Proceedings of the Society for Computation in Linguistics

Volume 2 Article 42

2019

Place and Position are Computationally Different

Charlie O'Hara
University of Southern California, charleso@usc.edu

Recommended Citation

O'Hara, Charlie (2019) "Place and Position are Computationally Different," *Proceedings of the Society for Computation in Linguistics*: Vol. 2 , Article 42.
DOI: https://doi.org/10.7275/2jhc-dx44
Available at: https://scholarworks.umass.edu/scil/vol2/iss1/42

Place and Position are Computationally Different*

Charlie O'Hara
University of Southern California
charleso@usc.edu

1 Introduction

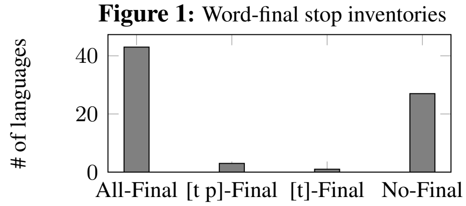
Pater and Moreton (2012) and others have argued that learning biases against complex patterns lead to underrepresentation of such patterns cross-linguistically. For the purposes of this paper, complexity can be reduced to featural complexity, that is, the fewer features needed to describe a pattern, the simpler it is. However, computational features do not necessarily map cleanly onto the classic sets of phonological features needed to distinguish different segments. Intuitively, an inherent featural property of a segment like place of articulation is different than a contextually derived property, like its syllable position. However, a null hypothesis would suggest that a segment being [+dorsal] need not be computationally distinct from that segment being [+coda]; and thus a number of previous studies have implemented these two properties as computationally equivalent features.

O'Hara (2018) shows that following intuition, the typology of voiceless stops demonstrates a difference in attestation rate between patterns based on place of articulation compared to patterns based on syllable position. Further, O'Hara (2018) demonstrates that a set of constraints motivated in order to capture the factorial typology in classical optimality theory interact with a model of phonological learning to predict the observed typology. This paper extends this result, analyzing exactly how different these features need to be encoded in order to capture the typological results.

2 Probabilistic Typology

O'Hara (2018) reports on a typological survey of word initial and word final stop inventories from 170+ language grammars available at the libraries at USC and UCLA and online. In order to have some amount of control for the amount of data available to language learners, I focus on the 77 languages (from 25 language families) that allow only three supralaryngeal places of articulation for stops. Of these 77 languages, 90.9% allow either all or none of the stops available word-initially to be available word-finally, as shown in Figure 1.

Figure 1: Word-final stop inventories



This result suggests that patterns defined just using syllable position are better attested than patterns that require the interaction of syllable position and place of articulation. This result supports the hypothesis that featurally simple patterns should be well attested—because the No-Final pattern can be defined using just [coda] as in (2), but the [pt]-Final requires an interaction of [coda] and [dorsal] as in (3).

(1) ***Simple:*** *All-Final*
($\frac{43}{77} = 56\%$ languages)

tV	pV	kV
Vt	Vp	Vk

(2) ***Simple:*** *No-Final, *[coda]*
($\frac{27}{77} = 35\%$ languages)

tV	pV	kV
~~Vt~~	~~Vp~~	~~Vk~~

(3) ***Complex:*** *[pt]-Final, *[coda]&[dorsal]*
($\frac{3}{77} = 3.9\%$ languages)

tV	pV	kV
Vt	Vp	~~Vk~~

*Thanks to Karen Jesney, Rachel Walker, Arto Anttila, Coral Hughto, Gaja Jarosz, Joe Pater, Jason Riggle, and Caitlin Smith; as well as audiences at USC, AMP 2017, LSA 2018, and UMass Amherst. All errors are my own.

Proceedings of the Society for Computation in Linguistics (SCiL) 2019, pages 342-345.
New York City, New York, January 3-6, 2019

On the other hand, this hypothesis would also predict that the No-Dorsal pattern in (4) would be similarly well attested. O'Hara (2018) supplements the survey looking at all the languages in UPSID (Maddieson, 1984) that lack one of these places of articulation, as well as languages noted by de Lacy (2006) to have a subset of the [p t k] inventory. The result is that no language of the No-Dorsal type exist. Any language that has less than three supralaryngeal places of articulation for stops in all positions also bans final stops altogether; e.g. Tahitian (Tryon, 1970), which has the inventory in (5).

(4) ***Simple:*** *No-Dorsal, *[dorsal]*
(**Unattested**)

tV	pV	~~kV~~
Vt	Vp	~~Vk~~

(5) ***Complex:*** *[pt]-Initial, *[dorsal]∨[coda]*
(At least 11 languages (O'Hara, 2018))

tV	pV	~~kV~~
~~Vt~~	~~Vp~~	~~Vk~~

Thus, we see that the typology does not straightforwardly replicate a bias towards featurally-simple patterns; at least if place of articulation and position are equal features. Languages are more likely to define their inventories using just position than an interaction of position and place; but these inventories that are defined through an interaction of the features are more common than inventories that only use place of articulation. This result suggests that syllable position is somehow more powerful for the purpose of learnability than the place of articulation feature.

3 Model

I used a generational model of learning to model the effect of learning biases on typology, using the Soft Typology Tool, available on my website.[1] Harder to learn languages incur more transmission errors, which in turn leads to instability across generations. Simulations were run (following Dowman et al. (2006); Staubs (2014)) where a learner was exposed to a limited number of forms from the target grammar, sampled randomly. After this limited number of forms, the learner "matured" and taught a new learner the grammar it had learned, and so on for 20 generations. Each learner was a MaxEnt grammar trained with the truncated Perceptron algorithm (Magri, 2015).

[1] dornsife.usc.edu/ohara/softtyptool/

In the papers presented in this paper, the number of forms presented to each learner were 3600, the learning rate for all constraints was .05, and learners were initially set with output-oriented constraints weighted at 50 and faithfulness constraints weighted at 1. These numbers were selected somewhat arbitrarily; however it was checked that all learners in the first generation were closer to the target pattern than any other pattern before iteration 3300. This ensures that all patterns *are learnable* in the time allotted. As long as this is true, these relative stability metrics remain constant across different settings of the parameters.

For this paper, in order to test how different constraint sets treat these features differently, I focused on both of the so-called "simple" patterns: No-Final and No-Dorsal, as well as All-Final. Simulations were run with each pattern 50 times, and the closest categorical pattern after 20 generations was noted. Patterns that stay the same more often across twenty generations are easier-to-learn, and less likely to decay across generations, one factor in whether that pattern would be common.

4 Simulations

Simulations were run using three sets of constraints. First, in the Unbiased Condition, a set of constraints with no substantive biases was tested (6); this model treats all features *and feature values* as equivalent; no markedness hierarchies are encoded. For any markedness constraint *K, an anti-markedness constraint +K exists that rewards dorsals. In this case, all forms are equally likely to be produced initially (0.576 probability). If the target pattern includes a repair (e.g. /kV/→[ʔV]), this probability will go down, and that repair will gain probability. Similar to the results seen in Pater and Moreton (2012)'s GMECCS model, both featurally simple patterns are found to be near equally stable, as seen in Table 1. Thus this model fails to capture a bias against the No-Dorsals pattern over the No-Finals pattern.

(6) Unbiased Condition Constraints:
*kpt, *k, *p, *t, NoCoda, Onset, +Coda, NoOnset, +k, +p, +t +kpt, Max, Ident[Pl]

One issue in the unbiased condition is that there are no mechanisms to capture the markedness implications well known in the literature. The next two conditions tested learners with constraints meant to encode the markedness hierarchies observed on these two scales in previous literature shown in (7-8).

(7) Positional Hierarchy (Goldsmith, 1990)
Onset $\succ$[2] Coda

(8) Place Hierarchy (Kean, 1975; Lombardi, 2001; de Lacy, 2006)
t $\succ$ p $\succ$k

First, the constraints were defined as in (9) so as to make the place of articulation and positional scales equivalent. In this EQUIVALENT CONDITION, specific (and stringently related) markedness constraints encoded the markedness (NOCODA, *k, *kp); and specific faithfulness constraints favored less marked items (IDENT/ONS, IDENT/T, IDENT/TP; based on positional faithfulness (Beckman, 1998)). While this type of CON has rarely if ever been proposed in the literature, it is the result of treating place and position as equivalent features. The results of this simulation, again in Table 1, are even worse than in the unbiased condition, with the nonattested No-Dorsal pattern being the most stable, and the common No-Final pattern being very unstable.

(9) Equivalent Condition Constraints:
*KPT, *KP, *K, NOCODA, MAX, IDENT, IDENT/ONS, IDENT/PT, IDENT/T

Finally, in the DISTINCT CONDITION these hierarchies were encoded as in (10), where the differences between place and position that have been discovered in the literature are all encoded. There are two major distinctions between the two hierarchies, one encoded in markedness, one encoded in faithfulness. First, onsets are protected by an additional markedness constraint ONSET; so deleting an onset stop is worse on two constraints than deleting a coda stop (ONSET and NOCODA). Secondly, while specific faithfulness constraints protect privileged (or unmarked) positions; they protect MARKED places of articulation, a la de Lacy (2006)'s Marked Faithfulness (IDENT-K, IDENT-KP). Table 1 shows that this model successfully captures the observed typology; finding the All-Final and No-Final patterns more stable than the unattested No-Dorsal pattern.

[2] $x \succ y$ here means x is more harmonic than (less marked than) y.

Table 1: Simulation Results (# Stable out of 50)

Constraint Set	All-Final	No-Final	No-Dorsal
Unbiased Condition	47	40	43
Equivalent Condition	30	0	47
Distinct Condition	48	41	11

(10) Distinct Condition Constraints:
*KPT, *KP, *K, NOCODA, ONSET, MAX, IDENT, IDENT/ONS, IDENT/KP, IDENT/K

5 Discussion

The simulations above show that the observed typology can only be modeled as a learning bias by encoding the place and positional markedness hierarchies differently. To understand why these three constraint sets perform so differently on these patterns, it can be useful to think about how the set of constraints contribute to the learning dynamics for each form in the simulation.

The probability of a form surfacing faithfully is completely dependent on the harmonic difference between the faithful candidate and each of the repair candidates, given the current constraint weights. Therefore, it is useful to think of the elementary weighting conditions (or EWCs) of each potential error. The probability of the target candidate does not approach 1 until its harmony is sufficiently greater than each of its competitors (the harmonic difference of the target candidates EWCs is sufficiently positive). In order to intuit about these dynamics it is important to understand both the INITIAL HARMONIC DIFFERENCE (IHD) of all of these EWCs, as well as the expected rate of change.

Because markedness constraints are initially weighted high in these simulations, they contribute greatly to the IHD. If the target candidate violates a markedness constraint that a competitor does not, this lowers the IHD by 50; but if the competitor violates one that the target satisfies, the IHD goes up by 50.

The expected rate of change of the harmonic difference of an EWC is dependent on the expected rate of change of all constraints that the target candidate and the competitor differ on. The rate of change of a constraint itself is dependent on how likely an observed error will cause an update on the weight of that constraint, and the probability of an error is itself dependent on the current harmonic difference of the EWC represented by that error. While the expected rate of change on the next iteration given a current set of constraint weightings is simple to calculate, the function of the expected rate of change over time is much more difficult, lending itself to simulation rather than analysis. However, simplifying things, for each constraint that the target candidate and competitor differ on, the EWC moves faster, though the amount and direction of this is dependent on how *general* and *consistent* that constraint is. (For the most part, more EWCs differ on general constraints than specific constraints; and consistent constraints are updated in the same direction by more observed errors.) Because faithfulness constraints make a small contribution to IHD, most of their effect is observed in the expected rate of change of the EWCs that differ on them.

Consider the Distinct Condition. The IHD between /kV/→[kV] and its debuccalization competitor /kV/→[ʔV] is -146. The difference is larger in magnitude than the difference between the target faithful candidate and the deletion competitor (-99), so debuccalization is the learner's initially preferred repair. Every time the learner observes [kV] and produces [ʔV], the harmonic difference updates by seven times the learning rate. Making debuccalization errors on /pV/ and /tV/ will also update this harmonic difference by a factor of five and three respectively. This leads to fast learning of the No-Final pattern in this condition.

Compare this to the Equivalent Condition. The IHD between faithful [kV] and debuccalization is comparable to the Distinct Condition (-148); though the lack of ONSET in this constraint set means deletion is a near equally likely repair initially (-149). The closeness of these two competitors means that the errors observed will be less consistent about what constraints get updated (IDENT constraints will update at least half as fast). Further, there are no longer as many faithfulness constraints preventing debuccalization of /kV/, meaning that a debuccalization error will only update the harmonic difference between [kV] and [ʔV] by a factor of five, rather than seven as in the Distinct Condition. Both the competition between errors and the slower update speed of the debuccalization EWC lead to slower learning of No-Final in the Equivalent condition than the Distinct Condition, helping lead to the instability of this pattern.

In order to capture the observed typology through this learning bias, it is necessary that not only are the place and position markedness hierarchies encoded in the constraint set, but that they are encoded in distinct ways. There is something substantively different about place of articulation and syllable position that is visible in the probabilistic typology of stop inventories. This paper warns that syllable position does not act equivalently to distinctive features, and this difference should be heeded in learning and computational work.

References

Jill N. Beckman. 1998. *Positional Faithfulness*. Ph.D. thesis, University of Massachusetts Amherst, Amherst.

Mike Dowman, Simon Kirby, and Thomas L. Griffiths. 2006. Innateness and culture in the evolution of language. In *The Evolution of Language: Proceedings of the 6th international conference on the Evolution of Language*.

John A. Goldsmith. 1990. *Autosegmental and Metrical Phonology*. Blackwell.

Mary-Louise Kean. 1975. *The Theory of Markedness in Generative Grammar*. Ph.D. thesis, Massachusetts Institute of Technology.

Paul de Lacy. 2006. *Markedness: Reduction and Preservation in Phonology*. Cambridge University Press, Cambridge.

Linda Lombardi. 2001. Why place and voice are different: Constraint-specific alternations in Optimality Theory. In Linda Lombardi, editor, *Segmental phonology in Optimality Theory: Constraints and Representations*, pages 13–45. Cambridge University Press, Cambridge.

Ian Maddieson. 1984. *Patterns of sounds*. Cambridge University Press, Cambridge.

Giorgio Magri. 2015. How to keep the HG weights non-negative: the truncated perceptron reweighting rule. *Journal of Language Modeling*, 3(2):345–375.

Charlie O'Hara. 2018. Emergent learning bias and the underattestation of simple patterns. Ms. University of Southern California.

Joe Pater and Elliott Moreton. 2012. Structurally biased phonology: complexity in language learning and typology. *The EFL Journal*, 3(2):1–44.

Robert Staubs. 2014. *Computational modeling of learning biases in stress typology*. Ph.D. thesis, University of Massachusetts Amherst, Amherst.

D. T. Tryon. 1970. *Conversational Tahitian: An Introduction to the Tahitian Language of French Polynesia*. Australian National University Press.