	p t k	p t k
Kewa	Trans-New Guinea	Engan	Franklin (1971)	p t k	0
Kiowa	Kiowa-Tanoan	Kiowa-Tanoan	Watkins (1980)	p t k	p t
Korean	Korean	Korean	Chang (1996)	p t k	p t k
Kutenai	Kutenai	Kutenai	Garvin (1948)	p t k	p t k
Lakota	Siouan	Core Siouan	Rood and Taylor (1996)	p t k	p k
Lango	Eastern Sudanic	Nilotic	Noonan (1992)	p t k	p t k
Lavukaleve	Solomons East Papuan	Lavukaleve	Terrill (1999)	b t k	t k
Ma'anyan	Austronesian	Barito	Gudai (1985)	p t k ?	p t k ?
Mandarin	Sino-Tibetan	Chinese	Lin (2001)	p t k	0
Meithei	Sino-Tibetan	Kuki-Chin	Thoudam (1980)	p t k	p t k
Oneida	Iroquoian	Northern Iroquoian	Julian (2010)	t k k ^w	t k k ^w
Oromo (Mecha) [West-Central]	Afro-Asiatic	Lowland East Cushitic	Tola (1981)	p t k	p t
Otomí (Mezquital)	Oto-Manguean	Otomian	Wallis (1968)	p t k	0
Persian	Indo-European	Iranian	Alamolhoda (2000)	p t k	p t k
Pirahã	Mura	Mura	Everett (1986)	p t k	0
Quechua (Imbabura)	Quechuan	Quechuan	Cole (1985)	p t k	0
Sanuma	Yanomam	Yanomam	Borgman (1990)	p t k	0
Sierra Popoluca	Mixe-Zoque	Mixe-Zoque	de Jong Boudreault (2009)	p t k	p t k
Spanish	Indo-European	Romance	Maddieson et al. (2014-2019)	p t k	0
Supyire	Niger-Congo	Gur	Carlson (2011)	p t k	0
Tagalog	Austronesian	Greater Central Philippine	Llamzon (1976)	p t k	p t k
Tukang Besi	Austronesian	Celebic	Donohue (2011)	p t k	0
Turkish	Altaic	Turkic	Kornfilt (1997)	p t k	p t k
Wichita	Caddoan	Caddoan	Rood (1975)	t k k ^w	t k k ^w
Yagua	Peba-Yaguan	Peba-Yaguan	Payne (1985)	p t k	0
Yaqui	Uto-Aztecan	Cahita	Dedrick and Casad (1999)	p t k	p t k

B Estimated r-volumes of each pattern.

Table 10: Predicted attestation rates based on R-volume

	Initial			Final			Percentage
No-Stops	✗	✗	✗	✗	✗	✗	22.2
[t]-Initial	✗	tV	✗	✗	✗	✗	28.9
[pt]-Initial	pV	tV	✗	✗	✗	✗	17.2
No-Final	pV	tV	kV	✗	✗	✗	14.1
[t]-Final	pV	tV	kV	✗	Vt	✗	5.1
[pt]-Final	pV	tV	kV	Vp	Vt	✗	2.1
All-Final	pV	tV	kV	Vp	Vt	Vk	.4
Only-[t]	✗	tV	✗	✗	Vt	✗	1.6
[pt]-Initial+[t]-Final	pV	tV	✗	✗	Vt	✗	4.8
Only-[pt]	pV	tV	✗	Vp	Vt	✗	.4
[p]-Initial	✗	pV	✗	✗	✗	✗	0.9
[k]-Initial	✗	✗	kV	✗	✗	✗	0.5
[tk]-Initial	tV	✗	kV	✗	✗	✗	0.2
[pk]-Initial	✗	pV	kV	✗	✗	✗	0.6
[p]-Final	pV	tV	kV	✗	Vp	✗	0.3
[pk]-Final	pV	tV	kV	✗	Vp	Vk	0.4
[pk]-Initial	pV	✗	kV	✗	✗	✗	0.6
[pk]-Initial+[p]-Final	pV	✗	kV	✗	Vp	✗	0.1
Only-[p]	pV	✗	✗	✗	✗	Vp	0.1

References

- Alamolhoda, Seyyed Morteza. 2000. *Phonostatistics and phonotactics of the syllable in modern Persian*, Vol. 89. Finnish Oriental Society.
- Anttila, Arto. 1997. Deriving variation from grammar. In *Variation, change and phonological theory*, eds. Frans Hinskens, Roeland van Hout, and Leo Wetzels, 35–68. Philadelphia, PA: John Benjamins.
- Anttila, Arto. 2008. Gradient phonotactics and the complexity hypothesis. *Natural Language & Linguistic Theory* 26 (4): 695–729.
- Arvinita, Amalia. 2007. Greek phonetics: The state of the art. *Journal of Greek Linguistics* 8 (1): 97–208.
- Bane, Max, and Jason Riggle. 2008. Three correlates of the typological frequency of quantity-insensitive stress systems. In *Proceedings of the tenth meeting of the ACL special interest group on computational morphology and phonology*.
- Bane, Max, and Jason Riggle. 2009. The typological consequences of weighted constraints. In *CLS 45*, Vol. 45, 13–27.
- Beckman, Jill N. 1998. Positional faithfulness. PhD diss, University of Massachusetts Amherst, Amherst.
- Bell, A. 1971. Some patterns of the occurrence and formation of syllabic structure. In *Working papers on language universals*, Vol. 6, 23–138.
- Berry, Keith, and Christine Berry. 1999. *A description of Abun: a West Papuan language of Irian Jaya*. Pacific linguistics.
- Blevins, Juliette. 1993. Klamath laryngeal phonology. *International Journal of American Linguistics* 59 (3): 237–279.
- Blevins, Juliette. 1995. The syllable in phonological theory. In *Handbook of phonological theory*, ed. John Goldsmith, 206–244. Blackwell.
- Blevins, Juliette. 2004. *Evolutionary phonology: the emergence of sound patterns*. Cambridge University Press.
- Boersma, Paul, and Joe Pater. 2016. Convergence properties of a gradual learning algorithm for Harmonic Grammar. In *Harmonic Grammar and Harmonic Serialism*, eds. John J.

- McCarthy and Joe Pater. Equinox.
- Borgman, Donald M. 1990. Sanuma. *Handbook of Amazonian languages 2*: 15–248.
- Breen, Gavan, and Rob Pensalfini. 1999. Arrernte: A language with no syllable onsets. *Linguistic Inquiry* 30 (1): 1–25.
- Bright, William. 1957. The Karok language.
- Bruner, J. S., J. J. Goodnow, and G. A. Austin. 1956. *A study of thinking*. John Wiley and Sons.
- Butskhrikidze, Marika. 2002. The consonant phonotactics of Georgian. PhD diss, Netherlands Graduate School of Linguistics.
- Carlson, Robert. 2011. *A grammar of Supyire*, Vol. 14. Walter de Gruyter.
- Carroll, Lucien. 2010. Cumulativity in constraint grammars: An iterated learning view of probabilistic typology. Manuscript UCSD.
- Chang, Suk-Jin. 1996. *Korean*, Vol. 4. John Benjamins Publishing.
- Clements, G. N. 2003. Feature economy in sound systems. *Phonology* 20: 287–333.
- Coetzee, Andries. 2002. Between-language frequency effects in phonological theory. ms. University of Massachusetts Amherst.
- Cole, Peter. 1985. *Imbabura Quechua*. Israeli Hebrew.
- de Jong Boudreault, Lynda J. 2009. *A grammar of Sierra Popoluca (Soteapanec, a Mixe-Zoquean language)*. ERIC.
- de Lacy, Paul. 2002. The formal expression of markedness. PhD diss, University of Massachusetts Amherst, Amherst, MA.
- de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology*. Cambridge: Cambridge University Press.
- Dedrick, John M, and Eugene H Casad. 1999. *Sonora Yaqui language structures*. University of Arizona Press.
- Dixon, R. M. W. 1988. *A grammar of Boumaa Fijian*. The University of Chicago Press.
- Donohue, Mark. 2011. *A grammar of Tukang Besi*, Vol. 20. Walter de Gruyter.
- Dryer, Matthew S., and Martin Haspelmath, eds. 2013. *WALS online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/>.
- Edmiston, Pat, and Mel Edmiston. 2003. Alambak Organised Phonology Data. SIL Language and Culture.
- Everett, Daniel. 1986. Piraha. In *Handbook of Amazonian Languages*, eds. Desmond C. Derbyshire and Geoffrey K. Pullum, Vol. I. Mouton de Gruyter.
- Facundes, Sidney da Silva. 2000. The Language of the Apurinã People of Brazil (Maipure/Arawak).
- Feldman, Jacob. 2006. An algebra of human concept learning. *Journal of Mathematical Psychology* 50: 339–368.
- Finley, Sara. 2012. Typological asymmetries in round vowel harmony: Support from artificial grammar learning. *Language and Cognitive Processes* 27: 1550–1562.
- Fischer, Markus. 2005. A Robbins-Monro type learning algorithm for an entropy maximizing version of stochastic optimality theory. Master's thesis, Humboldt University, Berlin.
- Flack, Kathryn. 2007. The sources of phonological markedness. PhD diss, University of Massachusetts Amherst, Amherst.
- Franklin, Karl James. 1971. *A grammar of Kewa, New Guinea. Pacific linguistics, series c*. Canberra: Australian National University.
- Garvin, Paul L. 1948. Kutenai i: phonemics. *International Journal of American Linguistics* 14 (1): 37–42.
- Glewwe, Eleanor. 2019. Substantive bias in phonotactic learning: Positional extension of an obstruent voicing contrast. In *Proceedings of CLS 53*. Chicago, IL.
- Gnanadesikan, Amalia. 2004. Markedness and faithfulness in child phonology [ROA-67]. In *Fixing priorities: Constraints in phonological acquisition*, eds. René Kager, Joe Pater, and Wim Zonneveld, 73–108. Cambridge: Cambridge University Press.
- Goldsmith, John A. 1990. *Autosegmental and metrical phonology*. Blackwell.
- Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a Maximum Entropy model. In *Proceedings of the workshop on variation within optimality theory*. Stockholm University.
- Greenberg, Joseph H. 1978. Diachrony, synchrony, and language universals. In *Universals of human language, volume 1 method and theory*, eds. Joseph H. Greenberg, C. A.

- Ferguson, and E. A. Moravcsik, 61–91. Stanford, CA: Stanford University Press.
- Gudai, Darmansyah H. 1985. A grammar of Maanyan: A language of Central Kalimantan. PhD diss, Australian National University.
- Harms, R. T. 1964. *Finnish structural sketch. Indiana University publications. Uralic and Altaic series.* Indiana University. <https://books.google.com/books?id=a8kYAQAIAAJ>.
- Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: the early stages. In *Fixing priorities: Constraints in phonological acquisition*, eds. René Kager, Joe Pater, and Wim Zonneveld. Cambridge: Cambridge University Press.
- Hayes, Bruce, and Donca Steriade. 2004. Introduction: The phonetic basis of phonological markedness. In *Phonetically-based phonology*, eds. Bruce Hayes, Robert Kirchner, and Donca Steriade, 1–32. Cambridge: Cambridge University Press.
- Hayes, Bruce, and James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic Inquiry* 44: 45–75.
- Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39: 379–440.
- Hockett, Charles F. 1955. *A manual of phonology.* Baltimore, Maryland: Waverly Press.
- Honeybone, Patrick. 2005. Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications. In *The internal organization of phonological segments*, 317–352. Mouton de Gruyter.
- Householder, Frederic W. 1964. Three dreams of modern greek phonology. *Word* 20 (3): 17–27.
- Hughto, Coral. 2018. Investigating the consequences of iterated learning in phonological typology. In *Proceedings of the society for computation in linguistics*, Vol. 1.
- Hughto, Coral. 2019. Emergent typological effects of agent-based learning models in maximum entropy grammar. PhD diss, University of Massachusetts Amherst.
- Hughto, Coral, and Joe Pater. 2017. Emergence of strict domination effects with weighted constraints. Talk given at the CLS Workshop on Dynamic Modeling in Phonetics and Phonology.
- Hughto, Coral, Robert Staubs, and Joe Pater. 2014. Typological consequences of agent interaction. Presented at the Northeast Computational Phonology Circle (NECPhon 8).
- Hume, Elizabeth. 2003. Language specific markedness: The case of place of articulation. *Studies in Phonetics, Phonology and Morphology* 19 (2): 295–310.
- Hunt, E. B., and C. I. Hovland. 1960. Order of consideration of different types of concepts. *Journal of Experimental Psychology* 59 (4): 220–225.
- Inkelas, Sharon. 1995. The consequences of optimization for underspecification. In *Nels*, Vol. 25, 287–302. <http://roa.rutgers.edu/files/40-1294/40-1294-INKELAS-0-0.PDF>.
- Itkonen, Terho. 1964. *Proto-Finnic final consonants: Their history in the Finnic languages with particular reference to the Finnish dialects.* Vol. 138 of *Suomalais-ugrilaisen seuran toimituksia.* Helsinki.
- Itô, Junko. 1986. Syllable theory in prosodic phonology. PhD diss, University of Massachusetts Amherst, Amherst.
- Jäger, Gerhard. 2007. Maximum entropy models and stochastic optimality theory. In *Architectures, rules, and preferences: A festschrift for Joan Bresnan*, 467–479. CSLI.
- Jakobson, Roman, and Morris Halle. 1956. *Fundamentals of language.* The Hague: Mouton.
- Jesney, Karen, and Anne-Michelle Tessier. 2011. Biases in harmonic grammar: The road to restrictive learning. *Natural Language & Linguistic Theory* 29.
- Jones, Wendell, and Paula Jones. 1991. *Barasano syntax. Publications in linguistics.* Summer Institute of Linguistics and The University of Texas Arlington.
- Julian, Charles. 2010. A history of the Iroquoian languages. PhD diss, University of Manitoba.
- Jun, Jongho. 1995. Perceptual and articulatory factors in place assimilation: an optimality theoretic approach. PhD diss, University of California, Los Angeles.
- Kean, Mary-Louise. 1975. The theory of markedness in generative grammar. PhD diss, Massachusetts Institute of Technology.
- Keyser, S. J, and Paul Kiparsky. 1984. Syllable structure in Finnish phonology. In *Language sound structure*, eds. Mark Aronoff and Richard T. Oehrle, 7–31. Cambridge, MA: MIT

- Press.
- Kimper, Wendell. 2016. Asymmetric generalisation of harmony triggers. In *Proceedings of the 2015 Annual Meeting on Phonology*, eds. Gunnar Hansson, Ashley Farris-Trimble, Kevin McMullin, and Douglas Pulleyblank.
- Kingston, John. 1985. The phonetics and phonology of the timing of oral and glottal events. PhD diss, University of California, Berkeley.
- Kiparsky, Paul. 1993. Variable Rules. Handout distributed at the Rutgers Optimality Workshop.
- Kiparsky, Paul. 1994. Remarks on Markedness. Presented at TREND 2.
- Kiparsky, Paul. 2008. Universals constrain change; change results in typological generalizations. In *Linguistic universals and language change*, ed. Jeff Good, 23–53. Oxford University Press.
- Kirby, Simon. 2017. Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin and Review* 24: 118–137.
- Kirby, Simon, and James Huford. 2002. The emergence of linguistic structure: An overview of the iterated learning model. In *Simulating the evolution of language*, eds. A Cangelosi and D. Parisi, 121–148. London: Springer. Chap. 6.
- Knowles, Susan Marie. 1984. A descriptive grammar of Chontal Maya: San Carlos dialect. PhD diss, Tulane University.
- Kornfilt, Iaklin. 1997. Turkish. *London and New York: Routledge*.
- Lapoliwa, Hans. 1981. *A generative approach to the phonology of Bahasa Indonesia*. Dept. of Linguistics, Research School of Pacific Studies, Australian National University for Linguistic Circle of Canberra Canberra.
- Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky. 1990. Harmonic Grammar - a formal multi-level connectionist theory of linguistic wellformedness: an application. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, ed. Lawrence Erlbaum, 884–891. Cambridge, MA.
- Legendre, Géraldine, Antonella Sorace, and Paul Smolensky. 2006. The optimality theory-harmonic grammar connection. In *The harmonic mind: From neural computation to optimality-theoretic grammar*, eds. Paul Smolensky and Géraldine Legendre, 339–402. Cambridge, MA: MIT Press.
- Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig, eds. 2013. *Ethnologue: Languages of the world, seventeenth edition*, 17 edn. Dallas, Texas: SIL International. <http://www.ethnologue.com/language/kl>.
- Lin, Hua. 2001. *A grammar of Mandarin Chinese*, Vol. 344. Lincom Europa.
- Lin, Yu-Leng. 2016. Sonority effects and learning bias in nasal harmony. PhD diss, University of Toronto.
- Llamzon, Teodoro A. 1976. *Modern Tagalog: A functional-structural description*. The Hague: Mouton.
- Lombardi, Linda. 1998. Coronal epenthesis and unmarkedness. In *University of Maryland working papers in linguistics*, Vol. 5, 156–175.
- Maddieson, Ian, and Kristin Precoda. 1992. The UPSID database. UCLA.
- Maddieson, Ian, Sébastien Flavie, Egidio Marsico, and François Pellegrino. 2014–2019. LAPSyD: Lyon-Albuquerque Phonological Systems Databases, Version 1.1. <http://www.lapsyd.dcl.cnrs.fr/lapsyd/>.
- Magri, Giorgio. 2015. How to keep the HG weights non-negative: the truncated perceptron reweighting rule. *Journal of Language Modeling* 3 (2): 345–375.
- Martin, Alexander, and Sharon Peperkamp. 2020. Phonetically natural rules benefit from a learning bias: a re-examination of vowel harmony and disharmony. *Phonology* 37 (1): 65–90.
- Martinet, André. 1955. *Économie des changements phonétiques*. Berne: Francke.
- McCarthy, John J., and Alan Prince. 1995. Faithfulness and reduplicative identity. *University of Massachusetts Occasional Papers* 18: 249–384.
- Moran, Steven, and Daniel McCloy. 2019. PHOIBLE 2.0. (Available online at <http://phoible.org>, Accessed on 2019-09-04.).
- Moreton, Elliott. 2008. Analytical bias and phonological typology. *Phonology* 25: 83–127.
- Moreton, Elliott, and Joe Pater. 2012a. Structure and substance in artificial-phonology learning part I, structure. *Language and Linguistics Compass* 6 (11): 686–701.

- Moreton, Elliott, and Joe Pater. 2012b. Structure and substance in artificial-phonology learning part II, substance. *Language and Linguistics Compass* 6 (11): 702–718.
- Moreton, Elliott, and Katya Pertsova. 2014. Pastry phonotactics: Is phonological learning special? In *Proceedings of NELS 43*, eds. Hsin-Lun Huang, Ethan Poole, and Amanda Rysling, Vol. II, 1–12.
- Moreton, Elliott, Joe Pater, and Katya Pertsova. 2017. Phonological concept learning. *Cognitive Science* 41 (1): 4–69. doi:10.1111/cogs.12319. <http://dx.doi.org/10.1111/cogs.12319>.
- Morley, Rebecca. 2015. Deletion or epenthesis? on the falsifiability of phonological universals. *Lingua* 154: 1–26.
- Noonan, Michael. 1992. *A grammar of Lango*. Vol. 7 of *Mouton grammar library*. Mouton de Gruyter.
- O’Hara, Charlie. 2017. How abstract is too abstract: Learning abstract underlying representations. *Phonology* 34 (2): 325–345.
- Paradis, Carole, and Jean-François Prunet. 1991. Introduction: asymmetry and visibility in consonant articulations. In *The special status of coronals: internal and external evidence*, eds. Carole Paradis and Jean-François Prunet. Vol. 2 of *Phonetics and phonology*. Academic Press.
- Pater, Joe. 2012. Serial Harmonic Grammar and Berber syllabification. In *Prosody matters: Essays in honor of Elisabeth O. Selkirk*, eds. Toni Borowsky, Shigeto Kawahara, Takahito Shinya, and Mariko Sugahara, 43–72. Equinox Press.
- Pater, Joe. 2016. Universal grammar with weighted constraints. In *Harmonic Grammar and Harmonic Serialism*, eds. John J. McCarthy and Joe Pater, 1–46. London: Equinox.
- Pater, Joe, and Elliott Moreton. 2012. Structurally biased phonology: complexity in language learning and typology. *The EFL Journal* 3 (2): 1–44.
- Payne, Doris Lander. 1985. *Aspects of the grammar of Yagua: A typological perspective*.
- Pertsova, Katya. 2012. Linguistic categorization and complexity. In *Proceedings of the twelfth meeting of the acl-sigmorphon: Computational research in phonetics, phonology and morphology*, 72–81.
- Popjes, Jack, and Jo Popjes. 1986. Canela-Krahô. In *Handbook of Amazonian languages 1*, eds. Desmond C. Derbyshire and Geoffrey K. Pullum, 128–199. Berlin: Mouton de Gruyter.
- Prickett, Brandon. 2018. Complexity and naturalness biases in phonotactics: Hayes and White (2013) revisited. In *Proceedings of the 2017 Annual Meeting on Phonology*, eds. Gillian Gallagher, Maria Gouskova, and Sora Yin. Washington, DC: Linguistics Society of America.
- Prince, Alan. 1999. Paninian relations. Handout, University of Marburg.
- Prince, Alan, and Paul Smolensky. 1993/2004. *Optimality theory: Constraint interaction in generative grammar*. Oxford: Blackwell.
- R Core Team. 2014. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rice, Keren. 2008. Review of de Lacy 2008: Markedness: Reduction and preservation in phonology. *Phonology* 25 (2): 361–371.
- Romeo, Luigi. 1964. Toward a phonological grammar of Modern Spoken Greek. *Word* 20 (sup2): 60–78.
- Rood, David S. 1975. The implications of Wichita phonology. *Language* 51 (2): 315–337.
- Rood, David S., and Allan R. Taylor. 1996. Sketch of Lakhota, a Siouan language. In *Handbook of North American Indians*, ed. Ives Goddard, Vol. 17: Languages, 440–482. Washington: Smithsonian Institute.
- Rosenblatt, F. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review* 65: 386–408.
- Seiden, W. 1960. Chamorro phonemes. *Anthropological Linguistics* 2 (4): 6–35.
- Seiler, Walter. 1985. *Imonda, a Papuan language*. Department of Linguistics, Research School of Pacific Studies, The Australian National University.
- Seinhorst, Klaas T. 2017. Feature economy vs. logical complexity in phonological pattern learning. *Language Sciences* 60: 69–79.
- Shepard, R. N., C. I. Hovland, and H. M. Jenkins. 1961. Learning and memorization of classifications. *Psychological Monographs: General and Applied* 75 (13): 1–42.

- SIL. 2011a. Bukiyip Organised Phonological Data. SIL Language and Culture.
- SIL. 2011b. Daga Organised Phonology Data. SIL Language and Culture Archives.
- Stanton, Juliet. 2016. Learnability shapes typology: the case of the midpoint pathology. *Language* 92 (4): 753–791.
- Staubs, Robert. 2014. Computational modeling of learning biases in stress typology. PhD diss, University of Massachusetts Amherst, Amherst.
- Staubs, Robert, Michael Becker, Christopher Potts, Patrick Pratt, John J. McCarthy, and Joe Pater. 2010. OT-Help 2.0. Software package.
- Terrill, Angela. 1999. Lavukaleve: A Papuan language of the Solomon Islands. PhD diss, Australian National University, Canberra. 9783110178876.
- Tesar, Bruce, and Paul Smolensky. 2000. *Learnability in optimality theory*. MIT Press.
- Thoudam, Purna Chandra. 1980. A grammatical sketch of Meiteiron.
- Tola, Waqo. 1981. The Phonology of Mecha Oromo. *Unpublished MA Thesis. Addis Ababa: Addis Ababa University*.
- Topintzi, Nina, and Andrew Nevins. 2017. Moraic onsets in Arrente. *Phonology* 34 (3): 615–650.
- Voorhoeve, Clemens L. 1965. *The Flamingo Bay Dialect of the Asmat Language*. Vol. 46 of *Verhandelingen van het Koninklijk Instituut voor Taal, Land en Volkenkunde*. The Hague: M. Nijhoff.
- Wallis, Ethel E. 1968. The word and the phonological hierarchy of Mezquital Otomi. *Language*.
- Watkins, Laurel J. 1980. A grammar of Kiowa. PhD diss, University of Kansas.
- White, James. 2014. Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130 (1): 96–115.
- Wilson, Colin. 2006. Learning phonology with a substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* 30 (5): 945–982.
- Wolfart, H. Christoph. 1996. Sketch of Cree, an Algonquian Language. In *Handbook of American Indians. Volume 17: Languages*, ed. Ives Goddard, Vol. 17, 390–439. Washington: Smithsonian Institute.