Using simulated learning to account for tone typology

Breteler, J.; Boersma, P.

DOI
10.21437/TAL.2018-11
10.21437/TAL.2018

Publication date
2018

Document Version
Final published version

Published in
Proceedings of TAL2018, Sixth International Symposium on Tonal Aspects of Languages

TAL2018, Sixth International Symposium on Tonal Aspects of Languages
18-20 June 2018, Berlin, Germany

Using simulated learning to account for tone typology

Jeroen Breteler[1], *Paul Boersma*[1]

[1]University of Amsterdam, the Netherlands
j.m.w.breteler@uva.nl, paul.boersma@uva.nl

Abstract

This study reports on the first investigation into the learnability of tone spread and shift patterns, as determined by computer simulations of the learning process. Our simulations are cast within the context of a synchronic analytical framework developed in earlier work. The framework uses licensing constraints and foot structure to drive various kinds of tonal reassociation. One problem with the framework was that it was able to generate various unattested patterns. We address this problem through the learnability simulations; our results show that the representable-but-unattested patterns are harder to learn, explaining their non-attestation. This way, we demonstrate that learnability simulations are a meaningful tool for a typological account of tonal phenomena.

Index Terms: typology, phonological learning, tone shift, tone spread, tonal reassociation, foot structure

1. Introduction

Bantu languages display a variety of tonal phenomena where tone surfaces in positions that it did not occupy in the lexical form, even when there is no apparent phonological trigger to do so (such as vowel elision). These processes are commonly known as *tone shift* and *tone spread*, but some tone patterns do not exactly fit either of these labels. Consequently, we will refer to these patterns collectively as cases of *tonal reassociation* (TR). Our study is aimed at a typological account of tonal reassociation. We will do this using a novel tool: computer simulations that test the learnability of TR patterns in a given analytical framework.

A common way to account for the typology of some phonological phenomenon is to develop a theory of its synchronic analysis [1, 2, 3]. A synchronic framework accounts for a typology to the extent that it offers analyses for attested patterns while *not* accomodating the analysis of unattested patterns. However, in practice, there will be a less than perfect fit between a given pair of framework and typological facts. In particular, the framework might *overgenerate*, meaning that it offers analyses for patterns that are unattested (and that a linguist deemed unattestable). In the face of overgeneration, learnability results can be an additional source of typological explanation. This is the case when learnability results distinguish between attested and analyzable-but-unattested patterns. If simulations show higher rates of learning success for attested patterns than for analyzable-but-unattested ones, this can explain why patterns that are perfectly analyzable within a synchronic framework might still not be attested. Consequently, learnability simulations offer a means of tightening a set of typological predictions, going beyond the commonly employed criterion of analyzability in a synchronic framework.

Our study builds on results from theoretical work by the first author [4], which developed a framework that uses foot structure to drive and direct tonal reassociation [5]. While this framework successfully accounts for a range of attested TR patterns, it also allows for the representation of some TR patterns that have not been attested. In other words, the framework suffers from *overgeneration*. In this study, we improve the foot-based typological account by showing that representable-but-unattested patterns are less consistently learnable than attested patterns.

In the next section, we show some examples of tonal reassociation and outline the typological facts. In section 3 we overview synchronic analytical approaches to TR typology. Section 4 discusses previous literature about using learning for typological investigations, as well as the methodology for our learning simulations. Section 5 presents the results.

2. Varieties of tonal reassociation

A major distinction made for TR patterns is whether they are *bounded* or *unbounded*. This term refers to the distances across which tonal reassociation takes place. Bounded patterns operate over a span of material that is defined relative to the position of the lexical tone — we will call this the *sponsor* position. For example, Rimi displays a bounded shift pattern where tone targets the position immediately following the sponsor [6]. An example alternation is shown in (1) [7]. We have interpreted Olson's transcriptions from [6] to approach IPA, writing [o] for the second-highest back vowel among four levels of vowel height. Accents on vowels denote High tone; we assume other vowels are toneless, which we do not mark in any way. Hyphens denote morpheme boundaries.

(1) Bounded rightward shift in Rimi
  a. mo-nto ‘person’
  b. <u>ɾa</u>-mó-nto ‘of a person’

The datum in (1a) is toneless throughout, yet the same string surfaces with a High tone in (1b). Consequently, the most plausible source for this High tone is that it originated on the prefix [ɾa], and shifted to its surface position.

For *unbounded* TR patterns, the positions affected by reassociation are defined not only relative to the sponsor, but also to the edge of the relevant prosodic domain. For example, Phuthi shows an unbounded spreading pattern that runs from the sponsor to the antepenultimate syllable of the word [8, 9]. We show an example alternation in (2). We made an interpretation approaching IPA transcriptions based on [8, pp. 68–73, 487]; Donnelly reports that a tenseness contrast in mid vowels is not reflected in his transcriptions. In (2), swapping out the toneless subject prefix [si] for the High-toned prefix [ɓá] triggers tone spreading over an additional five syllables, up to the antepenultimate position.

(2) Unbounded spread to the antepenult in Phuthi
  a. si-ja-lima-lim-el-aː-na
  ‘we cultivate for each other now and then’
  b. <u>ɓá</u>-já-límá-lím-él-aːn-a
  ‘they cultivate for each other now and then’

10.21437/TAL.2018-11

3. Analyses of tonal reassociation

Traditionally, metrical accent (though not always foot structure) was the tool of choice for the analysis of unbounded TR patterns; e.g. [10, 11]. In addition, a foot-based account was proposed for the bounded tone shift pattern of Sukuma [12]. However, with the advent of Optimality Theory (OT) [13], attention in the literature turned towards alternative representations based on featural domains, namely Optimal Domains Theory [14, 15] and Headed Spans [16] — though one notable exception is Bickmore's analytical framework based on alignment effects [17].

One possible reason for the diminished interest in foot-based solutions is that OT runs into complications in the analysis of foot-based bounded tone shift, which is an opaque analysis. Informally, such a shift can be understood as mapping through an intermediate form where the foot is positioned relative to the tone but the tone itself has not yet reassociated. That is, with σ representing a syllable, the mapping can be written as $/\sigma\acute{\sigma}\sigma\sigma/ \rightarrow \sigma(\acute{\sigma}\sigma)\sigma \rightarrow [\sigma(\sigma\acute{\sigma})\sigma]$. However, OT does not have a concept of intermediate forms. More generally, through the restricted nature of its faithfulness constraints, OT limits the possibility of incorporating lexical information into decisions about surface forms. Consequently, analyzing bounded tone shift using feet is problematic in OT.

A second challenge for approaches using (binary) feet is that those feet are not a good tool to cover sequences of three tone-bearing units. This is a problem because some bounded TR patterns operate over such ternary spans [18, 19].

Previous work by the first author addressed these problems using recent advances in phonological theory [4]. Here, we leave out several innovations that are not pertinent to the learnability studies. One aspect of the approach that we wish to elaborate on is the use of licensing constraints [20]. In our framework, feet acts as licensors, and tones are motivated to reassociate to footed positions [21]. One constraint that has this effect is LICENSE(H, FT-R), defined in (3) below.

(3) LICENSE(H, FT-R) Assign one * for each tone that is not associated to the rightmost position in some foot.

Together with regular constraints on foot placement and faithfulness, licensing constraints can give rise to a variety of tonal reassociation effects. We demonstrate this in Table 1, which includes LICENSE(H, FT-R) as well as ALL-FT-RIGHT, which pulls feet rightward, and DEP-LINK, which militates against creating new tonal association links. Depending on the ranking of the constraints, this constraint set can generate static tone as in candidate 1a, bounded tone spreading in 1b, or unbounded tone spreading as in 1c.

Table 1: *Tone licensing for bounded and unbounded spread*

$/\acute{\sigma}\sigma\sigma\sigma\sigma/$	$\mathcal{L}$(H, FT-R)	ALL-FT-R	DEP-LINK
a. $\acute{\sigma}\sigma\sigma\sigma\sigma$	*		
b. $(\acute{\sigma}\acute{\sigma})\sigma\sigma\sigma$		***	*
c. $\acute{\sigma}\acute{\sigma}\acute{\sigma}(\acute{\sigma}\acute{\sigma})$			****

Previous work also calculated a factorial typology for the framework, to determine the kinds of patterns it can represent. In addition to accounting for a range of tonal reassociation patterns, the framework generates various types of unattested patterns. In particular, the framework predicts irregular activity for tone at the edges of the prosodic domain. The most extreme among these predicted, unattested patterns are so-called "initial-only" patterns, where tone in the initial position of a form triggers different tonal behavior than tone in other positions. For example, an "initial-only final spreading" pattern shows spreading to the final position if tone was underlyingly initial, while it is static in other positions. One of the interests of the present study is to see whether such extreme patterns are learnable at all, and whether their unattestedness can be explained as a consequence of poor learnability.

4. Simulated learning

4.1. Learning and typology

Since the beginning of research into Optimality Theory, there has been interest in the question of how to learn OT grammars [22, 23]. Most OT learning studies were about the learning challenge itself, but there has also been an interest in the informativeness of learning *failure*, i.e. the imperfect transfer between (simulated) adults and learners [24]. Staubs [25, 26] shows that learning biases explain the limited size of stress windows, obviating the need for the explicit stating of such size limits in grammar theory. Similarly, Stanton [27] shows that learning considerations can provide an alternative account for the typological finding that no language places stress on the middle positions of a stress domain. Specifically, learning such middle-stress patterns is complicated by the fact that exposure to longer words is necessary to disambiguate the pattern from edge-based patterns, and that the used learning algorithm extracts conflicting information from the various data types it is exposed to. Stanton's finding could vindicate theories of grammar that allow the representation of such systems, which were previously criticized for displaying this "midpoint pathology" [28]. To our knowledge, there is no previous work on the learnability of tone patterns, neither for its own sake nor within the context of typological questions. As our study demonstrates, tonology offers interesting technical challenges for learnability theory, and learnability offers new avenues for addressing tone typology.

4.2. Methodology

Our simulations use the Gradual Learning Algorithm [29, 30], which is an *error-driven* algorithm. This means that a simulated learner checks whether their behavior matches that of the adult they are learning from. Whenever the learner finds that this is not the case, they conclude that some aspect of their grammatical beliefs must be erroneous, and proceed to update their grammar in accordance with the error. The algorithm is called "gradual" because the learner is careful not to overgeneralize from a single error. A single learning update will only entail a small change in the learner's grammar; the learner approaches the target grammar over the course of learning from multiple mistakes. In our simulations, learners typically needed some hundreds of examples to converge on the target grammar, although we provided a more liberal 40,000 tokens in total to each learner, to ensure that any learning failure was not the result of a lack of examples.

The adult examples that the learner processes are pairs of what we call a "morphological" and an "overt" form. This pairing symbolizes the learning situation that children are in;

they hear utterances, and have some idea of the semantic context that is concurrent with the utterance. The overt form represents the level closest to the acoustic reality that learners are exposed to. In our simplified model, it is an impoverished phonological form; it contains the number of syllables and discrete pitch information (High vs. toneless), but not autosegmental structure or foot structure. The morphological form contains a sequence of semantic units that the learner can relate to lexical forms. Crucially, we follow earlier work in keeping some intermediate levels of representation *hidden* [23, 31, 32]. We do not inform the learner about the phonological structure or the lexical forms that the adult grammar generates; the learner should deduce such knowledge themselves [23]. For example, a learner might be presented with the pair of an overt form $[\acute{\sigma}\acute{\sigma}\sigma\sigma\sigma]$ and a morphological form <A+b+c+d+e>. From just the overt form, the learner cannot tell if there are one or two tonal autosegments, and whether there are any feet anywhere in the adult's structure for this form. The morphological form contains no phonological information (hence our use of meaningless letters of the alphabet), but it does tell the learner that a specific sequence of five morphemes is associated with this overt form. These morphological forms become informative when the learner has to deal with *alternations*. So, if the learner also processes pairs of $[\sigma\sigma\sigma\sigma\sigma]$ with <a+b+c+d+e>, it has the possibility of spotting a minimal pair: Forms with <A> have High tone on the first two syllables, and forms with <a> do not! A successful learner will deduce that the lexical material associated with <A> is a High-toned syllable, $|\acute{\sigma}|$ (we use pipes || to denote lexical forms). Furthermore, the successful learner will adopt a constraint ranking so that $|\acute{\sigma}\sigma\sigma\sigma\sigma|$ triggers a spreading process, presumably by using a foot, so that the surface structure comes out as $/(\acute{\sigma}\acute{\sigma})\sigma\sigma\sigma/$ (here we use slashes // for phonological surface forms).

4.3. Target patterns

We tested learning for five attested and six unattested patterns, all of which were representable in the foot-based framework discussed in section 3. In Table 2, we list the adult examples for all of these patterns. Each pattern has three overt forms surfacing with High tone, and one "toneless" overt form that surfaces without any High tone. Each overt form is paired with its own morpheme form, so that the learner can deduce phonological structure and lexical forms by considering alternations. Since the Generalized Learning Algorithm is also sensitive to frequencies in the input, we have boosted representation for the toneless pairing, so that it occurs equally often as all of the High-toned forms put together.

Among the attested patterns is one bounded pattern (Binary Spreading), and four unbounded patterns that target either the final or penultimate syllable with either a spreading or shifting pattern. The unattested patterns include the "initial-only" patterns discussed above, where forms with initial sponsors behave differently from all other forms. In addition, there are two patterns that show "edge doubling", a pattern where doubling occurs only if the sponsor is sufficiently close to the edge. That is, these patterns show spreading for forms in the third column, to the exclusion of spreading in the other columns.

5. Results

We ran 100 learning trials for each of the eleven patterns listed in Table 2. At the end of a learning trial, we tested if the virtual learner managed to consistently reproduce the target

Table 2: *Sets of overt–morphology pairs for all patterns*

<A+b+c+d+e>	<a+B+c+...>	<a+b+C+...>	<a+b+c+...>
Attested patterns			
Binary Spreading			
$[\acute{\sigma}\acute{\sigma}\sigma\sigma\sigma]$	$[\sigma\acute{\sigma}\acute{\sigma}\sigma\sigma]$	$[\sigma\sigma\acute{\sigma}\acute{\sigma}\sigma]$	$[\sigma\sigma\sigma\sigma\sigma]$
Penult Spreading			
$[\acute{\sigma}\acute{\sigma}\acute{\sigma}\acute{\sigma}\sigma]$	$[\sigma\acute{\sigma}\acute{\sigma}\acute{\sigma}\sigma]$	$[\sigma\sigma\acute{\sigma}\acute{\sigma}\sigma]$	$[\sigma\sigma\sigma\sigma\sigma]$
Final Spreading			
$[\acute{\sigma}\acute{\sigma}\acute{\sigma}\acute{\sigma}\acute{\sigma}]$	$[\sigma\acute{\sigma}\acute{\sigma}\acute{\sigma}\acute{\sigma}]$	$[\sigma\sigma\acute{\sigma}\acute{\sigma}\acute{\sigma}]$	$[\sigma\sigma\sigma\sigma\sigma]$
Penult Shift			
$[\sigma\sigma\sigma\acute{\sigma}\sigma]$	$[\sigma\sigma\sigma\acute{\sigma}\sigma]$	$[\sigma\sigma\sigma\acute{\sigma}\sigma]$	$[\sigma\sigma\sigma\sigma\sigma]$
Final Shift			
$[\sigma\sigma\sigma\sigma\acute{\sigma}]$	$[\sigma\sigma\sigma\sigma\acute{\sigma}]$	$[\sigma\sigma\sigma\sigma\acute{\sigma}]$	$[\sigma\sigma\sigma\sigma\sigma]$
Unattested patterns			
Final Doubling Shift			
$[\sigma\sigma\sigma\acute{\sigma}\acute{\sigma}]$	$[\sigma\sigma\sigma\acute{\sigma}\acute{\sigma}]$	$[\sigma\sigma\sigma\acute{\sigma}\acute{\sigma}]$	$[\sigma\sigma\sigma\sigma\sigma]$
Initial-Only Binary Spreading			
$[\acute{\sigma}\acute{\sigma}\sigma\sigma\sigma]$	$[\sigma\acute{\sigma}\sigma\sigma\sigma]$	$[\sigma\sigma\acute{\sigma}\sigma\sigma]$	$[\sigma\sigma\sigma\sigma\sigma]$
Initial-Only Binary Shift			
$[\sigma\acute{\sigma}\sigma\sigma\sigma]$	$[\sigma\acute{\sigma}\sigma\sigma\sigma]$	$[\sigma\sigma\acute{\sigma}\sigma\sigma]$	$[\sigma\sigma\sigma\sigma\sigma]$
Initial-Only Final Spread			
$[\acute{\sigma}\acute{\sigma}\acute{\sigma}\acute{\sigma}\acute{\sigma}]$	$[\sigma\acute{\sigma}\sigma\sigma\sigma]$	$[\sigma\sigma\acute{\sigma}\sigma\sigma]$	$[\sigma\sigma\sigma\sigma\sigma]$
Edge Doubling			
$[\acute{\sigma}\sigma\sigma\sigma\sigma]$	$[\sigma\acute{\sigma}\sigma\sigma\sigma]$	$[\sigma\sigma\acute{\sigma}\acute{\sigma}\sigma]$	$[\sigma\sigma\sigma\sigma\sigma]$
Penult Shift, Edge Doubling			
$[\sigma\sigma\sigma\acute{\sigma}\sigma]$	$[\sigma\sigma\sigma\acute{\sigma}\sigma]$	$[\sigma\sigma\acute{\sigma}\acute{\sigma}\sigma]$	$[\sigma\sigma\sigma\sigma\sigma]$
Relative frequency			
1	1	1	3

adult behavior. If so, this counted as one case of successful convergence. Table 3 shows our results, listing successful convergence rates for all the patterns.

Table 3: *Successful convergence rates (N=100)*

Pattern	*Successful learners*
Attested patterns	
Binary Spread	96%
Final Spread	66%
Penult Spread	63%
Final Shift	85%
Penult Shift	79%
Unattested patterns	
Final Doubling Shift	71%
Edge Doubling	18%
Init-Only Bin. Spread	0%
Init-Only Bin. Shift	24%
Init-Only Final Spread	0%
Penult Shift, Edge Doubling	23%

Overall, the results show the typologically desired outcome that the attested patterns are more easily learnable than the unattested patterns. Attested patterns have an average successful convergence rate of 78%, while unattested patterns average only 23%. One of the unattested patterns, Final Doubling Shift, behaves more in line with the attested group (both in this result and other results we are not reporting here where we used other variations of the learning algorithm). Based on our findings, we consider this pattern to be typologically plausible.

6. Possibilities for future work

Our approach has involved many simplifications; here we identify some of those simplifications that we think might be interesting to dispense with in future research. Firstly, we agree with one reviewer who noted that our study has limited typological coverage. We did not consider any TR patterns with ternary phenomena, or with syllable quantity effects, etc. Moreover, we did not offer any sets of adult forms that require an analysis with multiple lexical tones in a single word (although our learner was free to posit such analyses for the patterns we did present). This is a simplification because in multi-tone situations, some TR patterns treat the rightmost tone differently from others, and many TR patterns have special outcomes for situations where lexical tones are in close proximity to one another.

Our approach should also be expanded through the addition of more phonetic detail. In the present study, our most "phonetic" form, the overt form, still contains perfect information about syllabification, and has only the minimal pitch contrast of high versus not-high. In reality, adult forms contain much richer infomation about tonal realization; such information could have a drastic impact on the learnability of TR patterns, and by extension, the typological picture that stems from considering learnability.

A more subtle simplification we made here is to restrict morphemes to lexical forms on a one-to-one basis; every morpheme form has five morphemes in it, and every lexical form has five syllables. Future work could allow for the possibility of multi-syllable morphemes, as well as morphemes consisting only of a floating tonal autosegment.

7. Conclusion

In this paper, we demonstrated for the first time that tonal reassociation patterns are learnable, within an Optimality Theory context using licensing constraints and foot structure. Moreover, we found that learnability corresponds with typological attestation; the attested patterns in our study had far higher learning success rates than the representable-but-unattested patterns.

We look forward to a wider application of learnability simulations to understand the typology of tonal reassociation. Although the present learning task already achieved some degree of realism by hiding phonological and lexical information from the learner, future work is needed to make the learning task resemble real-world situations even more closely.

8. Acknowledgements

We thank two anonymous reviewers for their comments on an earlier version of this paper. We also thank René Kager, for participation in research meetings where this work was discussed. The first author's research was funded by the Netherlands Organisation for Scientific Research (NWO) through the graduate program of the Netherlands Graduate School of Linguistics (LOT), in the context of the project "Language — from cognition to communication" (NWO project number 022.004.015).

9. References

[1] M. Gordon, "A factorial typology of quantity-insensitive stress," *Natural Language & Linguistic Theory*, vol. 20, no. 3, pp. 491–552, 2002.

[2] R. Kager, "Rhythmic licensing theory: An extended typology," in *Proceedings of the 3rd Seoul International Conference on Phonology*. Seoul National University, 2005, pp. 5–31.

[3] M. Gordon, "Typology in Optimality Theory," *Language and Linguistics Compass*, vol. 1, no. 6, pp. 750–769, 2007.

[4] J. Breteler, "Deriving bounded tone with layered feet in Harmonic Serialism: The case of Saghala," *Glossa: A journal of general linguistics*, vol. 2, no. 1, 2017, 57.

[5] ——, "A foot-based typology of tonal reassociation: Perspectives from synchrony and learnability," Ph.D. dissertation, University of Amsterdam, forthcoming.

[6] H. S. Olson, "The phonology and morphology of Rimi," Ph.D. dissertation, Hartford Seminary Foundation, 1964.

[7] S. Myers, "OCP effects in Optimality Theory," *Natural Language & Linguistic Theory*, vol. 15, no. 4, pp. 847–892, 1997.

[8] S. S. Donnelly, "Aspects of tone and voice in Phuthi," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2009, corrected Ms.

[9] S. Donnelly, "Tone and depression in Phuthi," *Language Sciences*, vol. 31, no. 2–3, pp. 161–178, 2009.

[10] J. Goldsmith, "Tone and accent, and getting the two together," in *Annual Meeting of the Berkeley Linguistics Society*, J. Aske, N. Beery, L. Michaelis, and H. Filip, Eds., vol. 13, 1987, pp. 88–104.

[11] L. Downing, "Local and metrical tone shift in Nguni," *Studies in African Linguistics*, vol. 21, no. 3, pp. 261–317, 1990.

[12] B. M. Sietsema, "Metrical dependencies in tone assignment," Ph.D. dissertation, Massachusetts Institute of Technology, 1989.

[13] A. Prince and P. Smolensky, *Optimality Theory: Constraint interaction in Generative Grammar*. Malden, MA: Blackwell Publishing, 1993, technical Report number 2, Rutgers Center for

Cognitive Science, Rutgers University. Published in book form in 2004.

[14] J. Cole and C. Kisseberth, "An optimal domains theory of harmony," *Studies in the Linguistic Sciences*, vol. 24, no. 2, pp. 1–13, 1994.

[15] F. Cassimjee and C. Kisseberth, "Optimal domains theory and bantu tonology," in *Theoretical aspects of Bantu tone*, C. Kisseberth and L. Hyman, Eds. Stanford, CA: CSLI Publications, 1998, pp. 33–132.

[16] M. Key, "Headed Spans and Bantu tonology," 2007, ms, University of Massachusetts Amherst.

[17] L. S. Bickmore, "Bantu tone spreading and displacement as alignment and minimal misalignment," Ms., University at Albany, 1996.

[18] C. Patin, "Tone shift and tone spread in the Saghala noun phrase," *Faits de Langues – Les Cahiers*, vol. 1, pp. 229–244, 2009.

[19] L. S. Bickmore and N. C. Kula, "Ternary spreading and the OCP in Copperbelt Bemba," *Studies in African Linguistics*, vol. 42, no. 2, pp. 101–132, 2013.

[20] C. C. Zoll, "Parsing below the segment in a constraint based framework," Ph.D. dissertation, University of California, Berkeley, 1996.

[21] Y. Kang, "Tone in Sukuma," in *PF: Papers at the interface*, ser. MIT Working Papers in Linguistics, B. Bruening, Y. Kang, and M. McGinnis, Eds. Cambridge, MA: MITWPL, 1997, no. 30, pp. 49–95.

[22] B. Tesar and P. Smolensky, "The learnability of Optimality Theory: An algorithm and some basic complexity results," 1993, ms., University of Colorado at Boulder. ROA-2.

[23] ——, *Learnability in Optimality Theory*. Cambridge, MA: MIT Press, 2000.

[24] P. Boersma, "Review of Bruce Tesar and Paul Smolensky (2000): Learnability in Optimality Theory," *Phonology*, vol. 20, no. 3, pp. 436–446, 2003.

[25] R. Staubs, "Computational modeling of learning biases in stress typology," Ph.D. dissertation, University of Massachusetts Amherst, 2014.

[26] ——, "Learning bias in stress windows: Frequency and attestation," in *Proceedings of the 2014 Annual Meeting on Phonology*, A. Albright and M. A. Fullwood, Eds., vol. 2, 2015.

[27] J. Stanton, "Learnability shapes typology: The case of the midpoint pathology," *Language*, vol. 92, no. 4, pp. 753–791, 2016.

[28] R. Kager, "Stress in windows: Language typology and factorial typology," *Lingua*, vol. 122, no. 13, pp. 1454–1493, 2012.

[29] P. Boersma, "How we learn variation, optionality, and probability," in *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, vol. 21, 1997, pp. 43–58.

[30] P. Boersma and B. Hayes, "Empirical tests of the Gradual Learning Algorithm," *Linguistic Inquiry*, vol. 32, no. 1, pp. 45–86, 2001.

[31] D. Apoussidou, "The learnability of metrical phonology," Ph.D. dissertation, University of Amsterdam, 2007.

[32] P. Boersma and J.-W. Van Leussen, "Efficient evaluation and learning in multilevel parallel constraint grammars," *Linguistic Inquiry*, vol. 48, no. 3, pp. 349–388, 2017.