




Emergent learning bias and the underattestation of simple patterns

Charlie O'Hara¹ 

Received: 30 January 2018 / Accepted: 4 December 2022 / Published online: 18 January 2023

© The Author(s), under exclusive licence to Springer Nature B.V. 2023

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.

Keywords Phonological learning · Maximum Entropy Harmonic Grammar · Soft typology · Simplicity bias · Onset-coda asymmetries · Place of articulation · Iterated learning

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 theory of grammar projects a grammatical typology, the set of phonological systems that can be modeled by said theory. Historically, this grammatical typology has

✉ C. O'Hara
charlieoharaphd@gmail.com

¹ University of Michigan, Lorch Hall, 611 Tappan Ave, Ann Arbor, MI 48109, USA

been the major way scholars attempted to define the range of possible phonological systems.

One of the most clear pieces of empirical data that offers insight towards the set of possible phonological patterns is 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 (in the sense of Moreton 2008, 2010), as evidence for our phonological theories.

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 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). Bane and Riggle (2009) further extend the idea of *r-volume* to systems with weighted constraints, such as Harmonic Grammar and MaxEnt.

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 in order to be modeled in a weighted constraint grammar are harder to learn than those that do not, Bane and Riggle (2009) and Carroll (2010) note that such patterns tend to have relatively small *r-volumes* in Harmonic Grammar. From these studies, it is not clear whether the learning bias is just a result of the skews in *r-volume* or whether other

factors are at play as well. One extreme hypothesis therefore is that the soft typology of the world's phonological patterns is correlated to the r-volume of those patterns modeled with a particular set of constraints and that the typological niches 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 phonological patterns in MaxEnt using a minimally restrictive constraint set. Pater and Moreton (2012) and Moreton et al. (2017) show that MaxEnt learners with simple, symmetrical, and unrestricted sets of constraints (GMECCS or Gradual Maximum Entropy with a Conjunctive Constraint Schema) are biased towards featurally simpler patterns, resembling typological generalizations seen in Clements' (2003) Feature Economy. 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, leaving 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 these two components like the r-volume or GMECCS models do. I show that the model based on the interaction of learning and grammar makes closer matches to typological skews than a model based on the r-volume of patterns, or a model based on the learnability of those patterns with a minimal set of constraints. The learner and the grammar interact in ways that are not apparent from focusing on either one in isolation.

The interactive model of learning and grammar implicates types of learning biases that cannot be predicted by the GMECCS model with a simple symmetrical constraint set. The GMECCS model predicts that learning biases 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]∨[+G]

For example, Pater and Moreton (2012) consider the typology of obstruent stop inventories. As seen below in (2)–(4), languages that have three voiceless stops [p], [t], and [k] may allow all, none, or a subset of the voiced stops [b d g]. 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 (Bruner et al. 1956; Hunt and Hovland 1960; Shepard et al. 1961), and has 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 better 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 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; Finley 2012; White 2014; Kimper 2016a; Lin 2016; Martin and Peperkamp 2020). These substantive biases have also been observed in experiments probing speakers learning of natural languages in non-artificial settings (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 might be explained solely as the result of some channel bias, and not as a direct learning bias. However, this paper argues that not all typological skews are either just structural or just 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 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. Encoding substantive information, such as markedness hierarchies, in a set

of constraints, may be meant to handle a small set of substantive phonetically natural skews. However, when this substantive information interacts with the learning algorithm and other markedness hierarchies, this substantive information may create downstream effects, affecting the learnability of phonological patterns in ways that are not immediately apparent from their encoding, leading to substantive biases 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 separate components by a learning model, a simplicity bias would predict that patterns that make restrictions based on only syllable-position or only place of articulation should be more easily learned than patterns that require an interaction of both of these dimensions, and therefore the simple patterns should be better attested. Section 2 presents a typological survey of place and position in stop inventories. The survey's results are 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 Sect. 4, two alternative approaches to capturing these typological generalizations are explored: the GMECCS model, a learning-based model without substantive biases in the constraints themselves, and the r-volume model, which attempts to capture typological tendencies through grammatical bias without learning (Anttila 1997; Coetzee 2002; Bane and Riggle 2008). 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 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 phonological 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 Sect. 2.3 will present the results and demonstrate evidence for the all-or-nothing and positional priority skews.

2.1 Positional stop inventories

The hypothesis that typological attestation is based only on the simplicity bias 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's (2006) 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.¹

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

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).² 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

¹ < here denotes the relation "is less harmonic (or more marked) than."

² 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.

hierarchies. The typological survey in this paper is too small to directly test this hierarchy, however see O'Hara (2021, Sect. 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 codaful syllables (Jakobson and Halle 1956; Blevins 1995).³ 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: onsetful syllables 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 Sect. 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 Sect. 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 (e.g. WALSH, 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 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 $V\underline{C}CV$ context, but distinguishing $VC.CV$ syllabification from $V.CCV$ can be difficult. If a [pr] cluster appears intervocally it is ambiguous (especially from a transcribed rather than phonetic form) whether the cluster is a complex

³Cf. Breen and Pensalfini (1999) who argue that Arrernte has obligatory codas and no onsets, though see Topintzi and Nevins (2017) for a recent reanalysis of Arrernte.

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 first 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 usually is able to surface before any possible word-initial 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.⁴ 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, including three Bantu languages. 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 documentation of some variety⁵ of that language that listed which consonants were available in word-initial

⁴Available at <https://wals.info/languoid/samples/100> as of August 14, 2019.

⁵Different 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.

and word-final position.⁶ 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).⁷ I was unable to find sufficient documentation on Rama (or any closely related Chibchan languages), so it was excluded from the sample, resulting in a sample of 93 languages.

In this paper, I will discuss only those languages that have no supralaryngeal places of articulation beyond [t], [p], and [k]. There are 44 languages in the sample that allow no more places of articulation than these three, and all but two of these 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 pattern with four places of articulation is well enough attested to control for such pressures.⁸ 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.⁹

⁶In 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 Alamlak (Edmiston and Edmiston 2003), Arapesh (Mountain) (SIL 2011a), and Daga (SIL 2011b).

⁷The 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).

⁸Overall, this decision was made to control for potential interactions between the learning of secondary places of articulation and the primary places of articulation. Considering additional places of articulation raises questions about how to best represent the relative markedness between secondary places of articulation and primary ones, and might interact with learning in myriad ways i.e. would the presence of [t] in the training data enhance the learning of [t] more than [c] or [q] would? However, languages with more than three contrasting places of articulation for stops still appear to be largely subject to the all-or-nothing skew. Such languages exhibit subset inventories, i.e. some but not all of the word-initial stops are licit in word-final position, at non-significantly higher rates than languages with only [p t k] word-finally. See (O'Hara 2021, Sect. 2.4.3) for more details.

⁹The presence of an affricate [tʃ] in a language 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; Arvaniti 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.

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 (e.g. being banned in just onset or just coda, or appearing in utterance-initial or utterance-final positions) and b) the markedness relationship between glottal stops and [p], [t], and [k] is complex, without a clear typological implication (de Lacy 2006). These properties make glottal stops qualitatively different than any fourth supralaryngeal place of articulation such as [q] or [c], which tend to show distributional qualities similar to [p], [t], and [k]. As such, a language that allows [p], [t], [k], and [ʔ] in a position is treated the same in this survey as a language that allows [p], [t], and [k] but not glottal stop in that position, 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.¹⁰ 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); Finnish is classified as not allowing word-final [k] in this survey. 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 appears in Appendix A.

2.3 Results

There are two crucial results from the study. First, in Sect. 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, Sect. 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

¹⁰de Lacy (2006, 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.

Table 1 Example languages of each pattern

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

[pap]	pop	[tap]	top	[kap]	cop
[pap]	pop	[pat]	pot	[pak]	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 ^h luk ^h]	‘cabbage’
6. Language with labial and dorsal stops word-finally: Lakota ([pk]-Final) (Rood and Taylor 1996)

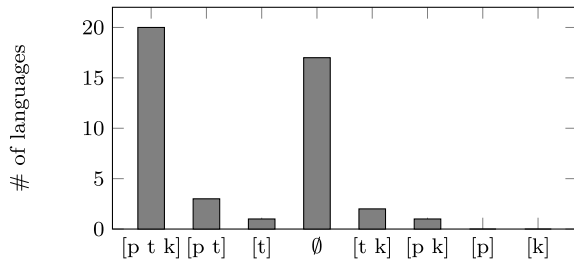
[p ^h i]	‘liver’	[t ^h i]	‘to live’	[k ^h i]	‘to reach home there’
[t ^h ob]	‘four’	*[t ^h ot]		[pa ^h ag]	‘stopping’

(1993/2004), de Lacy (2006). An inventory of word-final stops is harmonically complete if any particular segment is allowed word-finally, then every less marked stop must also be allowed word-finally. 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, (based on the markedness hierarchy in (6)). 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 Fig. 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

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

Word-Final	Number	Languages
[p t k]	20	Abun, Alamblak, Asmat, Canela-Krahô, Chamorro, Cree (Plains), Chontal Mayan, Daga, English, Georgian, Karok, Koasati, Korean, Lango, Ma'anyan, 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	44	

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

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 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 44 languages considered, 37 languages, or 84.1%, 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 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 ($\frac{37}{44}$, $p < .001$ 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

($p < .001$).¹¹ 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 types 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 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.¹² 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], so at its simplest, it 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, as shown in Fig. 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.

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

¹¹All p-values testing pairwise independence of stops are from Fisher's Exact Test for Count Data.

¹²This metric is a simplification of the algebraic complexity metric proposed by Feldman (2006). This simplification performs 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 O'Hara (2021).

<p>(a) All-Final Simple <i>Complexity</i> = 0 $\frac{20}{44} = 45.4\%$</p>	<p>(b) [pt]-Final Complex <i>Complexity</i> = 2 $\frac{3}{44} = 6.8\%$</p>	<p>(c) [t]-Final Complex <i>Complexity</i> = 2 $\frac{1}{44} = 2.3\%$</p>	<p>(d) No-Final Simple <i>Complexity</i> = 1 $\frac{17}{44} = 38.6\%$</p>
<p>pV tV kV Vp Vt Vk</p>	<p>pV tV kV Vp Vt Vk</p>	<p>pV tV kV Vp Vt Vk</p>	<p>pV tV kV Vp Vt Vk</p>

Fig. 2 Complexity example

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 |

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 allows 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. 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 equally 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). These languages both 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 ($\frac{17}{19}$, $p < 0.001$).¹³ In sum, there are no languages in the WALs sample that clearly allow only two contrasting supralaryngeal places of articulation in their phonemic inventory.

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 (O'Hara 2021, Sect. 2.6.3).

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. This skew is a substantive skew, because both patterns are equally simple so defining the requires referring to the *substance* of these dimensions, either word-position vs. place of articulation. 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 presented in Sect. 3.1. The results of the generational simulations are presented in Sect. 3.2, and shown to capture the typological skews observed in Sect. 2.

3.1 Modeling learning bias

The phonological learning problem is difficult. The phonological grammar is a mapping between underlying forms to surface forms, defining which forms are licit and

¹³In fact, the difference between these categories remains significant with up to seven languages of the Only-[pt] type.

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 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 Sect. 2. This paper uses a Generational Stability Model (GSM)¹⁴ 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 in theory, some constraints could be defined to assign non-positive real numbers (Smolensky et al. 2014; McCollum 2018) or positive integers though care must be taken to avoid the "infinite goodness" problem (Kemper 2011, 2016b; Kaplan 2018).

- (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

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

¹⁴Also known as an Iterative Learning Model (Kirby and Hurford 2002; Hughto 2018).

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 step in the learning procedure, an input form x is 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 are 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

$$\eta \text{ is a learning rate constant between 0 and 1.}$$

$$w_C(t + 1) = w_C(t) + \eta * (C(x, y_L) - C(x, y_T))$$

There are two sources of stochasticity in this algorithm, first the input form chosen at each iteration of learning is randomly selected, and second 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 two sources of randomness prevent the algorithm from being deterministic—any two learners could have seen a slightly different sequence 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 will 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 grammars that differ slightly from 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 Hurford (2002), Staubs (2014), Hughto (2018), Ito and Feldman (2022). 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* into a teacher, and a new learner is initialized and 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. Because MaxEnt is a probabilistic model of grammar, most learners ended up learning slightly different grammars, so I classified these patterns into particular bins for comparison. 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 Hugtto (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 for the next generation to learn a more categorical pattern. In other words, generational chains are very unlikely to reverse changes across these bins, but they are likely to oscillate back and forth within these bins—though they are attracted towards more categorical patterns. 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 (O'Hara 2021, Sect. 3.3.5).

In order to encode the place of articulation hierarchy, I will make use of stringently related constraints, following Kiparsky (1994), Prince (1999), de Lacy (2002). The three markedness constraints I included are listed in (11), based on Lacy's 2002 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 Sect. 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 redundant for purposes of factorial typology in this small domain: ONSET causes segments to be protected in onset while they would delete in codas, and NOCODA drives segments to delete in codas while they would not in onsets. One could omit either of these constraints and would still be able to capture the same range of onset/coda stop inventories. This redundancy is important however, 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]).¹⁵ 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.

¹⁵While 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 changes in manner.

Using just deletion, MAX and the markedness constraints discussed above, and considering only faithful and deletion candidates, all of the surface patterns consistent with the markedness hierarchies can be modeled in Harmonic Grammar (and even in 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 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. Whereas 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 the Place of Articulation Scale
- a. IDENT{X} Assign a violation for a segment in 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 Sect. 2, allowing for both deletion and debuccalization to be used as repairs for the stops.

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

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	X	X	X	X	X	X	X	
[t]-Initial	X	✓	X	X	X	X	X	
[pt]-Initial	✓	✓	X	X	X	X	✓	Tahitian (Tryon 1970)
No-Final	✓	✓	✓	X	X	X	✓	Spanish
[t]-Final	✓	✓	✓	X	✓	X	✓	Finnish
[pt]-Final	✓	✓	✓	✓	✓	X	✓	Kiowa
All-Final	✓	✓	✓	✓	✓	✓	✓	English
Only-[t]	X	✓	X	X	✓	X	X	
[pt]-Initial+[t]-Final	✓	✓	X	X	✓	X	X	
Only-[pt]	✓	✓	X	✓	✓	X	X	

This set of constraints predicts 108 distinct categorical grammars in Harmonic Grammar according to OT-Help (Staubs 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 language with no place contrasts for stops at all is extremely impractical for communication), but some (like the Only-[pt] pattern discussed in Sect. 2) are more surprising. Ten of these patterns are harmonically complete, meaning they fully respect the place of articulation markedness hierarchy in (6), 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. To see that the all-or-nothing skew discussed in Sect. 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. 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.¹⁶

¹⁶Though see the predictions of r-volume approaches in Sect. 4.2.

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 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 (Tesar and Smolensky 2000; Gnanadesikan 2004; Jesney and Tessier 2011).

The setting of these parameters affects 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. The parameter settings used here were chosen relatively arbitrarily with four goals in mind: one, learners should (usually) be closer to the categorical target pattern than any other pattern after the first generation to ensure they have enough data to learn the patterns in question; two, the learning rate is very small relative to initial markedness constraint weights to ensure weak convergence; three, most patterns were not categorically stable or unstable, so that the relative stability of patterns is comparable; and four, simulation time was relatively short (<8 hours).

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, 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. 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 (Inkelas 1995; Tesar and Smolensky 2000) and 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 candidate if it was a final stop, and to its debuccalization candidate for initial stops. As a result, when the initial teacher produces an unfaithful mapping it matches the learner's initial unfaithful

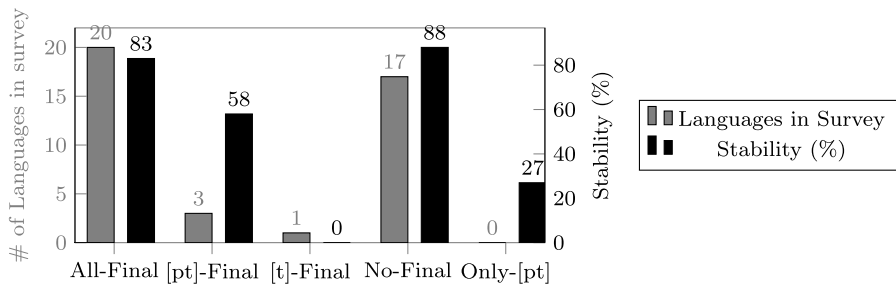


Fig. 3 Well attested patterns are stable in the simulations, but poorly attested ones are not

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 /V_k/→[V?]) would be unable to reconstruct the underlying form /V_k/ without sufficient evidence from alternations.

The Python code utilized to run these simulations is available at <https://github.com/charlieohara/softtypologytool>.

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.

Please note, the 0% stability of the [t]-Final pattern is only reflective of the fact that that pattern is very *unlikely* to be stable, rather than it being impossible for the [t]-Final pattern to be stable in these sorts of simulations. Under these parameter settings, the [t]-Final pattern is learned stably across the first five generations in the majority of runs, and remains stable in some runs for up to 14 generations. If we increased the number of iterations per generation, the stability of all patterns, including the [t]-Final pattern would increase.

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

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 supralaryngeal stops being banned. Incomplete learning involves learning only a subset of the target pattern's 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.¹⁷ 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 (by elimination, because there are no coda stops). Runs trained on Only-[pt] lose final [Vp].

However, for the [t]-Final and [pt]-Final patterns, there is no 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

¹⁷Noticeably, 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.

the [t]-Final pattern instead. This result can help explain why the less stable [t]-Final pattern is more common than the seemingly more stable Only-[pt] pattern. 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 Sect. 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 is never accidentally learned by learners attempting to learn other patterns. 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 learned 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 in the digital version of this article) 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.

To measure how long it took to learn each form when learners were trained on the All-Final pattern, I found the number of learning iterations it took for each form to be produced with a probability of at least .9 in each of the 100 runs. There was a significant difference via a 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; $t(99) = -13.8$, $p < .001$. This shows that /kV/ is consistently learned faster than /Vt/, when trained on the All-Final pattern, even though they are both learned in rapid succession. Thus, across each of these trained patterns, the learning order is quite consistent across runs. Because patterns tend to lose forms across generations in reverse order of learning, the fact that learners tend to learn forms in the same order means that 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

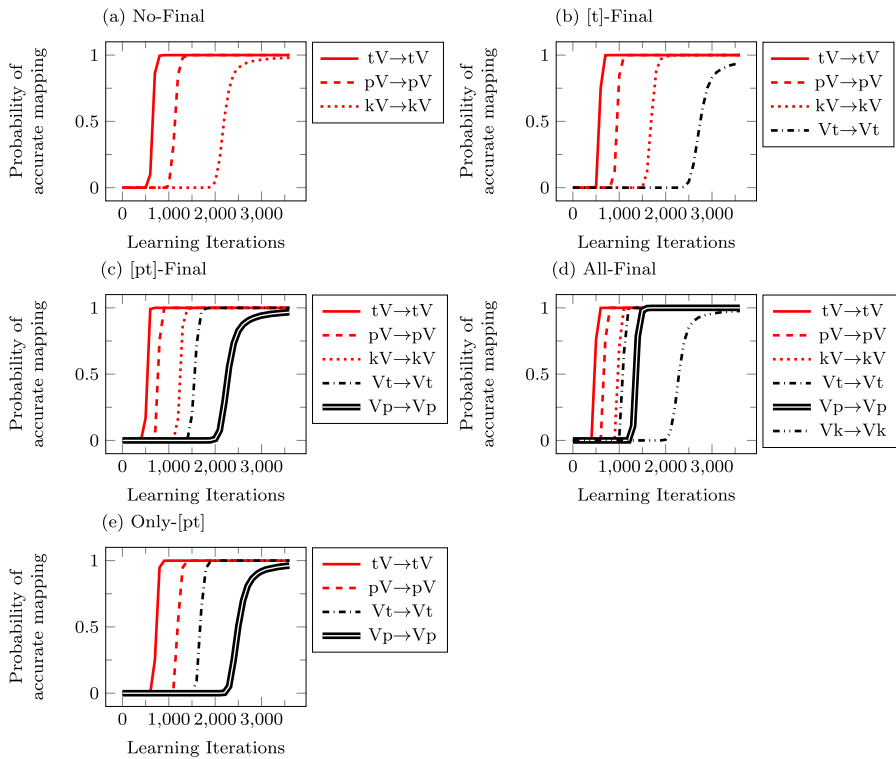


Fig. 4 Learning paths for five different patterns (averaged over 100 runs). The legend of each plot shows the mappings in the order that they are learned. (Unfaithful mappings that remain at one throughout the learning process are omitted for ease of reading) (Color figure online)

the learning order is so consistent in these simulations, 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 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 ($p < .001$). Across all five patterns, as the average number of learning iterations required to learn a pattern

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

increases, the stability decreases, showing the clear correlation between stability and first-generation behavior.

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]), replicating the positional priority 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 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 generational agent-based learning models that focus 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, Sect. 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. 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 structural 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.

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 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).¹⁸ These constraints are defined as to create no substantive bias in the grammar. Therefore, for any markedness constraint NOCODA or *T, there is an inverse constraint CODA or +T, that rewards the behavior the other constraint penalized. 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

¹⁸The 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 the model I use in this section 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.

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	

other constraints at 50. Each generation was exposed to 2500 forms, and each run went for 20 generations.¹⁹

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

The results of these simulations can be found in Table 6. 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.

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 significantly less stable than the language that only bans final [Vk]. The typological data is better matched by the model with substantively biased constraints than this model with unbiased constraints.

In the previous section, in the substantively biased model, I noted that even though the Only-[pt] pattern was stable, it was never innovated by learners attempting to learn other patterns, even the similar [pt]-Final pattern. However, the model with unbiased constraints shows a different pattern. Unstable runs in this model frequently involve learning more simple patterns, without any priority for position over place—18 runs became the Only-[pt] pattern, losing initial [kV], but only 2 runs changed into the [t]-Final pattern, losing [Vp]. If anything, the typologically frequent No-Finals pattern is less likely to be innovated, rather than the typologically unattested Only-[pt] pattern.

While not shown directly here, this model also has difficulty capturing implicational relations between places of articulation or syllable-positions. Since there is no

¹⁹Because 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.

substantive bias in the grammar, all places of articulation are 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 Sect. 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 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 mirror 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 structurally-based model used in this paper. Hughto et al. (2014), Hughto and Pater (2017) show that an iterated learning MaxEnt model disprefers patterns that require gang effects in order to be modeled, though Bane and Riggle (2009) and Carroll (2010) note that patterns that require gang effects tend to have relatively 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.²⁰ The constraints used were those used

²⁰Using 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

Table 7 Predicted attestation rates based on R-volume

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

in the simulations in Sect. 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 7.²¹ 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] 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 there are 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 to [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 the 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

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 affect the relative r-volume of any of these patterns.

²¹A full table including the r-volume of each pattern that was sampled in the r-volume simulations can be found in 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.

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 added 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 Sect. 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 must 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 Sect. 3 has a similar r-volume to the well-attested and easy to learn All-Final pattern. Without allowing for 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 in Table 8. The positional priority skew observed in the typological data is predicted by the interactive model in Sect. 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 0.74, but the other two models fit the typology far worse, with the structural bias model having an R^2 of 0.31, and the r-volume model having an R^2 of 0.13.

The evidence suggests that more languages of the world favor the position hierarchy above the place hierarchy—i.e. languages that ban dorsals initially, lack final

Table 8 Comparison of the three models to the typology

Model	Positional Priority skew	All-or-nothing skew	R ²
Interactive Model (Sect. 3)	✓	✓	0.74
Learning: Structural Bias (Sect. 4.1)	✗	✓	0.31
Grammar: r-volume (Sect. 4.2)	✓	✗	0.13
Typology (Sect. 2)	✓	✓	

stops. This is evidence favoring the Interactive Model to the structural 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 Interactive 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, on the grammar and on their interaction, rather than a model that focuses on deriving inductive biases from either grammar or learning alone.

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 within a theory of universal CON, they should also consider how different sets of constraints may predict that different patterns would be 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 occur (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 phonetically-natural substantive typological skews 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. But the positional priority skew found in the learning models in this paper is a downstream result of the way markedness hierarchies were encoded in the grammar. If the markedness hierarchies emerge out of channel bias, will the same non-phonetically natural substantive biases, such as the positional priority skew emerge as a second order result from that channel bias, will the positional priority skew emerge from interaction with other external factors, such as morphology or lexicon learning, or does the positional priority skew only emerge when markedness hierarchies are modeled via substantively grounded constraints?

Appendix A: Information about Language Survey

Table 9 lists the 44 languages used in the empirical studies in this paper. I list the language, family and genus classifications from WALS, the source I used to find information about positional consonant inventories, and the inventory of available supralaryngeal stops in initial and final position.

Table 10 lists the languages and sources for the languages with more than three places of articulation available. In some of these cases, I was unable to find clear

Table 9 Languages in typological survey with 3 places of articulation

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
Apurinã	Arawakan	Purus	Facundes (2000)	p t k	∅
Arapesh (Mountain)	Torricelli	Kombio-Arapesh	SIL (2011a)	p t k	∅
Asmat	Trans-New Guinea	Asmat-Kamoro	Voorhoeve (1965)	p t k	p t k
Barasano	Tucanoan	Tucanoan	Jones and Jones (1991)	p t k	∅
Canela-Krahô	Macro-Ge	Ge-Kaingang	Popjes and Popjes (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

Table 9 (continued)

Language	Family	Genus	Source	Initial	Final
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	∅
Hixkaryana	Cariban	Cariban	LAPSyD	p t k	∅
Imonda	Border	Border	Seiler (1985)	p t k	t k
Indonesian	Austronesian	Malayo-Sumbawan	Lapoliwa (1981)	p t k	p t
Japanese	Japanese	Japanese	LAPSyD	p t k	∅
Karok	Karok	Karok	Bright (1957)	p t k	p t k
Kewa	Trans-New Guinea	Engan	Franklin (1971)	p t k	∅
Kiowa	Kiowa-Tanoan	Kiowa-Tanoan	Watkins (1980)	p t k	p t
Koasati	Muskogean	Muskogean	Kimball (1985)	p t k	p t k
Korean	Korean	Korean	Chang (1996)	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	∅
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	∅
Persian	Indo-European	Iranian	Alamolhoda (2000)	p t k	p t k
Pirahã	Mura	Mura	Everett (1986)	p t k	∅
Quechua (Imbabura)	Quechuan	Quechuan	Cole (1985)	p t k	∅
Sanuma	Yanomam	Yanomam	Borgman (1990)	p t k	∅
Sierra Popoluca	Mixe-Zoque	Mixe-Zoque	de Jong Boudreault (2009)	p t k	p t k
Spanish	Indo-European	Romance	LAPSyD	p t k	∅
Supyire	Niger-Congo	Gur	Carlson (1994)	p t k	∅

Table 9 (continued)

Language	Family	Genus	Source	Initial	Final
Tagalog	Austronesian	Greater Central Philippine	Llamzon (1976)	p t k	p t k
Tukang Besi	Austronesian	Celebic	Donohue (2011)	p t k	∅
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	∅
Yaqui	Uto-Aztecan	Cahita	Dedrick and Casad (1999)	p t k	p t k

Table 10 Languages in typological survey with more than 3 places of articulation

Language	Family	Genus	Source	Additional stops
Abkhaz	Northwest Caucasian	Northwest Caucasian	Hewitt (1979)	t ^w k ^j k ^w q
Acoma	Keresan	Keresan	Miller (1965)	t ^j
Amele	Madang	Trans-New Guinea	Roberts (2016)	gb
Arabic (Gulf)	Semitic	Afro-Asiatic	Qafisheh (1977)	t ^ʕ q
Bagirmi	Bongo-Bagirmi	Central Sudanic	Thayer (1974)	t c
Basque	Basque	Basque	Hualde et al. (2010)	c
Berber (Tashlhiyt)	Berber	Afro-Asiatic	Ridouane (2014)	t ^ʕ k ^w q q ^w
Burmese	Burmese-Lolo	Sino-Tibetan	Okell (1969)	c
Burushaski	Burushaski	Burushaski	Yoshioka (2012)	t q
Chukchi	Northern Chukotko-Kamchatkan	Chukotko-Kamchatkan	Dunn (1999)	q
Dani (Lower Grand Valley)	Dani	Trans-New Guinea	Bromley (1961)	k ^w
Gooniyandi	Bunuban	Bunuban	McGregor (1990)	d j
Grebo	Kru	Niger-Congo	LAPSYD	c kp
Greenlandic (West)	Eskimo	Eskimo-Aleut	Sadock (2003)	q
Guaraní	Tupi-Guaraní	Tupian	LAPSYD	k ^w
Hausa	West Chadic	Afro-Asiatic	Newman (2000)	c k ^w k ^j
Hindi	Indic	Indo-European	Ohala (1983)	t
Hmong Njua	Hmong-Mien	Hmong-Mien	Mortensen (2004)	c q
Humburi Senni	Songhay	Songhay	Heath (2014)	c
Kannada	Southern Dravidian	Dravidian	Sridhar (1990)	t




















Table 10 (continued)

Language	Family	Genus	Source	Additional stops
Kayardild	Tangkic	Tangkic	Round (2009)	t̥ c
Khalkha	Mongolic	Altaic	Svantesson et al. (2005)	p ^j t̥ ^j g ^j
Khoekhoe	Khoe-Kwadi	Khoe-Kwadi	Hagman (1977)	k k k! k‡
Krongo	Kadugli	Kadu	LAPSYD	t̥ c
Kutenai	Kutenai	Kutenai	Garvin (1948)	q
Lezgian	Lezgitic	Nakh-Daghestanian	LAPSYD	t ^w k ^w q q ^w
Makah	Southern Wakashan	Wakashan	LAPSYD	k ^w q q ^w
Mangarrayi	Mangarrayi	Mangarrayi-Maran	Merlan (1982)	d̥ j
Mapudungun	Araucanian	Araucanian	Zúñiga (2000)	t̥
Maricopa	Yuman	Hokan	LAPSYD	t̥ k ^w q q ^w
Martuthunira	Western Pama-Nyungan	Pama-Nyungan	Dench (1987)	c
Maung	Iwaidjan	Iwaidjan	Capell and Hinch (1970)	d̥ j
Meithei	Kuki-Chin	Sino-Tibetan	Thoudam (1980)	c
Mixtec (Chalcatongo)	Mixtecan	Oto-Manguean	Macaulay (1996)	k ^w
Ngiyambaa	Southeastern Pama-Nyungan	Pama-Nyungan	LAPSYD	t̥ c
Paiwan	Paiwan	Austronesian	Chen (2009)	c q
Russian	Slavic	Indo-European	Yanushevskaya and Bunčić (2015)	p ^j t̥ ^j k ^j
Sango	Ubangi	Niger-Congo	Samarin (1967)	kp
Slave	Athapaskan	Na-Dene	Rice (1976)	k ^w
Swahili	Bantoid	Niger-Congo	Polomé (1967)	c
Thai	Kam-Tai	Tai-Kadai	Noss (1964)	c
Tiwi	Tiwan	Tiwan	Anderson and Maddieson (1994)	t̥ t̥
Vietnamese	Viet-Muong	Austro-Asiatic	Thompson (1965)	c
Warao	Warao	Warao	Romero-Figeroa (1997)	k ^w
Wari'	Chapacura-Wanham	Chapacura-Wanham	Everett and Kern (2002)	k ^w
Wichí	Matacoan	Matacoan	Avram (2008)	k ^j k ^w
Yoruba	Defoid	Niger-Congo	Adesola (No date)	kp

information about word-initial vs. word-final or onset vs. coda distributions, but because these languages had more than three places of articulation, I stopped my search. Information about the consonant inventories of some of these languages was available in the LAPSYD database (Maddieson et al. 2014–2019), which usually does not include information about positional inventories, but if LAPSYD showed that the language had more than three places of articulation I terminated my search there.

Appendix B: Estimated r-volumes of each pattern

Table 11 Predicted attestation rates based on R-volume

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

Acknowledgements Particular thanks go to Karen Jesney and Rachel Walker for their guidance on this paper and its related projects since its inception. I would also like to thank the audiences at AMP 2017, LSA 2018, SCAMP 2017, the Analyzing Typological Structure workshop, the Workshop on the Emergence of Universals, AMP 2018, SCiL 2019, USC, UMass Amherst, UC Berkeley, Miami University, Florida State University for helpful feedback. Further thanks go to discussion and comments from Arto Anttila, Morteza Deghani, Eleanor Glewwe, Jeremy Goodman, Coral Hughto, Larry Hyman, Khalil Iskarous, Hayeun Jang, Gaja Jarosz, Myriam Lapierre, Joe Pater, Katya Pertsova, Betsy Pillion, Brandon Prickett, Jason Riggle, Stephanie Shih, Caitlin Smith, Juliet Stanton, and Jason Zevin.

Declarations

Competing Interests The author declares no competing interests.

References

- Adesola, Oluseye. No date. Yoruba: A grammar sketch version 1.0. Available at <https://www.africananaphora.rutgers.edu/images/stories/downloads/casefiles/YorubaGS.pdf>.
- Alamolhoda, Seyyed Morteza. 2000. *Phonostatistics and phonotactics of the syllable in modern Persian*, Vol. 89. Helsinki: Finnish Oriental Society.

- Anderson, Victoria B., and Ian Maddieson. 1994. Acoustic characteristics of Tiwi coronal stops. *UCLA Working Papers in Phonetics* 87: 131–162.
- 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: John Benjamins.
- Anttila, Arto. 2008. Gradient phonotactics and the complexity hypothesis. *Natural Language and Linguistic Theory* 26(4): 695–729.
- Arvaniti, Amalia. 2007. Greek phonetics: The state of the art. *Journal of Greek Linguistics* 8(1): 97–208.
- Avram, Megan L. Z. 2008. A phonological description of Wichí: The dialect of Misi3n La Paz. Master thesis. Salta, Argentina Orienta33o: Ver3nica Grondona Master of Arts, Eastern Michigan University.
- 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*, 29–38. Columbus: Association for Computational Linguistics.
- Bane, Max, and Jason Riggle. 2009. The typological consequences of weighted constraints. In *Proceedings of Chicago linguistics society 45*, Vol. 45, 13–27.
- Beckman, Jill N. 1998. Positional faithfulness. PhD diss, University of Massachusetts, Amherst.
- Bell, A. 1971. Some patterns of the occurrence and formation of syllabic structure. In *Working papers on language universals*, Vol. 6, 23–138. Language Universals Project, Stanford University.
- Berry, Keith, and Christine Berry. 1999. *A description of Abun: A West Papuan language of Irian Jaya*. Canberra: 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. Cambridge, MA & Oxford: Blackwell.
- Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: 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. Sheffield: Equinox.
- Borgman, Donald M. 1990. *Sanuma*, Vol. 2 of *Handbook of Amazonian Languages*, 15–248. Berlin, New York: Mouton de Gruyter.
- 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*. University of California.
- Bromley, H. Myron. 1961. *The phonology of Lower Grand Valley Dani: A comparative structural study of skewed phonemic patterns*. Vol. 34 of *Verhandelingen van het koninklijk instituut voor taal, land en volkenkunde*. s-Gravenhage: Martinus Nijhoff.
- Bruner, Jerome S., Jacqueline J. Goodnow, and George A. Austin. 1956. *A study of thinking*. New York: John Wiley and Sons.
- Butskhrikidze, Marika. 2002. The consonant phonotactics of Georgian. PhD diss, Netherlands Graduate School of Linguistics.
- Capell, Arthur, and Heather E. Hinch. 1970. *Maung grammar. Janua linguarum, series practica*. The Hague: Mouton de Gruyter.
- Carlson, Robert. 1994. *A grammar of Supyire*, Vol. 14. Berlin, New York: De Gruyter Mouton.
- Carroll, Lucien. 2010. Cumulativity in constraint grammars: An iterated learning view of probabilistic typology. Ms., UC San Diego.
- Chang, Suk-Jin. 1996. *Korean*, Vol. 4. Amsterdam: John Benjamins Publishing.
- Chen, Chun-Mei. 2009. Documenting Paiwan phonology: Issues in segments and non-stress prosodic features. *Concentric: Studies in Linguistics* 35(2): 193–223.
- 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*. Amsterdam: North-Holland Publishing Company.
- de Jong Boudreault, Lynda J. 2009. A grammar of Sierra Popoluca (Soteapanec, a Mixe-Zoquean language). PhD diss, University of Texas at Austin.
- de Lacy, Paul. 2002. The formal expression of markedness. PhD diss, University of Massachusetts, Amherst.
- 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*. Tucson: University of Arizona Press.
- Dench, Alan Charles. 1987. *Martuthunira: A language of the Pilbara region of Western Australia*. PhD diss, The Australian National University.
- Dixon, Robert M. W. 1988. *A grammar of Boumaa Fijian*. Chicago, London: The University of Chicago Press.
- Donohue, Mark. 2011. *A grammar of Tukang Besi*, Vol. 20. Berlin, New York: Walter de Gruyter.
- Dryer, Matthew S., and Martin Haspelmath, eds. 2013. *WALS online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/>.
- Dunn, Michael. 1999. *A grammar of Chukchi*. PhD diss, Australian National University.
- Edmiston, Pat, and Mel Edmiston. 2003. Alambalak 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. Berlin, New York: Mouton de Gruyter.
- Everett, Daniel L., and Barbara Kern. 2002. *Wari*. London, New York: Routledge.
- Facundes, Sidney da Silva. 2000. *The language of the Apurinã people of Brazil (Maipure/Arawak)*. PhD diss, State University of New York, Buffalo.
- 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.
- Flack, Kathryn. 2007. *The sources of phonological markedness*. PhD diss, University of Massachusetts 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*. Colchester, VT: 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, Vol. 1 method and theory*, eds. Joseph H. Greenberg, C. A. Ferguson, and E. A. Moravcsik, 61–91. Stanford: Stanford University Press.
- Gudai, Darmansyah H. 1985. *A grammar of Maanyan: A language of Central Kalimantan*. PhD diss, Australian National University.
- Hagman, Roy Stephen. 1977. *Nama hottentot grammar*. Bloomington: Indiana University Press.
- Harms, Robert T. 1964. *Finnish structural sketch*. Vol. 42 of *Uralic and Altaic series*. Indiana University.
- 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.
- Heath, Jeffrey. 2014. *Grammar of Humburi Senni (Songhay of Hombori, Mali)*. Ms., Max-Planck Institute for Evolutionary Anthropology, Leipzig.
- Hewitt, B. George. 1979. *Abkhaz*. Vol. 2 of *Lingua descriptive studies*. Amsterdam: North-Holland.
- Hockett, Charles F. 1955. *A manual of phonology*. Baltimore: 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. Berlin, Boston: Mouton de Gruyter.
- Householder, Frederic W. 1964. Three dreams of Modern Greek phonology. *Word* 20(3): 17–27.
- Hualde, José Ignacio, Oihana Lujanbio, and Juan Joxe Zubiri. 2010. Goizueta Basque. *Journal of the International Phonetic Association* 40(1): 113–127. <https://doi.org/10.1017/S002510030990260>.
- Hughto, Coral. 2018. Investigating the consequences of iterated learning in phonological typology. 1: 182–185.
- 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, Earl B., and Carl 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 *Proceedings of NELS* 25, 287–302.
- 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.
- Ito, Chiyuki, and Naomi Feldman. 2022. Iterated learning models of language change: A case study of Sino-Korean accent. *Cognitive Science* 46: 13115.
- Itô, Junko. 1986. Syllable theory in prosodic phonology. PhD diss, University of Massachusetts Amherst.
- Jäger, Gerhard. 2007. Maximum entropy models and stochastic optimality theory. In *Architectures, rules, and preferences: A festschrift for Joan Bresnan*, 467–479. Stanford: CSLI Publications.
- 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 and Linguistic Theory* 29: 251–290.
- 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.
- Kaplan, Aaron. 2018. Positional licensing, asymmetric trade-offs and gradient constraints in Harmonic Grammar. *Phonology* 35(2): 247–286.
- 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: MIT Press.
- Kimball, Geoffrey David. 1985. A descriptive grammar of Koasati. PhD diss, Tulane University.
- Kimper, Wendell. 2011. Competing triggers: Transparency and opacity in vowel harmony. PhD diss, University of Massachusetts, Amherst.
- Kimper, Wendell. 2016a. 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. Linguistic Society of America. <https://doi.org/10.3765/amp.v3i0.3662>.
- Kimper, Wendell. 2016b. Positive constraints and finite goodness in harmonic serialism. In *Harmonic Grammar and Harmonic Serialism*, eds. John J. McCarthy and Joe Pater, 221–235. London: Equinox.
- 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 Rutgers Optimality Workshop 1, Rutgers University.
- Kiparsky, Paul. 1994. Remarks on markedness. Handout from Talk presented at TREND 2, Stanford University.
- Kiparsky, Paul. 2008. Universals constrain change; change results in typological generalizations. In *Linguistic universals and language change*, ed. Jeff Good, 23–53. Oxford: Oxford University Press.

- Kirby, Simon. 2017. Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin & Review* 24: 118–137.
- Kirby, Simon, and James Hurford. 2002. The emergence of linguistic structure: An overview of the iterated learning model. In *Simulating the evolution of language*, eds. Angelo Cangelosi and Domenico Parisi, 121–148. London: Springer.
- Knowles, Susan Marie. 1984. A descriptive grammar of Chontal Maya: San Carlos dialect. PhD diss, Tulane University.
- Kornfilt, Jaklin. 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*, 884–891. Hillsdale, NJ: Lawrence Erlbaum Associates.
- 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: MIT Press.
- Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig, eds. 2013. *Ethnologue: Languages of the world, seventeenth ed.* Dallas: SIL International. <http://www.ethnologue.com/>.
- Lin, Hua. 2001. *A grammar of Mandarin Chinese*, Vol. 344. Munich: 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.
- Macaulay, Monica Ann. 1996. *A grammar of Chalcatongo Mixtec*, Vol. 127. Berkeley, Los Angeles, London: University of California Press.
- Maddieson, Ian, and Kristin Precoda. 1992. The UPSID database. UCLA.
- Maddieson, Ian, Sébastien Flavien, Egidio Marsico, and François Pellegrino. 2014–2019. LAPSyD: Lyon-Albuquerque Phonological Systems Databases, Version 1.1. <http://www.lapsyd.ddl.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.
- McCollum, Adam. 2018. Vowel dispersion and Kazakh labial harmony. *Phonology* 35(2): 287–326.
- McGregor, William. 1990. *A functional grammar of Gooniyandi*. Amsterdam: John Benjamins.
- Merlan, Francesca. 1982. *Mangarayi*. London: Routledge.
- Miller, Wick R. 1965. *Acoma grammar and texts*. Vol. 40 of *University of California publications in linguistics*. Berkeley: University of California Press.
- Moran, Steven, and Daniel McCloy. 2019. PHOIBLE 2.0. Available at <http://phoible.org>; Accessed on 2019-09-04.
- Moreton, Elliott. 2008. Analytical bias and phonological typology. *Phonology* 25: 83–127.
- Moreton, Elliott. 2010. Underphonologization and modularity bias. In *Phonological argumentation: Essays on evidence and motivation*, ed. Steve Parker, 79–101. London: Equinox.
- 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. <https://doi.org/10.1111/cogs.12319>.
- Morley, Rebecca. 2015. Deletion or epenthesis? On the falsifiability of phonological universals. *Lingua* 154: 1–26.
- Mortensen, David. 2004. Preliminaries to Mong Leng (Hmong Njua) phonology. Ms., UC Berkeley.
- Newman, Paul. 2000. *The Hausa language: An encyclopedic reference grammar*. New Haven: Yale University Press.

- Noonan, Michael. 1992. *A grammar of Lango*. Vol. 7 of *Mouton grammar library*. Berlin, New York: Mouton de Gruyter.
- Noss, Richard B. 1964. *Thai Reference Grammar*. Department of State, Foreign Service Institute.
- Ohala, Manjari. 1983. *Aspects of Hindi phonology*. Delhi, Varanasi, Patna: Motilal Banarsidass Publishing House.
- O'Hara, Charlie. 2017. How abstract is too abstract: Learning abstract underlying representations. *Phonology* 34(2): 325–345.
- O'Hara, Charlie. 2021. *Soft biases in phonology: Learnability meets grammar*. PhD diss, University of Southern California.
- Okell, John. 1969. *A reference grammar of colloquial Burmese (two volumes)*. London: Oxford University Press.
- 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*, 1–28. San Diego: 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. United Kingdom: 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*. PhD diss, University of California, Los Angeles.
- 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. Association for Computational Linguistics.
- Polomé, Edgar C. 1967. *Swahili language handbook*. *Language handbook series*. Washington: Center for Applied Linguistics.
- 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: 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.
- Qafisheh, Hamdi A. 1977. *A short reference grammar of Gulf Arabic*. Tucson: University of Arizona Press.
- R Core Team. 2014. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Rice, Keren. 1976. *Hare phonology*. PhD diss, University of Toronto.
- Rice, Keren. 2008. de Lacy (2006). Markedness: Reduction and preservation in phonology [Review]. *Phonology* 25(2): 361–371.
- Ridouane, Rachid. 2014. Tashlhiyt Berber. *Journal of the International Phonetic Association* 44(2): 207–221. <https://doi.org/10.1017/S0025100313000388>.
- Roberts, John R. 2016. *Amele RRG Grammatical Sketch*. SIL International.
- Romeo, Luigi. 1964. Toward a phonological grammar of Modern Spoken Greek. *Word* 20: 60–78.
- Romero-Figeroa, Andrés. 1997. *Warao*. Munich: Lincom Europa.
- 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, Vol. 17: Languages*, ed. Ives Goddard, 440–482. Washington: Smithsonian Institution.
- Rosenblatt, F. 1958. The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* 65: 386–408.
- Round, Erich R. 2009. *Kayardild morphology, phonology and morphosyntax*. PhD diss, Yale University.
- Sadock, Jerrold M. 2003. *A grammar of Kalaallisut: (West Greenlandic Inuttut)*, Vol. 162. Munich: Lincom.
- Samarin, William J. 1967. *A grammar of Sango*. The Hague: Mouton.
- 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, Robert N., Carl I. Hovland, and Herbert 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 Archives.
- SIL. 2011b. Daga organised phonology data. SIL Language and Culture Archives.
- Smolensky, Paul, Matt Goldrick, and Donald Mathis. 2014. Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science* 38: 1102–1138.
- Sridhar, S. N. 1990. *Kannada: Descriptive grammar. Croom helm descriptive grammars*. London: Routledge.
- 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.
- Staubs, Robert, Michael Becker, Christopher Potts, Patrick Pratt, John J. McCarthy, and Joe Pater. 2010. OT-Help 2.0. Software package.
- Svantesson, Jan-Olof, Anna Tsendina, Anastasia Karlsson, and Vivan Franzén. 2005. *The phonology of Mongolian. Phonology of the world's languages*. Oxford: Oxford University Press.
- Terrill, Angela. 1999. Lavukaleve: A Papuan language of the Solomon Islands. PhD diss, Australian National University, Canberra.
- Tesar, Bruce, and Paul Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.
- Thayer, Linda Jean. 1974. A reconstructed history of the Chari languages. Comparative Bongo-Bagirmi-Sara segmental phonology with evidence from Arabic loanwords. PhD diss, University of Illinois, Urbana-Champaign.
- Thompson, Laurence C. 1965. A Vietnamese grammar. PhD diss, University of Washington.
- Thoudam, Purna Chandra. 1980. A grammatical sketch of Meiteiron. PhD diss, Jawaharlal Nehru University.
- Tola, Waqo. 1981. The phonology of Mecha Oromo. Master's thesis, Addis Adaba University.
- Topintzi, Nina, and Andrew Nevins. 2017. Moraic onsets in Arrente. *Phonology* 34(3): 615–650.
- Tryon, Darrell T. 1970. *Conversational Tahitian: An introduction to the Tahitian language of French Polynesia*. Canberra: Australian National University Press.
- 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* 44: 76–90.
- 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 Vol. 17: Languages*, ed. Ives Goddard, Vol. 17, 390–439. Washington: Smithsonian Institute.
- Yanushevskaya, Irena, and Bunčić. 2015. Russian. *Journal of the International Phonetic Association* 45(2): 221–228.
- Yoshioka, Noburu. 2012. A reference grammar of Eastern Burushaski. PhD diss, Tokyo University of Foreign Studies.
- Zúñiga, Fernando. 2000. *Mapudungun*. Munchen: Lincom Europa.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.