Computational Modeling of Nonfinality Effects on Stress Typology

Robert Staubs

1. Introduction

Attested linguistic patterns are not all equally frequent. This fact poses a difficulty: models of grammar cannot predict relative frequency differences on their own—something else is required. In this paper, I discuss an explicit model of one of those additional components: bias in perception and production. I show that a very simple model of how (mis)perception interferes with learning can predict large-scale trends in typology. In particular, I offer the model as an explanation of how proposed perceptual biases disfavoring final stress could explain several directional tendencies in stress typology.

This model has the advantage of being nearly "end-to-end." That is, given a perceptual bias, grammatical representation, and learning mechanism, we can examine predicted typological results. This allows the explicit testing of assumptions and proposed explanations.

2. Nonfinality

For the purposes of this paper, a nonfinality bias is any force in perception or production that deters final stress. A reason to suspect that such a pressure is active in these channels is that nonfinality as a concept has been robustly useful in understanding the grammatical side of stress. This includes notions such as extrametricality (Liberman & Prince, 1977), as well as more explicit nonfinality (Prince & Smolensky, 1993/2004) in constraint-based formalisms—both serve as grammatical ways to exclude final syllables from bearing stress or other metrical structure.

It is less clear why such a pressure should exist in the first place, although some potential answers have been proposed. As one possibility, the avoidance of final stress might be derived from an avoidance of clash between levels of tonal prominence and a preference for final higher-level prominences (Hyman, 1985; Gordon, 2000). As another, the intrinsic prominence due to final lengthening might disrupt the perception of stress in final position (Lunden, 2006). In this study, my goal is not to give new answers about why nonfinality pressures should arise. Instead, I show the kinds of typological implications can be derived from even very simple assumptions about these kinds of pressures.

Despite the apparent utility of nonfinality for grammatical analysis, it is not immediately clear whether final syllable stress is actually underattested crosslinguistically. This question can be addressed, at least in part, by a statistical view of the legal patterns in a stress typology. Here I take Heinz's (2007) Stress Pattern Database as my basis for frequency inference. Each stress pattern in this database has some frequency and some set of corresponding legal strings. We can use these measures to derive information about which properties of stress strings are markedly common or uncommon—either rare within individual patterns, or members of only rare patterns. Properties of interest should show up as significantly more or less attested than just creating strings at random.

Here I look only at patterns of one or two syllables anchored at the left or right word edge. In results here, a 0 indicates an unstressed syllable, a 1 indicates a stressed syllable. Sampling follows a simple procedure. First, a stress pattern is chosen according to its frequency from the Stress Pattern Database. Second, a word length is selected at random between one and eight syllables. Third, a word shape (i.e. a combination of syllable weights) is chosen uniformly at random. Finally, the above combinations of

---

[*] Robert Staubs, University of Massachusetts Amherst, rstaubs@linguist.umass.edu. This material is based upon work supported by the National Science Foundation under Grant No. S121000000211. Thanks to audiences at mfm 22, RUMMIT IV (April 2014), the 2014 Annual Meeting on Phonology, LSA 2015, and WCCFL 33 for comments on related talks. Thanks also to John Alderete, Luca Iacoponi, Karen Jesney, John Kingston, Anya Lunden, John McCarthy, Joe Pater, and others for valuable comments and discussion.

© 2016 Robert Staubs. Proceedings of the 33rd West Coast Conference on Formal Linguistics, ed. Kyeong-min Kim et al., 371-380. Somerville, MA: Cascadilla Proceedings Project.

stresses, unstressed syllables, and word edges occurring in the resulting string are tallied. These tallies can produce estimates of the mean values of each feature as well as estimates of confidence or variability.

Random selection of word length introduces one complication: what should the sampling distribution over word lengths look like? I present results from two assumptions. In the first, word length is sampled uniformly. This is not characteristic of real word length distributions, but might correspond to a learner focusing on "difficult" data in long words. In the second, short words are sampled exponentially (2^n) more often. This more closely echoes real-world distributions.

Figures 1 and 2 show the mean proportion of these tallies relative to chance, where chance is naively assumed to be 0.5 for single-syllable sequences and 0.25 for two-syllable sequences. The bars indicate 95% of the sampling distribution. Bars are marked with a significance star if 95% of the distribution does not include the chance (zero) line. These sequences reliably occur more (or less) often than would be expected by uniformly random concatenation of stressed and unstressed syllables. Table 1 lists patterns that deviate "significantly" from chance in the same direction under both sampling assumptions. These results can be considered reasonably robust to the exact distribution.

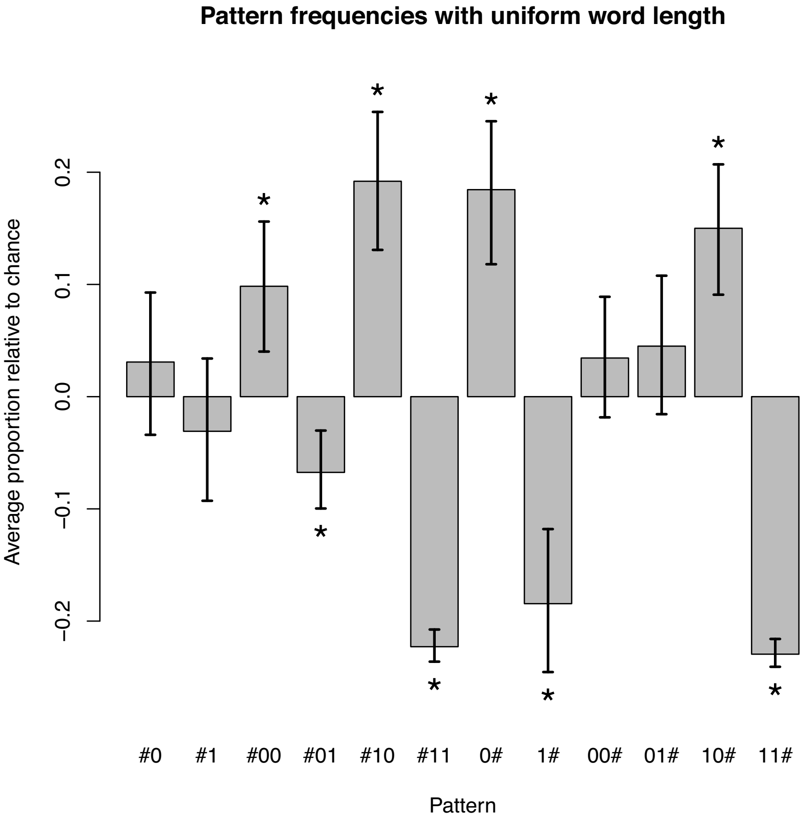

Figure 1: Deviation from chance for sequences sampled from the Stress Pattern Database (see text). Uniform sampling distribution.

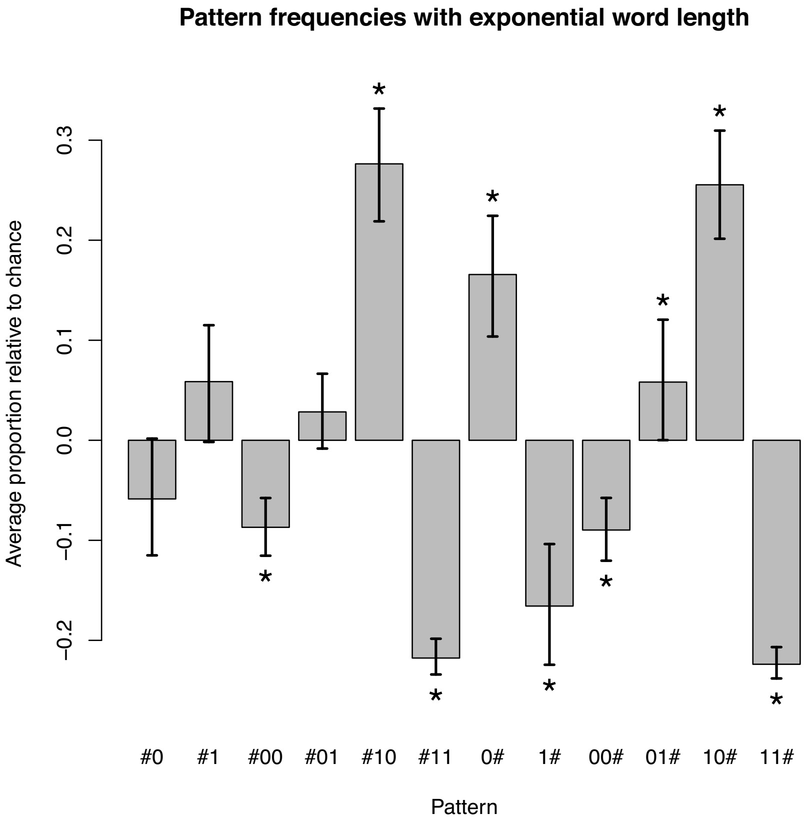

Figure 2: Deviation from chance for sequences sampled from the Stress Pattern Database (see text). Exponential distribution.

Pattern	Description	Tendency relative to chance
#10	Initial fall	favored
#11	Initial clash	disfavored
0#	Stressless final syllable	favored
1#	Stressed final syllable	disfavored
10#	Final fall	favored
11#	Final clash	disfavored

Table 1: Summary of preferences for one- and two-syllable sequences at word edges. Non-significant deviations from chance or deviations in opposite directions are not listed.

Although this kind of technique is useful for more than the present purposes, the most important tendencies here are for the sequences 0# and 1#. Under both sampling assumptions, 0# appears frequently and 1# appears infrequently. This is exactly what would be expected under a systematic pressure for nonfinality. It should be noted that most of these tendencies overlap—e.g. a general preference for falls and against clashes would provide a similar pattern of results. Notably, however, no single-syllable sequences at the left edge appear in this table—perhaps making preferences such as initial gridmark less appealing explanations (Hyde, 2008).

This statistical work serves to show that there is something worth modeling in nonfinality. The frequency of stress strings patterns in accordance with something like a nonfinality pressure, even if the precise details are unclear. In the next section, I turn to details of my model for nonfinality biases on typology.

3. Model

3.1. Grammar and learning framework

The grammatical model used here is Maximum Entropy grammar (MaxEnt; Goldwater & Johnson, 2003). In MaxEnt, a grammar consists of a weighted set of constraints. The grammar assigns harmonies to candidates as their weighted sum of violations—that is, it is a Harmonic Grammar (Legendre et al., 1990; Pater, 2009). MaxEnt further assigns probabilities to candidates proportional to the exponential of these harmonies. These explicit probability calculations are the chief virtue of MaxEnt for this work—they can record gradient confidence of the model in a particular hypothesis. This "confidence" affects production and perception but can also be examined by the analyst to determine how much learning a particular simulation has done—how much probability the model has moved onto the target forms.

In learning, a teacher produces a form according to a given categorical grammar—it samples from some stress pattern. The learner then probabilistically produces its own parse for that chosen word shape. If there is a mismatch between the output of the teacher and the output of the learner, the learner updates its constraint weights by the scaled difference between the violations of the two outputs (sampling SGA or perceptron; Jäger, 2007; Boersma & Pater, 2014). The set of constraints includes some which mention feet. Foot structure is not overt in the input to the learner, and thus must be inferred—here, by Robust Interpretive Parsing (Tesar & Smolensky, 2000; Boersma, 2003; Jarosz, 2013; Boersma & Pater, 2014).

The constraints used are a sample of those found in the stress literature (e.g. Alber, 2005; Kager, 2005). The goal here is not to predict much about frequency with the constraint set itself—the set must merely be capable of representing the relevant patterns. Instead, the constraint set is meant to be relatively unbiased on the question of nonfinality. To that end, two types of nonfinality and two types of noninitiality constraint are included. Thus the constraints themselves cannot bias a learner toward one edge rather than the other—if a nonfinality bias results from the constraints, noninitiality must as well.

- ALIGNHEADLEFT/RIGHT
- ALIGNFOOTLEFT/RIGHT
- "WEIGHTTOSTRESS"
- FOOTBINARITY
- Rhythmic: *CLASH, *LAPSE
- Foot headedness: IAMB and TROCHEE
- NONFINALITY and NONINITIALITY for foot and syllable

Some additional assumptions are relevant, though I have considered alternatives to each in additional simulations. Only words of length two to eight are allowed. Short words are sampled exponentially more often (2^n). For each word length, the candidates are all the valid parses for that length, with the exception of single stress systems. For these, only candidates with exactly one stress and at most one "heavy" syllable are considered. The distribution of "heavies" is assumed to be uniform where relevant. The learning rate for the update rule is 0.1 throughout. Weights start at a constant value of 10.

3.2. Incorporating nonfinality

The above model is necessarily unbiased with respect to direction—right is equivalent to left as far as the constraints are concerned. We must introduce some mechanism that biases the model towards nonfinality, modeling some view of how a nonfinality bias in production or perception might work. One such strategy is to assume that the final syllable of a word is probabilistically deficient. That is, when the teacher intends a form with final stress, this stress is occasionally lost on the way to the learner's grammatical update. This might be due to the teacher failing to adequately produce the stress, the learner failing to perceive it, or some other property of the noisy channel between the two grammars.

Crucially, this failure only happens on occasion. If it happened for every production and every teacher, no grammars violating nonfinality would be learnable. As this is clearly not a true fact about typology, the probability of transmission failure must be less than 1.0. In the results below, I consider a range of values for this parameter probability.

There is a residual issue with this type of strategy. Namely, if the final stress of a word is "lost," where does it go? In a foot-based constraint set, stress is necessary for any useful learning. A number of possible resolutions are possible, but here I will focus on only one. See Staubs (2014) for consideration of alternatives with similar results. If final stress is misperceived, it is instead perceived on the penult. This strategy is simple and seemingly plausible if part of the mechanism of nonfinality bias is something like peak delay.

4. Case studies

In this section, I demonstrate this model as applied to three aspects of stress typology: fixed stress position, alignment of stress windows, and foot headedness. In each case, I show that a nonfinality effect in the model can serve to break the left/right symmetry in a way aligned with typological observation.

4.1. Case study 1: Fixed stress

In a fixed stress system, stress always falls on a syllable a fixed number of syllables from the word edge. Typically, the primary stress is the focus in such descriptions. In the extreme case, fixed single stress systems have exactly one stress per word, always in a fixed location. In either focus, fixed stress typology seems to favor the right edge. Initial and final single stress (no separation from the word edge) do not show much asymmetry, but a distinction becomes apparent moving further into the word. In particular, penultimate stress is more frequent than peninitial (one syllable from the edge) and antepenultimate stress is more frequent than postpeninitial (two syllables from the edge). These patterns are shown in Table 2, with data from WALS and SPD.[1]

	SPD		WALS	
	#	%	#	%
initial	69	30.9	92	32.6
peninitial	12	5.4	16	5.7
postpeninitial	0	0.0	1	0.4
final	74	33.2	51	18.1
penultimate	60	26.9	110	39.0
antepenultimate	8	3.6	12	4.3

Table 2: Comparison of typologies of edge stress in Heinz (2007) and Goedemans & van der Hulst (2013).

These asymmetries are not explicable without some mechanism that makes left distinct from right for frequency predictions. A nonfinality bias provides this. Penultimate stress in particular clearly benefits from the perceptual error model given above: of all patterns that align with the right edge, penultimate aligns the most without relying on perceptually weak final stress data. Further, the model outlined above explicitly creates penultimate stress data out of final stress.

A bias in favor of antepenultimate stress is less obvious in this model, but does result from it. The key insight is that long initial stress patterns can actually be *final* in short words. In particular, postpenitial stress is final in words up to length three. It is therefore disadvantaged when compared with antepenultimate stress, which does not have this property.

The graph in Figure 3 shows patterns of error for six types of fixed stress. These results show how far a learner's grammar is from its teacher's after 10,000 pieces of learning data if it receives data from a teacher who "speaks" language X. Thus, higher numbers show more divergence from the teacher—these

---

[1] WALS counts are focused on fixed stress in general, while SPD counts are for fixed single stress. This (partly) accounts for an asymmetry between the two with respect to final stress.

languages are “harder” to learn. Here we see that as the probability of misperception increases, error on initial patterns also increases (as does final stress). This seems to offer an explanation for observed trends.

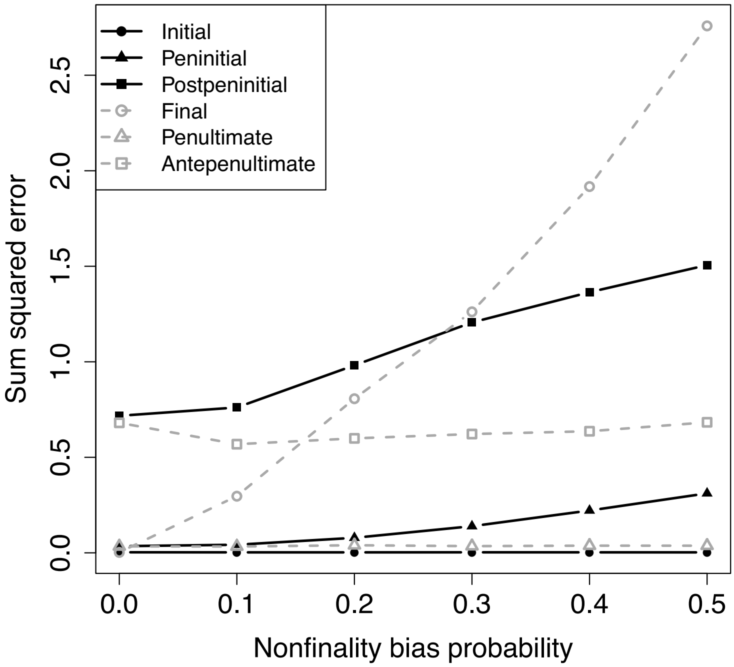

Figure 3: Error after 10,000 iterations for fixed stress patterns with a range of nonfinality parameter values.

However, error alone can lead to faulty frequency predictions (Rafferty et al., 2011)—we must model language transmission across generations, as in iterated learning (e.g. Kirby, 2002; Griffiths & Kalish, 2005). In the iterated learning model used here, we start a chain off with language X and see how likely it is to move to language Y in one generation. We can then exponentiate the resulting probabilities to extend results further into the “future.”

Learner 1 $\rightarrow$ Learner 2 $\rightarrow$ Learner 3 $\rightarrow$ Learner 4 $\rightarrow$ ...

Iterated learning results are shown in Figure 4. Here, predicted frequency of penultimate stress is shown to increase, both as time goes on and as the probability of misperception goes up.

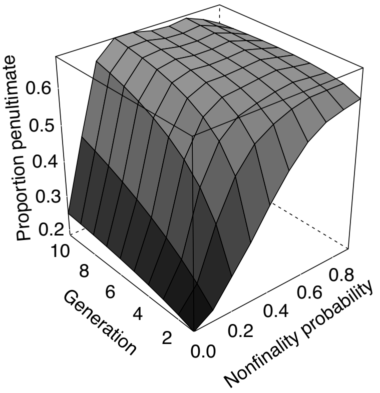

Figure 4: Iterated learning results for frequency of fixed stress systems.

4.2. *Case study 2: Foot headedness*

A dramatic directional asymmetry has been noted in favor of trochaic languages over iambic (e.g. by Hayes, 1995). That is, patterns compatible with left-headed feet are more common than those compatible with right-headed feet. This difference is stark: 153 versus 31 in WALS (Goedemans & van der Hulst, 2013).

Perceptual nonfinality offers a possible explanation. Iambs are inherently "final"—the smallest words showing iambs must have final stress. Many of the conceivable right-to-left systems will also show final stress in long words, as would left-to-right systems with degenerate feet. Trochees are different. Short words can easily have non-final stress, and longer words can have final stress *only* if degenerates are tolerated. A bias for nonfinality will thus decrease the frequency of iambs asymmetrically. It is especially damning that iambs are final in short words, likely the most common inputs. The frequency of some abstract systems are shown in Table 3, including how often they show final stress.

Foot type	Direction	Degenerate feet?	Final?	Count
Trochees	Left-to-Right	no	never	33
		yes	sometimes	22
	Right-to-Left	no	never	34
		yes	never	4
Iambs	Left-to-Right	no	sometimes	13
		yes	always	3
	Right-to-Left	no	always	2
		yes	always	3

Table 3: Parametric iterative stress in StressTyp.

Error patterns are shown in Figure 5. As expected, all iambic patterns become harder to learn with increased probability of final stress misperception. One trochaic pattern also becomes harder—left-to-right trochees with degenerate feet. Notably, this is not one of the most common trochaic types. An iterated learning simulation corroborates these results, showing that iambic frequency increases over generations and with increasing probability of misperception.

4.3. *Case study 3: Stress windows*

The final case study concerns a directional asymmetry in stress windows. In stress window systems, stress varies by a designated property (e.g. weight, sonority, lexical, etc.)—but only up to a given distance from the word edge. Windows can be at least two or three syllables long, but right edge systems are more common in both cases (Table 4). This situation is very similar to that of fixed stress: two-syllable windows are more common on the right edge than the left, as are three-syllable windows.

Error results are shown in Figure 7. A number of initial windows become more difficult with a nonfinality bias, but so do some final windows. Error is not enough for final-default right-edge windows. In particular, stress windows in which stress defaults to the last syllable are problematic.

There is reason to be hopeful for a larger model. Learning an initial window consists of learning both that alignment is leftward and that stress is sensitive to a designated property. A nonfinality effect disrupts this process: overall alignment is still leftward, but in short words where stress "should" be final, it instead appears on the penult. This type of misperception can produce data which is neither fully-aligned nor sensitive to a diacritic, contradicting the overall picture of learning.

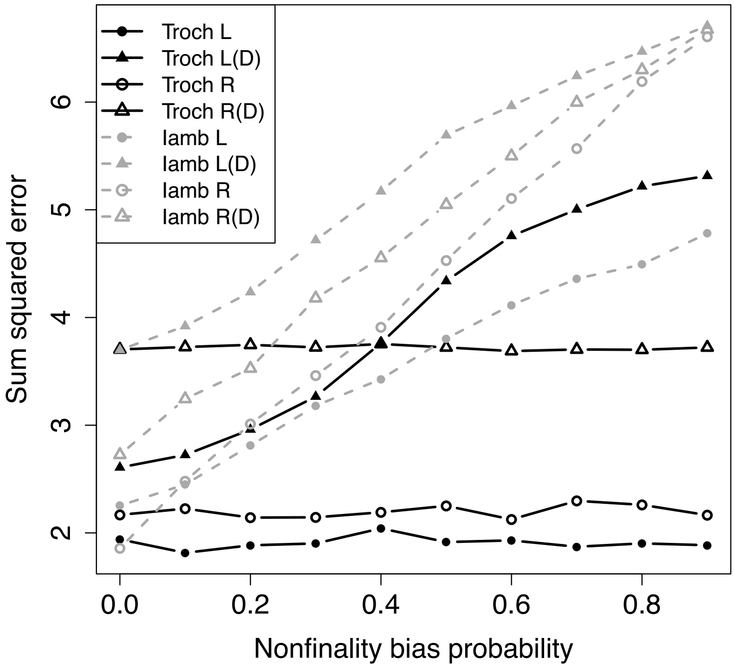

Figure 5: Error after 10,000 iterations for iterative feet patterns with a range of nonfinality parameter values. Key: foot type, left-to-right (L) or right-to-left (R), tolerance of degenerate feet (D).

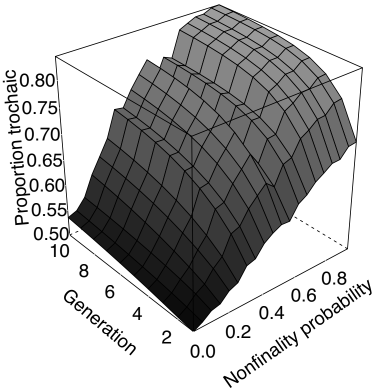

Figure 6: Iterated learning results for frequency of fixed stress systems.

Window type	#	
Final two syllables	82	e.g. Yapese (Jensen et al., 1977)
Final three syllables	38	e.g. Pirahã (Everett & Everett, 1984)
Initial two syllables	39	e.g. Malayalam (Asher & Kumari, 1997)
Initial three syllables	1	e.g. Comanche (Smalley, 1953)

Table 4: Frequency of window types, collapsing over defaults and designated properties (numbers from Kager, 2012).

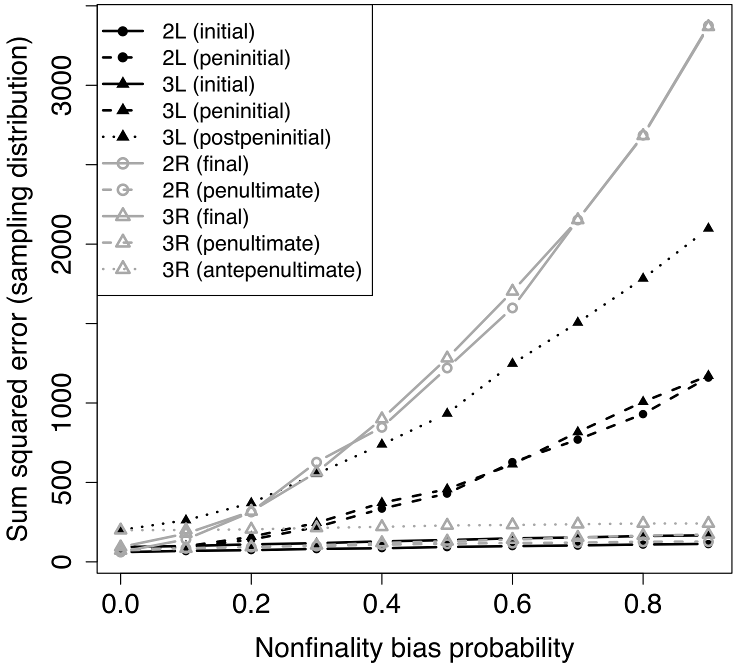

Figure 7: Error after 10,000 iterations for window stress patterns with a range of nonfinality parameter values. Key: window size, window on left (L) or right (R), position of default.

At this point, the learner must doubt something about its previous analysis. One option is to decrease the importance of left alignment. This potentially yields a large right-aligned stress window (the rightward solution). Alternatively, it could decrease the importance of designated property sensitivity (the quantity-insensitive solution). Both options are possible paths for this model, dictated by parameters of learning and random chance. Both the rightward and QI solutions are available, depending on model parameters. However, rightward option is disfavored—consistent alignment is easy to get right (see e.g. Staubs, 2014). Despite this, an overall typological bias moving windows right should result from a full model. A nonfinality bias can *occasionally* create a rightward solution, but it rarely results in a leftward movement. This could provide for an overall increase in the probability of final windows.

5. Discussion

In this paper, I showed that a simple model of a nonfinality bias in perception has useful typological consequences in an iterated learning model. This model predicts the relative frequency of penultimate and antepenultimate stress and the dominance of trochees over iambs. It also offers a tentative approach to the frequency of right-edge windows.

These results arise from a model in which most components (grammar, perception, learning, and diachrony) are explicitly modeled. This "end-to-end" approach is comparatively rare. Using these kinds of comprehensive models for inference about frequency allow us to better test theories of typological tendencies.

This model cannot give a full account of stress windows. Directions for future elaboration include several considerations, including different iterated learning models, different interpretations of the perceptual nonfinality bias, or different forms of formal bias. Stress typology in general can benefit from this type of model that integrates perception and learning, for example models of the perceptual basis of clash and lapse avoidance/tolerance or the correlation between direction and foot shape.

References

Alber, Birgit (2005). Clash, lapse and directionality. *Natural language & linguistic theory* 23:3, 485–542.

Asher, Ronald E. & T.C. Kumari (1997). *Malayalam*. Psychology Press.

Boersma, Paul (2003). Review of Tesar & Smolensky (2000): Learnability in Optimality Theory. *Phonology* 20, 436–446.

Boersma, Paul & Joe Pater (2014). Convergence properties of a gradual learner in Harmonic Grammar. McCarthy, John J. & Joe Pater (eds.), *Harmonic Grammar and Harmonic Serialism*, Equinox Press, London.

Everett, Dan & Keren Everett (1984). On the relevance of syllable onsets to stress placement. *Linguistic inquiry* 705–711.

Goedemans, Rob & Harry van der Hulst (2013). Fixed stress locations. Dryer, Matthew S. & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*, Max Planck Institute for Evolutionary Anthropology, Leipzig, URL http://wals.info/chapter/14.

Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Stockholm workshop on variation within Optimality Theory*, 111–120.

Gordon, Matthew (2000). The tonal basis of weight criteria in final position. *Regional Meetings, Chicago Linguistic Society*, vol. 36, 141–156.

Griffiths, Thomas L. & Michael L. Kalish (2005). A Bayesian view of language evolution by iterated learning. *Proceedings of the annual conference of the cognitive science society*, vol. 27, 827–832.

Hayes, Bruce (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.

Heinz, Jeffrey (2007). *Inductive learning of phonotactic patterns*. Ph.D. thesis, University of California, Los Angeles.

Hyde, Brett (2008). The rhythmic foundations of initial gridmark and nonfinality. *Proceedings of the North East Linguistics Society*, vol. 38.

Hyman, Larry M. (1985). *A theory of phonological weight*. Foris Publications Dordrecht.

Jäger, Gerhard (2007). Maximum entropy models and stochastic Optimality Theory. *Architectures, rules, and preferences: a festschrift for Joan Bresnan* 467–479.

Jarosz, Gaja (2013). Learning with hidden structure in Optimality Theory and Harmonic Grammar: Beyond Robust Interpretive Parsing. *Phonology* 30, 27–71.

Jensen, John Thayer, Leo David Pugram, John Baptist Iou & Raphael Defeg (1977). *Yapese reference grammar*. University Press of Hawaii Honolulu.

Kager, René (2005). Rhythmic licensing theory: an extended typology. *Proceedings of the third international conference on phonology*, Seoul National University, 5–31.

Kager, René (2012). Stress in windows: Language typology and factorial typology. *Lingua* .

Kirby, Simon (2002). Learning, bottlenecks and the evolution of recursive syntax. *Linguistic evolution through language acquisition: Formal and computational models* 173–203.

Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky (1990). *Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness*. University of Colorado, Boulder, Department of Computer Science.

Liberman, Mark & Alan Prince (1977). On stress and linguistic rhythm. *Linguistic inquiry* 8:2, 249–336.

Lunden, Stephanie Laura (2006). *Weight, final lengthening and stress: A phonetic and phonological case study of Norwegian*. Ph.D. thesis.

Pater, Joe (2009). Weighted constraints in generative linguistics. *Cognitive Science* 33:6, 999–1035.

Prince, Alan & Paul Smolensky (1993/2004). *Optimality Theory: Constraint interaction in generative grammar*. Blackwell, Malden, MA and Oxford, UK.

Rafferty, Anna N., Thomas L. Griffiths & Marc Ettlinger (2011). Exploring the relationship between learnability and linguistic universals. *Association for Computational Linguistics: Human Language Technologies 2011* .

Smalley, William A. (1953). Phonemic rhythm in Comanche. *International journal of American linguistics* 19:4, 297–301.

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

Tesar, Bruce & Paul Smolensky (2000). *Learnability in Optimality Theory*. The MIT Press.

Proceedings of the 33rd West Coast Conference on Formal Linguistics

edited by
Kyeong-min Kim, Pocholo Umbal,
Trevor Block, Queenie Chan,
Tanie Cheng, Kelli Finney, Mara Katz,
Sophie Nickel-Thompson, and Lisa Shorten

Cascadilla Proceedings Project Somerville, MA 2016

Copyright information

ISBN 978-1-57473-469-0 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Staubs, Robert. 2016. Computational Modeling of Nonfinality Effects on Stress Typology. In *Proceedings of the 33rd West Coast Conference on Formal Linguistics*, ed. Kyeong-min Kim et al., 371-380. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #3256.