Learning biases and primary stress orientation

Robert Staubs

University of Massachusetts Amherst

rstaubs@linguist.umass.edu

Introduction

The placement of primary stress is correlated across languages with the direction of footing—the “first” foot placed in parsing a word tends to be the one which bears primary stress.

Languages which show “opposite-edge” primary stress do exist. In StressTyp, 8% of languages are of this sort (Goedemans & van der Hulst 2005).

This is a strong tendency but not an absolute restriction. I seek to explain this.

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

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

Proposal: This typological generalization emerges from learning.

- Cues for primary and secondary stress can reinforce one another in learning.
- This reinforcement is strongest when primary stress correlates with the direction of footing.
- Such languages can be learned faster and come to dominate the typology.

Grammar

Maximum Entropy grammar (MaxEnt; Goldwater & Johnson 2003, Hayes & Wilson 2008):

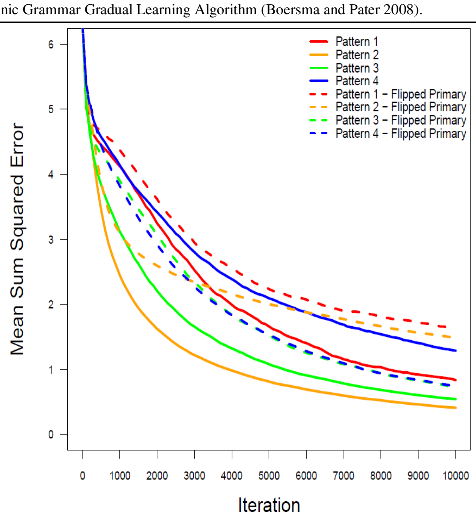
- Set of weighted constraints used to compute harmony scores for candidates.
- The probability of output candidates is proportional to the exponential of their harmony.

Such a model gives explicit values for the learner’s correctness at any given time.

Learning

On each iteration of learning the following occurs:

1. The learner receives a form sampled from an exponential distribution on word length.
2. The learner produces a form for that word shape according to its grammar.
3. If the learner’s form differs from the target, the learner updates its weights as in the Harmonic Grammar Gradual Learning Algorithm (Boersma and Pater 2008).



***n*-Gram Constraint Learning Trajectories**

Dotted lines indicate “opposite-side” patterns.

n-gram Constraints

The constraints match strings of 1 to 4 unstressed syllables, secondary stresses, primary stresses, and word edges.

Constraints which match secondary stress also match primary stress (but not vice versa).

Measuring learnability differences:

- Compare mean sum squared errors of 100 trials at the $10{,}000^{\text{th}}$ iteration (learning rate = 0.01).
- Subtract measure for a “flipped” pattern (with primary stress on the opposite edge) from the pattern’s measure.
- Negative if there is a bias for the pattern, positive if for the reverse, and zero if unbiased.

Schematic results:

Pattern	Strings	Learnability Difference
1. from edge	102020 1020202	-1.067
2. from edge, lapse	102020 1020200	-0.824
3. from one off edge	010202 0102020	-0.177
4. from one off edge, clash	010202 0102022	0.575

All but one pattern is learned faster with same-edge primary stress.

The exception would have a clash at the primary stress if flipped—likely marked on other grounds (Kager 2001).

Larger typology:

(Heinz 2007—only the quantity-insensitive languages):

	Mean Learning Diff.	Learning Diff. Range
Single stress	0.023	0.003 to 0.039
Same-edge primary	-0.349	-2.079 to 0.966
Opposite-edge primary	-0.664	-1.081 to -0.246

- 2 attested languages have primary stress opposite the direction of footing and get negative values.
  - These are bidirectional trochaic systems—just those we expect to have “opposite-edge” primary stress.
  - As desired, the system predicts primary stress should fall on the “first” foot—*not* the start of iteration.
- Some same-edge systems get positive learning values—they would be easier to learn if flipped.
  - These 4 languages would all have clash at the primary stress if this were done, however.
  - The remaining majority of languages in the typology are easier-learned as they are.

Mutually-Reinforcing Cues

Constraints on secondary stress double as constraints on primary stress.

Learning how a secondary stress constraint is weighted means (potentially) also learning about primary stress.

Patterns which are learned quickly are in some way consistent across word lengths for both primary and secondary stress: they might always have some initial stress, or penultimate, or so on.

If a consistent stress is primary, the learner’s evidence all points in the same direction: all word parities and both types of stress support the same location.

If primary stress falls on a more variable location the evidence is weaker and learning is slower.

Alignment Constraints

Constraints:

1. ALIGNFTLEFT/RIGHT: Assign a violation for every syllable between the left/right edge of a foot and the edge of the prosodic word.
2. 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. FTBIN: Assign a violation for every monosyllabic foot.
4. *CLASH: Assign a violation for every pair of adjacent stressed syllables.
5. *LAPSE: Assign a violation for every pair of adjacent unstressed syllables.
6. IAMB: Assign a violation for every trochaic foot.
7. TROCHEE: Assign a violation for every iambic foot.

Learning feet:

Foot structure is not overtly observable.

The learner uses Robust Interpretive Parsing (Tesar & Smolensky 1998): the target form used for learning is the learner’s best-scoring parse of the received overt form.

Results on the Heinz (2007) typology:

	Mean Learning Diff.	Learning Diff. Range
Same-edge primary	-0.349	-2.079 to 1.058
Opposite-edge primary (= bidirectional)	-2.372	-2.433 to -2.311

- Most same-edge patterns are negative.
- 4 of the exceptions are the same as before: languages which would have primary stress clash.
- A remaining 2 have near-zero differences. These languages are perfect grids—left-to-right trochees and right-to-left iambs, both with degenerate feet possible.

Generality:

Not all constraint sets work.

This set without IAMB and TROCHEE is less successful. Without these constraints the learner lacks some ways of expressing consistency of stress across word shapes (and stress type).

Conclusion

This approach to explaining the primary stress/footing correlation seems potentially appealing.

The bias produced is definite but not total, mirroring the incomplete nature of the generalization.

The learning explanation used is explicit and not designed specifically for the task.

This manner of projecting biases might be useful in comparing and evaluating different proposed constraint sets.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. BCS-0813829 to UMass Amherst and a Graduate Research Fellowship to the author.

Thanks are due especially to Jeff Heinz for his database and Max Bane for his processing of it.

Thanks also to John J. McCarthy, Claire Moore-Cantwell, Joe Pater, and audiences at NECPhon V and UMass Amherst.