Learning Stress with Feet and Grids

Seung Suk Lee, Cerys Hughes, Alessa Farinella, and Joe Pater
{seungsuklee, ceryshughes, afarinella, pater}@linguist.umass.edu
Annual Meeting on Phonology 2022

Assumption: are Feet a burden?

- Gordon (2011): "… adoption of a grid-based rather than a foot-based approach to stress avoids the **burden** of having to model the learning of the hidden structure implicit in a foot-based theory."
- What could "burden" mean?
  - **Foot-based approach may fail to learn some languages or take longer to learn languages**

Learning Simulations

- Data: 28/33 Quantity Insensitive languages (Gordon 2002, Hayes & Wilson 2008) that the foot-based approach can represent
- Question 1: Does the foot-based approach correctly converge on every language?
- Question 2: Is the foot-based or grid-based approach faster?
- Learner: MaxEnt learning with hidden structure (Pater & Prickett 2022)
  - Initial weight = 1, Gradient Descent (as opposed to L-BFGS-B in P&P 2022)
- Input/Output: 1-7 syllables in length, marked with $\sigma/\grave{\sigma}/\acute{\sigma}$
  For example,

$\acute{\sigma}\sigma$	Candidates	Constraints
Grid	$\acute{\sigma}\sigma, \sigma\acute{\sigma}, \grave{\sigma}\acute{\sigma}, \acute{\sigma}\grave{\sigma}$	Gordon (2002) (n=12)
Foot	$(\acute{\sigma}\sigma), (\sigma\acute{\sigma}), (\grave{\sigma})(\acute{\sigma}), (\acute{\sigma})(\grave{\sigma})$	M&P (1993), Kager (2005) (n=9)

  - Learning data and all aspects of learning for both constraint sets were identical
  - Only difference is that the foot-based learner had to assign prosodic structure to the learning data (see P&P 2022)
- Learning success: For each syllable-length, 0.90 prob on correct surface form
- Learning speed: The number of epochs it takes for the learner to succeed in learning
- Learning failure: no success until maximum number of epoch (1000)

Differences between Grid and Foot

- Foot introduces hidden structure, Grid doesn't

$\sigma\acute{\sigma}\sigma$		
Grid	Level 2	x
	Level 1	x
	Level 0	x x x
		σ σ σ
Foot	$\sigma(\acute{\sigma}\sigma)$	
	$(\sigma\acute{\sigma})\sigma$	

- Grid counts grid marks to the edge, Foot counts syllables to the edge

$\sigma\grave{\sigma}\sigma\acute{\sigma}\sigma\grave{\sigma}$			
Grid			Align (X_2, R)
	Level 2	x	*
	Level 1	x x x	
	Level 0	x x x x x x	
		σ σ σ σ σ σ	
Foot			AlignHeadRight
	$(\sigma\grave{\sigma})(\sigma\acute{\sigma})(\sigma\grave{\sigma})$		**

Grid vs. Foot

	Learned	Faster
Grid	28	11
Foot	28	**14**
Total	28	28 (3 equal)

- Foot correctly converges on all 28 languages it can represent
- Foot faster than or equal to Grid on 17/28 languages (61%)

Grid-main-revised

- One of the major differences between Grid and Foot is how main stress placement is penalized. So we adopted the syllable counting AlignHead constraint from the foot-based approach.

$\sigma\grave{\sigma}\sigma\acute{\sigma}\sigma\grave{\sigma}$		Grid-main-revised
Grid	Level 2 x Level 1 x x x Level 0 x x x x x x σ σ σ σ σ σ	Align (X_2, R) Syll **
Feet	$(\sigma\grave{\sigma})(\sigma\acute{\sigma})(\sigma\grave{\sigma})$	AlignHeadRight **

Grid-main-revised vs. Foot vs. Grid

- This change had a great consequence on learning
- Grid-main-revised faster than or equal to Grid on 25/28 languages (89%)

	Learned	Faster
Grid	28	3
Grid-main-revised	28	**19**
Total	28	28 (6 equal)

- Grid-main-revised faster than or equal to Foot on 19/28 languages (68%)

	Faster
Foot	9
Grid-main-revised	**15**
Total	28 (4 equal)

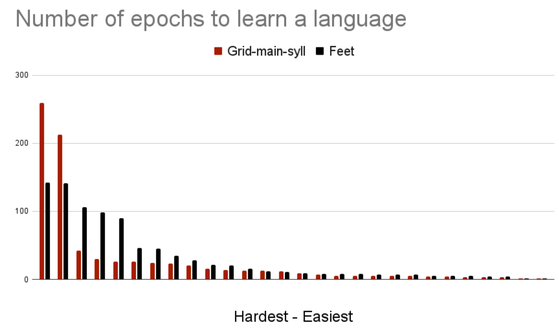
- Foot was faster than Grid or Grid-main-revised on 8 languages

References

Goedemans, R. W., Heinz, J., van der Hulst, H. (2015, April). StressTyp2, version 1. http://st2.ullet.net
Gordon, M. (2002). A factorial typology of quantity-insensitive stress. Natural Language Linguistic Theory, 20(3), 491-552.
Gordon, M. (2011). Stress systems. The handbook of phonological theory, 75.
Hayes, B., Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. Linguistic inquiry, 39(3), 379-440.
Kager, R. W. J. (2005). Rhythmic licensing theory: an extended typology. In Proceedings of the third international conference on phonology (pp. 5-31). Seoul National University.
McCarthy, J. J., Prince, A. (1993). Generalized alignment. In Yearbook of morphology 1993 (pp. 79-153). Springer, Dordrecht.
Pater, J., Prickett, B. (2022). Typological gaps in iambic nonfinality correlate with learning difficulty. In Proceedings of the Annual Meetings on Phonology (Vol. 9).

Learning speed (Number of epochs)

	Grid	Grid-main-revised	Foot
Mean	56.8	28.3	31.8
SD	87.6	59.9	42.5
Min	1	1	1
Max	401	259	142



- On average, Grid was slower than Foot,
- and Foot slower than Grid-main-revised
  - The hardest language for Grid-main-revised took longer than the hardest language for Foot

Conclusion

- Foot can correctly converge on all the languages it can represent
- Foot was **not slower** than Grid
- Grid learned faster with the **main stress** constraints that counts the number of **syllables**, rather than grid marks.
- Even with the revised constraints, Grid-main-revised was consistently slower on certain languages than Foot
- **The fact that Feet introduce hidden structure does not necessarily make learning harder for all languages in general.**
- **Details of the constraint set can have considerable effect on the speed of learning.**

Future work

- Will the current result hold if weights are initialized randomly?
- Quantity-sensitive languages in StressTyp2 (Goedemans et al. 2015)

Acknowledgement

This research was supported by NSF grant BCS-2140826 to UMass Amherst. We'd like to thank Brandon Prickett for providing the MaxEnt learner, all the people at the UMass Sound Workshop, as well as six anonymous AMP reviewers. All errors are ours.
The MaxEnt learner we used is publically available at https://github.com/blprickett/Hidden-Structure-MaxEnt