Learning and the Position of Primary Stress

Robert Staubs
University of Massachusetts Amherst

1. Introduction

Patterns of primary stress are not completely independent of secondary stress patterns. For a given secondary stress pattern, particular choices of primary stress are more dominant typologically. In a rule-based metrical framework the generalization can be stated as follows: primary stress tends to fall on the first foot placed in a metrical parse. Thus for a left-to-right parse the leftmost stress is primary and for a right-to-left it is the rightmost.

In this paper I present an explanation of this typological bias in favor of "same edge" primary stress: languages following this generalization are more accurately learned and in iterated learning come to outnumber languages violating the generalization. I give an explicit online model of stress learning based in Maximum Entropy grammar and a gradual learning algorithm.

A learning-based model of typology provides a satisfactory explanation for numerical tendencies. These tendencies are inherently non-categorical—they permit exceptions. A probabilistic, learning-based model allows us to capture generalizations while permitting such exceptions. I conclude by showing that such a model depends crucially on grammatical assumptions. Thus probabilistic typological models offer distinct sets of data on which grammatical hypotheses can be evaluated. I use primary stress to demonstrate the ability of such explanations to account for typological tendencies in general, a task not well addressed by most formal models.

2. Primary stress directionality

A striking tendency is found in the relationship between primary stress and directionality. Primary stress tends to fall on the "first" foot placed in a metrical parse. Iterative stress systems may be categorized in part by the edge at which iteration begins. The starting edge gives a place from which the system seems to count. In (1), the two patterns are easily explained as alternating stressed and unstressed syllables (trochees) starting at the left or right edge, respectively. An attempt using the opposite edge yields a comparatively clumsy description: stress is penultimate or antepenultimate depending on word parity (and iterative thereafter).

(1)

Left	Right
$(\grave{\sigma}\sigma)\sigma$	$\sigma(\grave{\sigma}\sigma)$
$(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)$	$(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)$
$(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)\sigma$	$\sigma(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)$
$(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)$	$(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)$
$(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)\sigma$	$\sigma(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)$

* This work benefited greatly from Jeff Heinz's stress database and Max Bane's processing of it. Thanks to Eric Baković, Ryan Bennett, John McCarthy, Joe Pater, and Anne-Michelle Tessier for useful insight and the audiences of the NECPhon IV, the 2nd UConn Workshop on Stress and Accent, and WCCFL 31 for helpful comments and questions. This material is based upon work supported by the National Science Foundation under Grant No. BCS-0813829 to the University of Massachusetts Amherst and DGE-0907995 as a Graduate Research Fellowship to the author.

 Proceedings of the 31st West Coast Conference on Formal Linguistics, ed. Robert E. Santana-LaBarge, 428-437. Somerville, MA: Cascadilla Proceedings Project.

In the great majority of cases these systems place primary stress on the "first" foot of a parse. That is, a system which counts from the left will place primary stress on the leftmost foot (*vice versa* for the right). However, some exceptions exist, placing primary stress on the "last" foot parsed.

(2)

First foot	Last foot
$(\acute{\sigma}\sigma)$	$(\acute{\sigma}\sigma)$
$(\acute{\sigma}\sigma)\sigma$	$(\acute{\sigma}\sigma)\sigma$
$(\acute{\sigma}\sigma)(\grave{\sigma}\sigma)$	$(\grave{\sigma}\sigma)(\acute{\sigma}\sigma)$
$(\acute{\sigma}\sigma)(\grave{\sigma}\sigma)\sigma$	$(\grave{\sigma}\sigma)(\acute{\sigma}\sigma)\sigma$
$(\acute{\sigma}\sigma)(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)$	$(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)(\acute{\sigma}\sigma)$
$(\acute{\sigma}\sigma)(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)\sigma$	$(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)(\acute{\sigma}\sigma)\sigma$

Parsing direction is very closely related to the position of primary stress, as seen in the StressTyp (Goedemans, 2010) database of stress patterns.[1]

(3)

	Left-to-right	Right-to-left
Left primary	63	12
Right primary	27	53

$\chi^2 = 38.1, p < 0.05$

The correlation is striking: primary stress predominantly falls on the foot at the start of iteration. Most of the exceptions are bidirectional stress systems in which a foot is placed *opposite* the start of iteration. The other feet iterate toward this single "opposite-edge" foot. Given other simple systems, one might expect that primary stress generally falls on the foot at the start of iteration. Instead, it is generally this opposite-edge stress which is primary.

Van der Hulst (1996) proposes the "Primary First" theory of stress assignment. Under this approach, primary stress is *always* the first stress assigned. Under such a theory, the primary stress of simple iterative systems must be at the start of iteration and bidirectional systems must place primary stress on the opposite edge—there are no other options. This theory accounts for the majority languages in (3), bidirectional systems, and systems where primary stress is opposite the start of iteration but fixed with respect to an edge. It does not easily generate languages where primary stress placement depends crucially on a secondary stress parse—on word parity. Such languages are very rare but examples exist—notably Cairene Arabic (McCarthy, 1979) and Nyawaygi (Dixon, 1983). Van der Hulst refers to these patterns as "counting systems." Only 15 of 158 iterative stress languages included in StressTyp are listed as having stress of this kind.[2]

The primary first *tendency*, in theory-neutral terms, states that primary stress placement is usually independent of the position of secondary stress. That is, in most languages primary stress can be placed without any reference to the placement of secondary stress—primary stress tends to be expressible in terms of a single privileged syllable (e.g. penultimate, initial, final heavy, etc.), rather than a particular privileged stress (e.g. the rightmost stress). In a derivational theory the tendency means that primary stress tends to fall on the first foot placed. An alignment-based translation is more nuanced because left- and right-alignment are not the same as left-to-right and right-to-left parsing (Alber, 2005). For example, a language with degenerate feet at the left edge may be better *aligned* with that edge but "parsed" right-to-left. This means that the generalization cannot always be precisely restated as for example "left-/right-aligning secondary stress systems tend to have left-/right-aligning primary stress" (but see Gordon, 2002 on related generalizations). The independence of primary stress is thus central to the tendency—this is the sense in which I describe a primary "first" tendency in this paper.

Van der Hulst's Primary First theory accounts for a wide range of the typological data on stress. The majority of stress systems do in fact place primary stress on their first foot. However, as the above counts and examples demonstrate, this tendency is not an absolute—Kager (1995), for example, points out this issue with Primary First but does not address an alternate approach. Despite the strength of the tendency, it is not categorical. However, it would be a mistake to leave such a strong generalization without an explanation. Instead, a full theory of primary stress and directionality must diverge from both approaches, giving an account of bias without excluding exceptions.

[1] Table modified from Goedemans' (2010) Table III. Yates' χ^2 added.

[2] The number drops to 11 when distinct varieties of Arabic are collapsed.

3. Learning biases in stress

Languages obeying the primary first tendency have an advantage in learning: stress data is more self-consistent. A single datum can provide evidence for many different properties of a hypothesized stress grammar. Data within a language can vary in the extent to which these properties are consistently specified, with important consequences for learning. If such properties are consistent, different data provide evidence for the same set of hypotheses. This means that learners effectively get more evidence—greater certainty in their hypotheses—out of consistent data compared with relatively inconsistent data. This is the situation that obtains with the primary first tendency: in tendency-obeying languages, primary stress is placed consistently—it does not vary with secondary stress. In these cases primary stress is independent of secondary stress placement, meaning that it will be placed consistently across word sizes. (2) shows this consistency (and lack thereof)—if primary stress is first, its placement does not vary with word parity. This consistency results in faster learning in a model of learning that can represent it.

Biases in learning can lead to biases in typology. One way in which this could happen is iterated learning (e.g. Kirby et al., 2007; Griffiths & Kalish, 2007). In this perspective learners impose their learning bias on the data they receive from teachers and then pass this biased result on to future generations. Successive generations receive biased data and bias it still further, eventually yielding languages which can be very dissimilar from the starting patterns. The languages which are likely to survive or be innovated in such a model are those which are better learned, accounting for greater typological frequency for such patterns. Moreton & Pater (2012) adapt such a model to a constraint-based framework to model feature economy effects in typology. I adopt a similar model. Another partial precedent for this work is found in Bane & Riggle (2008), who identify a link between the number of rankings or weightings describing a stress language and its relative attestation. This work follows Coetzee (2002) in linking numbers of rankings to predicted frequencies.

3.1. Grammatical model

The model I adopt for learning simulations uses Maximum Entropy grammar (MaxEnt; Goldwater & Johnson, 2003). MaxEnt is a probabilistic version of Harmonic Grammar (Legendre et al., 1990; Pater, 2009). A MaxEnt grammar consists of a set of weighted constraints, bearing similarity to ranked Optimality Theoretic grammars. MaxEnt differs from OT in that it induces a probability distribution over output candidates. More precisely, the probability of a candidate is proportional to the exponential of its harmony—the grammar-weighted sum of its violations.

$$(4) \quad p(x) \propto e^{H_x}$$

The probabilistic output of MaxEnt grammars is useful for modeling learning. In a categorical model learning can only succeed or fail, but if the result is a probability distribution the learner's degree of certainty in a hypothesis can be represented. A learning bias can be measured in the relative degree of certainty achieved in different learning problems. Such a bias can in turn directly affect successive learners in the way that it shapes the distribution of data given as input.

In the main simulations presented here I use constraints chosen to represent standard sorts of distinctions in stress grammars, building off the constraint sets of Alber (2005) and Kager (2005).

1. Stress alignment: ALIGNFTLEFT/RIGHT: Assign a violation for every syllable between the left/right edge of a foot and the edge of the prosodic word.

2. Primary stress alignment: ALIGNHEADLEFT/RIGHT: Assign a violation for every syllable between the left/right edge of the head foot and the edge of the prosodic word.

3. Foot size: FTBIN: Assign a violation for every monosyllabic foot.

4. Rhythmic: *CLASH/*LAPSE: Assign a violation for every pair of adjacent stressed/unstressed syllables.

5. Foot headedness: IAMB and TROCHEE: Assign a violation for every foot that is not strictly right/left headed.

3.2. Modeling learning bias

The learning model used here is online—the learner processes each datum it receives in turn, adapting its hypothesis. A teacher randomly selects a word length and produces a stress pattern for that word length based on its grammar. Shorter words are sampled exponentially more often than longer words, mirroring the distribution of word lengths in natural language. The learner considers a candidate set consisting of all metrical parses including the correct number of syllables with one and only one primary stress. The learner produces its own parse for that word length. If the learner's predicted stress pattern does not match the teacher's, the learner updates its constraint weights.

The learner's grammar is updated according to the HG-GLA (Boersma & Pater, In prep.), also known variously as the perceptron update rule or delta rule. In this update rule the old weights are adjusted by the difference between the violations of the learner's chosen candidate and the violations of the teacher's chosen candidate, scaled by a learning rate η.

(5) New Weights = Old Weights + $\eta \times$ (Learner Violations − Teacher Violations)

This update increases the weight of constraints violated more in the learner's erroneous form, penalizing such violations more heavily. It decreases the weight on constraints violated more in the teacher's chosen form, permitting such violations more. If there is no difference in violations for a particular constraint, no change is made.

Some of the constraints used in the simulations involve foot structure, which is not overtly observable. The learner must therefore make a decision about what hidden structure to use in evaluating the teacher's constraint violations. The learner presented here uses a probabilistic adaptation of Robust Interpretive Parsing (Tesar & Smolensky, 2000) to choose a likely hidden structure. In this version of RIP the hidden structure used for a particular overt form is probabilistically chosen according to the grammar from all hidden structures consistent with the form.

3.3. Results

A bias in learning for a particular feature, e.g. primary first stress patterns, can be illuminated by comparing the learning of languages possessing and lacking that feature. For the purposes of the simulations here, I start with the quantity-insensitive stress patterns used by Bane & Riggle (2008), supplementing Heinz (2007). All of these languages are primary first, so additional languages must be considered which would violate the generalization. In addition to these 26 stress patterns, I include the "flipped" version of 17 of them. These "flipped" patterns have identical secondary stress to some attested language but primary stress aligned toward the opposite edge. For example, (2) shows a language and its "flip"—both languages have trochees parsed left-to-right but they differ in whether the leftmost or rightmost stress is primary. The remaining 9 languages are not generally included because they do not "flip" non-vacuously (e.g. final stress).

This inclusion of flipping allows the comparison of languages differing only on their adherence to the primary first tendency. Results are included in Figure 1. In this graph the residual error after some amount of learning is compared between a language and its flipped counterpart. Significant deviation from the line indicates bias—one language or the other is better learned in the time allowed. A primary first bias is apparent here: the iterative languages are primarily above the line, indicating better learning in the unflipped, generalization-conforming pattern. I will return to the two exceptions in §3.4. These results are presented numerically in (6). In this table I present the difference between unflipped and flipped pattern errors. Negative values indicate primary first bias.

		Mean Diff.	Diff. Range
(6)	Same-edge primary	-0.664	-1.429 to 0.932
	Bidirectional	-1.145	-1.982 to -0.514

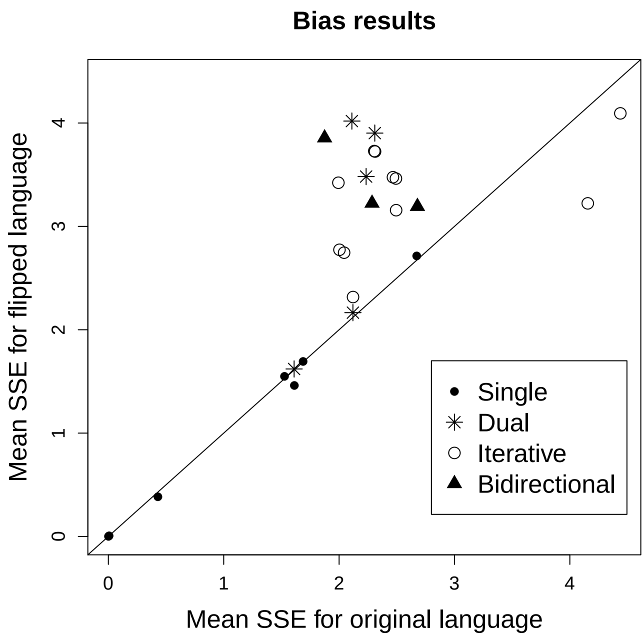

Figure 1: Learning bias in favor of generalization-conforming languages. Points above the line show bias in favor of the language in the typology; points below show bias in favor of the "flipped" language. Single stress languages have one stress per word, dual have at most two, iterative & bidirectional have stresses in proportion to word length. Single stress languages are included to estimate noise. $\eta = 0.1$, 100 trials, 1,000 iterations each.

Another way to look at learning predictions is to examine the actual result of an attempted learning instance rather than error alone. That is, we may ask what languages are likely to be learned when given certain kinds of input. In the simplest case, all languages are learned faithfully: given data from a particular language, a learner would only produce a grammar exactly consistent with its teacher's grammar. The degree to which individual languages differ from this ideal is their error rate, discussed above. We may also look at what languages are likely to be produced from particular sources other than the original language. Rafferty et al. (2011) show that in iterated learning models of typology this is a necessary step: simple error is not sufficient to describe typological trends.

Figure 2 shows results of this kind. This figure shows the probability of acquiring a particular language given a particular language as the data source, giving this probability as shading. Learning is carried out with a teacher for each of the languages in the augmented typology described above. The resulting grammar is then compared with the languages of the typology. The language which receives the maximum likelihood under the learned grammar is counted as the resulting language. The figure presents a confusion matrix of these results: the probability that a particular initial language yields a learned grammar that best describes some particular (potentially different) language. The rows represent particular starting languages, the columns particular resulting language, and the darkness of the square the frequency of a result. The diagonal of this matrix represents faithful learning. The boxed subsection represents the "flipped" languages, expected to be learned worse on average. The axis values are indices representing the 43 languages considered—their order is unimportant.

It is clear that these flipped languages are not completely impossible to learn in a single generation. This is as expected given the gradient nature of error in Figure 1. To further probe predictions of typology, iterated learning can be modeled. In such a model a learner's result grammar is used to generate data for

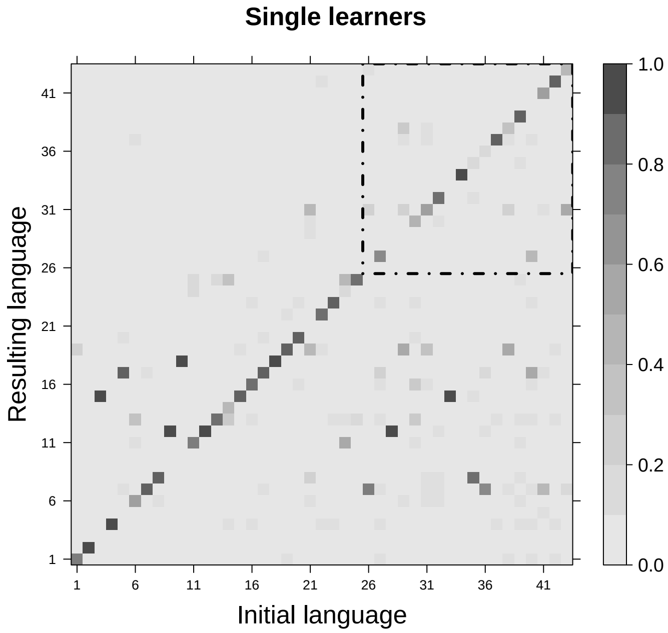

Figure 2: Single step learning results. Probability of ending at some language after starting at some (possibly different) language. "Flipped" languages are boxed. 500 trials with 10,000 iterations per trial.

a second generation learner, which in turn generates data for a third, and so on. Figure 3 shows results for this kind of learning model. A concentration of probability onto fewer languages is apparent, with a move particularly away from languages which contradict the primary first tendency. Graphically this presents as less faithful learning (on the diagonal) and some movement away from the boxed tendency-disobeying languages.

A final way of looking at this sort of bias systematically is to calculate the theoretical results of iterated learning based on a single instance of learning. To do this, we take the matrix in Figure 2 as representing the probability that a learner categorically learns a language based on some initial language. We can then calculate the probability of future generations acquiring each language by exponentiating this matrix. This method has the advantage of easily calculating long-term predictions with the tradeoff assumption being the categorical learning assumption. With this tool in hand, we can calculate the probability of arriving at an iterative language obeying the primary first tendency compared with one disobeying it. Figure 4 shows these theoretical probabilities over many generations assuming a uniform initial distribution over languages.[3] Despite an equal number of flipped and unflipped iterative stress languages, the probability distributed over the unflipped languages is greater.

3.4. Explaining the bias

In the above sections I have shown that the primary first bias emerges in learning stress languages. The origin of this bias can be better understood through consideration of violation vectors, the lists of constraint violations characterizing each candidate. A classic result by Novikoff (1962) gives a convergence guarantee for perceptron learning when applied to linearly-separable classes. This result shows a link between the speed of perceptron learning and properties of the vectors considered for classification. Learning speed decreases when the norm (size) of the largest vector under consideration increases. Speed *increases* when the margin between vectors in one class and those in another class increases.

[3] Results are qualitatively similar with other starting distributions such as restricting the starting languages to iterative stress and/or using the frequencies of the unflipped languages from Heinz (2007).

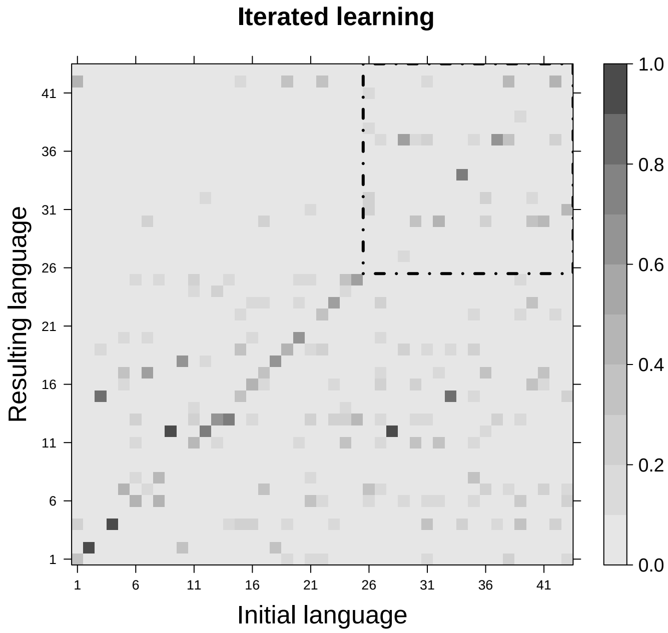

Figure 3: Simulated iterated learning results. Probability of ending at some language after starting at some (possibly different) language estimated over 100 trials. "Flipped" languages are boxed. $\eta = 0.1$, 10 generations with 10,000 iterations each.

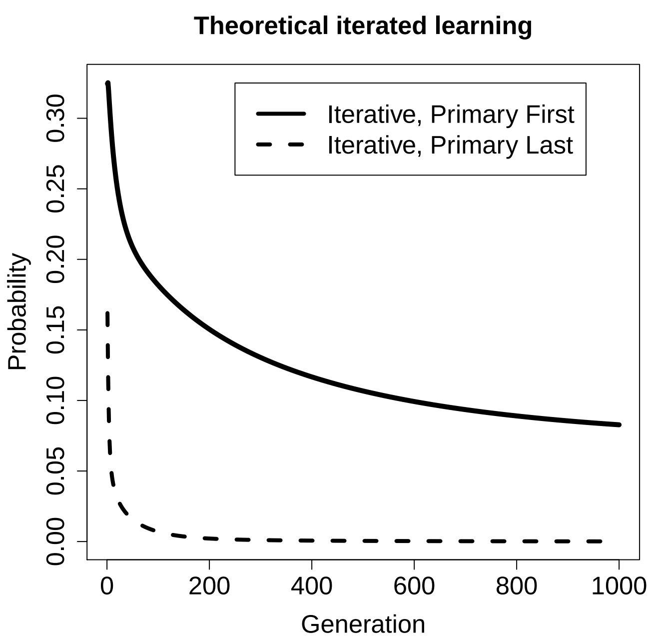

Figure 4: Theoretical bias of iterated learning. Probability distributed over all tendency-conforming iterative languages compared with all tendency disobeying ones. Calculated from learning results of Figure 2.

These two considerations apply simply to gradual learning of constraint-based grammars. The size of the largest vector considered corresponds roughly to the most-violating candidate that must be considered in learning. That is, the wider the range of candidates that need to be ruled out or accepted, the slower that learning proceeds. The margin between vectors corresponds to the distinctiveness of the candidates which are part of a language from those which are excluded from it. Thus the speed of learning increases when candidates in a language are all similar to one another (in terms of violations vectors) in ways in which they are dissimilar to candidates out of the language.

The important criterion for the primary first tendency is distinctiveness. The candidates in a tendency-obeying language are generally distinct from the candidates outside of it. The reason is that primary first languages are more *consistent* across word length in their placement of primary stress. The first foot in a directional parse will be aligned closely with a word edge in a similar or identical way across different word lengths. If primary stress falls on this foot, constraints referring to primary stress will be consistently violated (or unviolated) in a way distinct from candidates outside the language. The strings within a tendency-obeying language are more consistent than those outside of it: they are distinctive. This distinctiveness produces the learning bias presented here. The primary first bias in learning is thus closely related to Gordon's (2002) observation that stress patterns placing stress a uniform distance from the word edge are preferred.

The relationship of this idea of consistency or distinctiveness can be made more clear by considering two cases: bidirectional stress and primary stress clash. These two patterns are presented schematically in (7). Bidirectional stress is the chief reason for proposing that primary stress should actually *precede* secondary stress The most common sort of bidirectional stress language is like the one presented—primary stress falls on the "stranded" foot towards which secondary stress iterates. This aligns with learning results in Figure 1—bidirectional stress is better learned in its more attested form. This follows from consistency of primary stress placement: such bidirectional systems have extremely uniform primary stress, while their "flipped" patterns can vary in primary stress placement.

(7)

Bidirectional	Clash
$(\acute{\sigma}\sigma)$	$(\acute{\sigma}\sigma)$
$(\acute{\sigma}\sigma)\sigma$	$(\acute{\sigma})(\grave{\sigma}\sigma)$
$(\acute{\sigma}\sigma)(\grave{\sigma}\sigma)$	$(\acute{\sigma}\sigma)(\grave{\sigma}\sigma)$
$(\acute{\sigma}\sigma)\sigma(\grave{\sigma}\sigma)$	$(\acute{\sigma})(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)$
$(\acute{\sigma}\sigma)(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)$	$(\acute{\sigma}\sigma)(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)$
$(\acute{\sigma}\sigma)\sigma(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)$	$(\acute{\sigma})(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)(\grave{\sigma}\sigma)$

The other pattern here shows the type of language in which a single instance of learning predicts a reversal of the primary first tendency—the two languages noted in §3.3. One such language is schematized in (7). This language could be described as right-to-left trochees tolerating degenerate feet, but primary stress is on the "last" (leftmost) foot. This language—and its reversal, left-to-right iambs with degenerate feet and primary stress on the rightmost foot—are better-learned than their unflipped counterparts obeying the primary first generalization. These patterns have clash between a primary stress and a secondary stress. These languages are uncommon, as noted by e.g. Kager (2001). They are perceptually problematic—the two stresses must be correctly perceived as two stresses of different types. However, they are very consistent: primary stress always falls in exactly the same position with respect to the word edge. These languages are very rare but attested, for example South Conchucos Quechua (Hintz, 2006). Here perceptual and formal biases are in tension: the existence of such perceptually dispreferred languages may be due to their advantages in learning.

3.5. Types of bias

The nature of this sort of learning bias implies that not all constraint sets will behave similarly. To take advantage of consistent primary stress placement, the constraint set must be able to represent consistency of this type. This is true of many constraint sets which can align primary stresses based on syllable count—in addition to the alignment constraints presented here, n-gram constraints over syllables referring to levels of stress perform well. Some constraint sets do not make reference to syllables in connection to primary stress: for example, Gordon's (2002) constraint set for quantity insensitive stress. In Gordon's analysis, primary stress is placed determined by an end rule constraint (Prince, 1983) over secondary stresses. Primary first languages are identically consistent to primary last languages in such a constraint set—both have primary stress fully aligned within the secondary stresses.

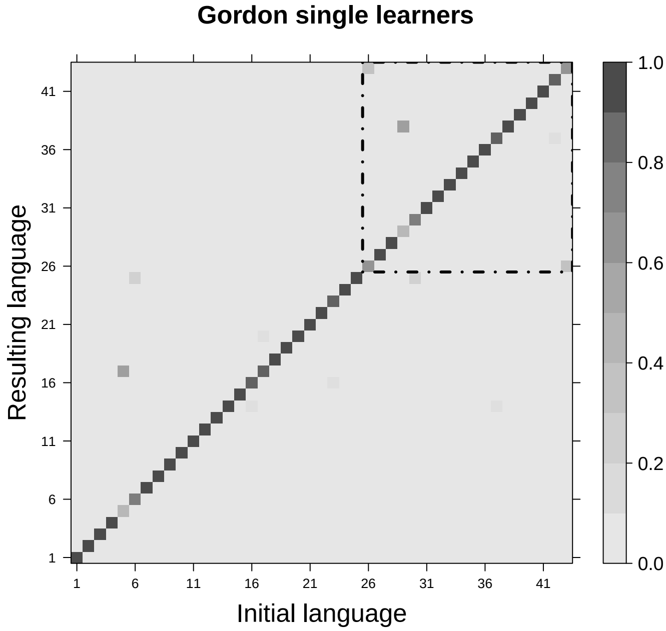

Figure 5: Single step learning results for Gordon (2002) constraint set. Probability of ending at some language after starting at another. “Flipped” languages are boxed. $\eta = 0.1$, 2,500 trials with 10,000 iterations per trial.

This model of stress learning makes other probabilistic predictions not connected to the particular choice of constraints. These tendencies arise not from properties of violation vectors but instead from the length of strings needed to learn particular patterns. For example, this model predicts that single stress systems should be more common than multiple stress systems, that binary systems should be more common than ternary ones, and that single stress systems should place stress close to the word edge. These properties of the model bear out observations on the size of strings needed for learning stress systems.

4. Conclusion

In this paper I presented an account of a probabilistic tendency in stress. Primary stress tends to be “first” in that primary stress placement is generally independent of secondary stress. I show that this sort of tendency is expected to arise from learning biases preferring consistent primary stress. I presented an explicit model of this bias couched in terms of an online learning model. This account formalizes notions of consistency (Gordon, 2002) and provides a natural link between formal properties of stress patterns Bane & Riggle (2008) and typology.

This account of a tendency produces probabilistic results. The tendency is accounted for but it is not rendered categorical. This advantage of a combined grammar and learning theory over a grammar-only theory follows from the probabilistic nature of the learning process and generalizes to other learning models and other phenomena of interest. Predictions about tendencies are projected from grammatical assumptions in combination with a learning model: no extra grammatical formalism is necessarily required. In the results presented here, a tendency for primary first stress follows from the types of grammatical hypotheses already supported on categorical grounds. By adding learning to the standard general account of stress typology, we derive the intended generalization without specifying anything additional about consistency in the grammar. Grammatical hypotheses differ in the probabilistic predictions they make because learning-based methods offer an additional way to test existing proposals as well as to guide the creation of new theories.

References

Alber, Birgit (2005). Clash, lapse and directionality. *Natural Language & Linguistic Theory* 23:3, 485–542.

Bane, Max & Jason Riggle (2008). Three correlates of the typological frequency of quantity-insensitive stress systems. *Proceedings of the Tenth Meeting of ACL Special Interest Group on Computational Morphology and Phonology*, Association for Computational Linguistics, 29–38.

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

Coetzee, Andries W. (2002). Between-language frequency effects in phonological theory.

Dixon, Robert M.W. (1983). Nyawaygi. Dixon, Robert M.W. & Barry J. Blake (eds.), *Handbook of Australian Languages*, Amsterdam: John Benjamins.

Goedemans, Rob (2010). A typology of stress patterns. van der Hulst, Harry, Rob Goedemans & Ellen van Zanten (eds.), *A survey of word accentual systems in the language of the world*, Mouton de Gruyter, Berlin, 647–666.

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 (2002). A factorial typology of quantity-insensitive stress. *Natural Language & Linguistic Theory* 20:3, 491–552.

Griffiths, Thomas L. & Michael L. Kalish (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science* 31:3, 441–480.

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

Hintz, Diane M. (2006). Stress in South Conchucos Quechua: a phonetic and phonological study. *International Journal of American linguistics* 72:4, 477–521.

van der Hulst, Harry (1996). Separating primary and secondary accent. Goedemans, Rob, Harry van der Hulst & Ellis Visch (eds.), *Stress patterns of the world*, Holland Academic Graphics, The Hague.

Kager, René (1995). The metrical theory of stress. Goldsmith, John A. (ed.), *A handbook of phonological theory*, Basil Blackwell, Oxford.

Kager, René (2001). Rhythmic directionality by positional licensing.

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

Kirby, Simon, Mike Dowman & Thomas L. Griffiths (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences* 104:12, 5241–5245.

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.

McCarthy, John J (1979). On stress and syllabification. *Linguistic Inquiry* 10, 443–465.

Moreton, Elliott & Joe Pater (2012). Structure and substance in artificial-phonology learning. *Language and Linguistics Compass* 6:11, 686–701.

Novikoff, Albert B.J. (1962). On convergence proofs for perceptrons. *Proceedings of the Symposium on the Mathematical Theory of Automata*, vol. 12.

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

Prince, Alan S (1983). Relating to the grid. *Linguistic inquiry* 19–100.

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* .

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

Proceedings of the 31st West Coast Conference on Formal Linguistics

edited by Robert E. Santana-LaBarge

Cascadilla Proceedings Project   Somerville, MA   2014

Copyright information

ISBN 978-1-57473-462-1 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