Typological gaps in iambic nonfinality and learning

Joe Pater and Brandon Prickett
University of Massachusetts Amherst

In this paper, we discuss gaps in stress typology that are unexpected from the perspective of a foot-based theory, and show that the patterns pose learning difficulties for gradient-based learning algorithms. The unattested patterns result from combining theoretical elements whose effects are generally well-attested: iambic footing, nonfinality, word edge alignment and a foot binarity requirement. One of these patterns is particularly descriptively simple: stress is final in disyllables that begin with a light syllable (and in monosyllables), and penultimate otherwise. The fact that it, and a set of related patterns, pose great difficulty to a class of learning algorithms suggests that their cross-linguistic absence may be due to learning factors.

These typological gaps have not received much prior attention, perhaps because they are so puzzling. The most explicit discussion appears to be that of Kager (2006: 24), who notes that the unattested patterns are produced by a standard set of Gradient Alignment constraints (GA; McCarthy and Prince 1993), but not by Rhythmic Licensing Theory (Kager 2001). Winner-loser pairs for the pattern already mentioned appear in Table 1. Foot boundaries are indicated with parentheses. L is a light syllable, H is heavy and X is either. Primary stress is shown with bolded italics – there is no secondary stress in this pattern, due to Align(Ft, Wd)-R over Parse-Syl.

W *vs.* L	Ft-Bin	Nonfinality	Align(Ft, Wd)-R
a. (L***X***) *vs.* (***L***)X	W	L	L
b. (***H***)X *vs.* H(***X***)		W	L
c. (X***X***)X *vs.* (***X***)XX			W

The winners will be picked with the constraints having ranks or weights matching their order in the table. A light-initial disyllable gets final stress due to the dominance of FtBin (and Iamb), and all other word shapes get penultimate stress due to Nonfinality and Align(Ft, Wd)-R. This is exactly the analysis Prince and Smolensky (1993/2002) provide for Latin, except with iambs replacing the Latin trochees. This suspension of nonfinality in short words is not particular to OT: parametric extrametricality accounts also invoke mechanisms to accomplish it (e.g. Hayes 1995).

A second unattested iambic pattern in Kager (2006) is primary stress on the penultimate syllable, with initial secondary stress in five syllable words, and on the second syllable in words of six syllables or more. In this case, Nonfinality is unviolated, and Ft-Bin is inactive. Align(Wd, Ft)-L place the foot in initial position, with Iamb preferring the head to be on the second syllable when there is sufficient space left by a final binary iamb. Kager (2006) points out that the analogue of this dual stress pattern is attested with trochees in Georgian, but is absent in this iambic form.

These patterns can be found amongst the 124 target stress systems constructed by Tesar and Smolensky (2000) as a test of their approach to hidden structure learning. They do not flag the systems' typological unattestedness, nor does any of the subsequent literature that uses this test set (Boersma 2003, Jarosz 2013, Boersma and Pater 2016). We became aware of their crosslinguistic absence when we found that they were part of a set of systems that our own learner failed on, and looked to the theoretical-typological literature to see if they existed.

Our learner operates with a Maximum Entropy grammar (Goldwater and Johnson 2003), and uses a form of Expectation Maximization to deal with hidden structure, as in Pater et al. (2012) (see Jarosz 2006, 2013, 2015 for other applications of EM in phonology). Learning is done in batch (as opposed to on-line), and error is minimized using LBFGSB, with no regularization. The success rate is comparable to the best of the results reported in Jarosz (2013). When the learner is initialized with constraint weights set to 1, 115/124 of the languages are learned (93%). When the initial weights are sampled from a uniform distribution from 0-10, the learner is successful 91% of the time across 10 runs of each of the languages.

We identified 11 languages as "hard": the nine languages that initialization at 1 failed on, and two more that random initialization never succeeded on. Of those 11 languages, all of which seem unattested, 8 are cases in which a correct analysis must use iambic feet with nonfinality to place primary stress on the penultimate syllable. In these 8 languages a large proportion of the data is consistent with a trochaic analysis. The two patterns described above are extreme examples: in the first, only the light-initial disyllables fail to be parsed correctly in a trochaic analysis, and in the second, only words of 6 syllables or longer require iambic feet. The failed learners wound up in local minima assigning trochaic feet in the highest probability candidates.

These results appear to be general for learners that change constraint weights, or stochastic OT rankings values, by reference to a gradient. The first pattern described above, which was the hardest for our learner, was noted by Jarosz (2013: fn. 10) to be problematic for the on-line stochastic OT and Noisy HG learners that she examined, which used the gradient-based methods Boersma and Pater (2016) refer to as the OT-GLA or HG-GLA (see Jäger on the connection to Stochastic Gradient Ascent). We also found that it failed to be learned by Stochastic OT, Noisy HG, and MaxEnt learners over 10 runs each using the method described in Boersma and Pater (2016). The second language described above, which was the second hardest for our learner, failed to be learned in our 10 runs with Stochastic OT, though Noisy HG succeeded all 10 times, and on-line MaxEnt in 8 of them. The results do not seem to be general for all constraint-based learners: the EDL learner described in Jarosz (2015) found the first pattern in 10/10 runs.

These results support Boersma's (2003) speculation that failures on this test set might be useful in explaining typology, and extend prior demonstrations that learning can shape stress typology in Staubs (2014) and Stanton (2016) to cases of ambiguity caused by hidden structure. Further work is needed to show that learning will in fact have the desired effect on typology (e.g. through agent-based simulations), and it also remains possible that the true source of the gaps is in the grammatical theory: the systems are not produced by metrical theories that place stress without reference to feet (Prince 1983, Bailey 1995, Gordon 2002, Kager 2006).