Representing and learning stress in a MaxEnt framework

Seung Suk Lee, Joe Pater, and Brandon Prickett, UMass Amherst

Take-aways: **A constraint can be sufficient for representing attested patterns, but not for learning, smaller isn't always better, new software (handout)**

General project

Comparison of grammatical theories in terms of success and efficiency of learning

Specific project

Comparison of theories of word stress placement with a MaxEnt learning framework

Stress theories

1. TS Original/OG Tesar and Smolenky's (2000) foot-based constraint set, based on Prince and Smolensky (1993/2004) and McCarthy and Prince (1993; Generalized Alignment), including these constraints:

FootNonfin: Assign a violation for every finally stressed foot: *(0 1), *(1)

Main-L/R: Assign a violation for every syllable between the L/R edge of the word and the L/R edge of the *foot* that has primary stress.

2. TS Revised/REV The TS constraint set with replacements:

Trochee: Assign a violation for every foot that is not initially stressed: *(0 1), ✓ (1)

Main-Syl- L/R: Assign a violation for every syllable between the L/R edge of the word and the L/R edge of the *syllable* that has primary stress.

And added **Nonfin-main** "no final main stress"

3. Gordon Original Gordon's (2002) grid-based constraint set, plus Weight-to-Stress

4. Gordon Revised Gordon (2002) + *W-to-S*, with revised edge oriented main stress constraints (syllable-counting instead of stress counting)

We focus here on the foot-based constraint sets, especially in learning

This research was supported by NSF grant BCS-2140826

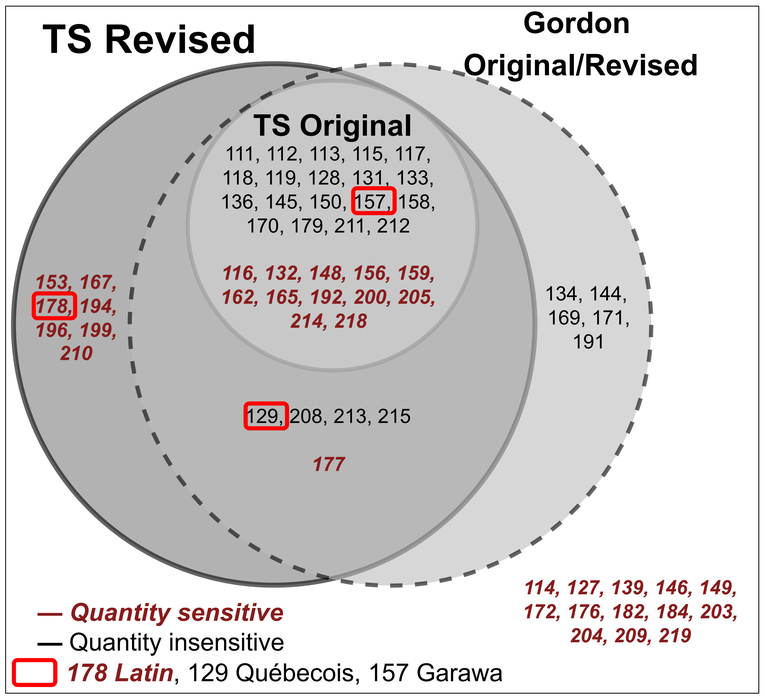
Target languages

61 languages from StressTyp 2 (Goedemans et al. 2015) that have FSTs and that can *in principle* be represented by these constraint sets (e.g. no ternary weight distinctions)

- FSTs used to generate stress patterns for all possible strings of heavy and light syllables of length 2-5, plus light syllables of length 6 and 7
- Candidates include all possible placements of main and secondary stress
- Violations assigned for each constraint set

Representational capacity

A language can't be learned if it can't be represented!



The original Tesar and Smolensky (TS) constraint set represents a proper subset of the others

- The revisions to the TS constraint set adds both QS (e.g. Latin) and QI languages (e.g. Québecois)
- The Gordon grid-based constraints add mostly QI languages, including ternary ones

Learning success

L-BFGS-B with initialization at 1 was successful for 42/43 = 98% of languages. The exception is Garawa (QI), which will be our focus below

Learning efficiency

Efficiency measured as the number of epochs using Gradient Descent with initialization at 1; comparisons are on the 24 languages that all theories successfully learned

- TS-OG quicker on quantity insensitive (QI)
- TS-REV quicker on quantity sensitive (QS)

	Constraint Set	Mean	Median	SD	Min	Max
QI	TS-OG	117.57	21.5	175.59	1	475
	TS-REV	199.86	32	305.86	1	803
QS	TS-OG	429.30	244	473.45	13	1273
	TS-REV	332.60	76.5	409.82	13	1018

Garawa

TS-OG (L-BFGS-B, initialization at 1)

FootNonfin	26.87
MainLeft	12.54
WordFoot-L	12.00
Parse	11.27
AllFeet-R	6.90
WordFoot-R	5.10
FtBin	4.93

Parses
(1 0)
(1 0) 0
(1 0) 0 (2 0)(2 0)

TS-REV (L-BFGS-B random initialization)

Nonfin	55.06
Parse	35.90
Main-Syl-L	30.70
Iambic	12.99
All-Feet-R	7.83
WordFoot-L	7.43
All-Feet-L	6.96
Nonfin-Main	5.16

Parses
(1) 0
(1 0) 0
(1 0) (02) (02) 0

Two "trochee" constraints? TS-REV + Ft-Nonfin (FNF)

	Constraint Set	L-BFGS-B (init 1)	L-BFGS-B (random init)	GD (init 1)
Garawa	TS-REV	Failure	2/10	Failure
	TS-REV + FNF	Success	10/10	163