23rd Manchester Phonology Meeting

May 28 2015

Learning as a window into the role of Faithfulness in stress systems

Aleksei Nazarov
University of Massachusetts Amherst

anazarov@linguist.umass.edu
blogs.umass.edu/anazarov

Overview

- For some languages, unclear whether stress is predominantly regular or predominantly lexical
- Computational model to predict regular vs. lexical stress: footing and role of Faithfulness learned in parallel
- Model tested on constructed languages: spectrum from fully regular to fully lexical stress

Regular vs. lexical stress

- regular (Markedness-governed) stress
  - Non-Finality >> Rightmost >> **Ident-Stress**
    /ˈpataka/ → [paˈtaka]
    /badaˈga/ → [baˈdaga]

- lexical (Faithfulness-governed) stress:
  - **Ident-Stress** >> Non-Finality, Rightmost
    /ˈpataka/ → [ˈpataka]
    /badaˈga/ → [badaˈga]

Regular vs. lexical stress

- For some languages: unclear whether some part of the stress system is regular or lexical
  - quantity-sensitive (QS) regular stress confusable with lexical stress

- E.g., West Germanic (English, German, Dutch; Domahs et al. 2014): debate on contribution of quantity-sensitivity vs. lexical stress

Regular vs. lexical stress

- West Germanic stress (Domahs et al. 2014 and work cited there):
  - Partially word-specific, partially predictable
  - Controversy: is there a QS default stress rule?

    (e.g. **English**: Kager 1989 vs. Booij & Rubach 1992;
    Dutch: van der Hulst 1984 vs. van Oostendorp 2012;
    German: Giegerich 1985 vs. Wiese 1996/2000)

Regular vs. lexical stress

- West Germanic stress (Domahs et al. 2014 and work cited there):
  - Partially word-specific, partially predictable
  - Controversy: is there a QS default stress rule?

Dutch

QS default			*QI default*	
/ˈkimono/	[ˈ**ki**.mo.no]		/ˈkimono/	[ˈ**ki**.mo.no]
/kasino/	[ka.ˈ**si**.no]	OR	/kasino/	[ka.ˈ**si**.no]
/vanitɑs/	[ˈ**va**.ni.*tas*]		/ˈvanitɑs/	[ˈ**va**.ni.*tas*]

Regular vs. lexical stress

- West Germanic stress (Domahs et al. 2014 and work cited there):
  - Partially word-specific, partially predictable
  - Controversy: is there a QS default stress rule?
    - No QS grammar will account for every word
    - But: tendencies towards quantity-sensitivity should be explained (e.g., Chomsky & Halle 1968)

Regular vs. lexical stress

- QS vs. lexical stress debate instantiates fundamental problem in phonology:
  - How much of the variance in the data comes from the grammar, how much from the lexicon? (Chomsky & Halle 1968, McCarthy 1981)
  - Similar to investigation by Albright & Hayes (2003) and others: to which extent does irregular past tense depend on rules vs. the lexicon?

Learning simulations

- One way of comparing alternative grammar hypotheses: learning simulations
  - Given Dutch data, will a learner represent quantity-sensitivity in the grammar?

- In this way, learning simulations are like an Evaluation Metric (Chomsky & Halle 1968, McCarthy 1981)

Learning simulations

- More direct way to test grammar hypotheses: behavioral experiments?

Learning simulations

- More direct way to test grammar hypotheses: behavioral experiments?
  - Domahs et al. (2014): speakers read nonce words with Heavy (closed) syllables in various positions
  - Asked to produce these words, stress pattern recorded
  - Result: weight of penult and final syllables is significant predictor of stress pattern for English, German, and Dutch speakers

Learning simulations

- However, behavioral experiments and learning simulations complement each other
  - Stochastically based (e.g., Maximum Entropy) grammar obtained from learning: more precise hypotheses and behavioral predictions than traditional analysis
  - Experiments provide more precise data for the learning model to match

(Hayes & Wilson 2008)

The model

- Standard hypothesis in OT:
  - Faithfulness competes with Markedness at the earliest stages (see also Hayes 2004, Prince & Tesar 2004)

- Thus: by default, Ident-Stress and phonological constraints on stress placement should compete in learning

The model

- Work on learning hidden foot structure (Tesar & Smolensky 2000, Pater & Boersma 2008, Jarosz 2013): no Ident-Stress constraint
  - Possibility of lexical stress introduces ambiguity:
  - For any observed stress pattern: does this pattern follow from general constraints on footing, or from Faithfulness?

The model

- I will show:
  - Competition between hidden (foot) structure and Faithfulness in learning allows exceptions to stress patterns to shape the grammar of stress
  - Type of exceptions matters for final product of learning

The model

- Model illustrated on the basis of constructed languages
  - Languages have QS stress pattern with various patterns of exceptions
  - Some languages learned as QS grammars, some as quantity-insensitive (QI) grammars

Languages

Constructed languages

- West Germanic data quite complex:
  - several types of irregularity which are patterns in themselves (e.g., see Nouveau 1994, van Oostendorp 1997 for Dutch)

- Therefore: simple constructed languages to narrow down the problem to the core
  - Given various kinds of exceptions to a QS pattern, will quantity-sensitivity still be in the grammar?

Constructed languages

- Series of 16 constructed languages
  - One language with Latin QS stress
  - One language with unpredictable stress
  - 14 languages with default Latin QS stress, but lexical exceptions present

Constructed languages

- The 8 words in each language:

paa.paa.paa	ta.paa.paa
paa.paa.ta	ta.paa.ta
paa.ta.paa	ta.ta.paa
paa.ta.ta	ta.ta.ta

All possible 3 syllable words with [ta], [paa]

Constructed languages

- The Latin stress language:

paa.(ˈpaa).paa	ta.(ˈpaa).paa
paa.(ˈpaa).ta	ta.(ˈpaa).ta
(ˈpaa).ta.paa	(ˈta.ta).paa
(ˈpaa).ta.ta	(ˈta.ta).ta

Penult stress iff penultimate syllable is heavy ([paa]); else antepenult stress.

Constructed languages

- The lexical stress language:

paa.(ˈpaa).paa	(ˈta.paa).paa
(ˈpaa).paa.ta	ta.(ˈpaa).ta
paa.(ˈta.paa)	(ˈta.ta).paa
(ˈpaa).ta.ta	ta.(ˈta.ta)

(No natural stress rule that applies here.)

Constructed languages

- Intermediate languages:
  - Introduce 1-3 exceptional forms into Latin stress language
  - 4 invariant forms across all languages

- These intermediate languages are of interest:
  - What will be the role of quantity-sensitive regular stress and lexical stress in the grammar learned for each of these languages?

Constructed languages

Invariant forms:

paa.ˈpaa.paa

ˈpaa.ta.ta

ta.ˈpaa.ta

ˈta.ta.paa

Variable forms:

		Latin 'paa.ta.paa		*Irregular* paa.'ta.paa	
		Latin 'ta.ta.ta	*Irregular* ta.'ta.ta	*Latin* 'ta.ta.ta	*Irregular* ta.'ta.ta
Latin paa.'paa.ta	*Latin* ta.'paa.paa	*Latin stress*	*1 exc.*	*1 exc.*	*2 exc.*
	Irregular 'ta.paa.paa	*1 exc.*	*2 exc.*	*2 exc.*	*3 exc.*
Irregular 'paa.paa.ta	*Latin* ta.'paa.paa	*1 exc.*	*2 exc.*	*2 exc.*	*3 exc.*
	Irregular 'ta.paa.paa	*2 exc.*	*3 exc.*	*3 exc.*	*Lexical stress*

Learning model

Learning constructed languages

- Grammar framework: Maximum Entropy (Goldwater & Johnson 2003, Hayes & Wilson 2008)
  - Constraints have weights rather than ranks
  - Every input yields statistical distribution over outputs (beneficial for learning)

- Learning: constraint weights optimized for all observed input → output mappings at once
  (software: Staubs' 2011 Solver)

Learning constructed languages

- Inputs are identical to winning outputs (cf. Tesar et al. 2003, Prince & Tesar 2004, Tesar 2006, Tesar 2013)
  - Latin stress language:
    correct form is [paaˈpaata] – input is /paaˈpaata/
  - Lexical stress language:
    correct form is [ˈpaapaata] – input is /ˈpaapaata/

- Markedness-over-Faithfulness bias (Smolensky 1996, Staubs 2011, Pater et al. 2012) keeps Faithfulness from taking over

Learning constructed languages

- For each input → output mapping I consider hidden structure (Tesar & Smolensky 2000 and subsequent work):
  - Every possible assignment of Heavy/Light to the output's syllables
  - Every possible footing of the output

  (see documentation for Staubs 2011; Pater et al. 2012 for math and implementation)

Learning constructed languages

- Constraints from Prince & Smolensky's (1993/2004) analysis of Latin stress:
  - Edgemost(L), Edgemost(R)
  - Foot-Binarity
  - Non-Finality
  - FootForm = Trochee, FootForm = Iamb
  - Weight-to-Stress Principle (WSP)

Learning constructed languages

- Additional constraints:
  - Ident-Stress

Learning constructed languages

- Additional constraints:
  - Ident-Stress
  - [paa]=Heavy: every syllable [paa] must be assigned Heavy
  - [ta]=Heavy: every syllable [ta] must be assigned Heavy

Learning constructed languages

- Additional constraints:
  - Ident-Stress
  - [paa]=Heavy: every syllable [paa] must be assigned Heavy
  - [ta]=Heavy: every syllable [ta] must be assigned Heavy
  - *Heavy : no syllable is assigned Heavy

Learning constructed languages

- Result of learning: Maximum Entropy grammar with a weight for every constraint

- Can be evaluated by “wug-testing” (Berko 1958)
  - How does the grammar respond to inputs not seen before?
    - What is the “intension” of the grammar?

Learning constructed languages

Constraint	Weight
Edgemost(L)	0.8319753
Edgemost(R)	25.7394494
Foot-Binarity	33.7328701
Non-Finality	40.1276016
FtForm = Trochee	36.8921376
FtForm = Iamb	0.0000000
Weigth-to-Stress Principle	8.0688030
Ident-Stress	0.0000000
[ta] = Heavy	16.5072300
[ti] = Heavy	0.0000000
*Heavy	33.1567912

Learning constructed languages

- Wug test for grammars:
  - Given input /abc/ not previously encountered by learner,
  - Compute output that wins in grammar learned for language X

- Wug-testing shows to which extent the patterns in the data are represented in the grammar (as opposed to the lexicon)

Learning constructed languages

- The wug inputs:

  1. underlying final stress (never observed in data)

  /paapaaˈpaa/ /paapaaˈta/ … /tataˈta/

  2. underlying stress that contradicts the words that have invariant stress across all languages

  /ˈpaapaapaa/ /paaˈtata/ /ˈtapaata/ /taˈtapaa/

Learning constructed languages

Invariant forms:

paa.ˈpaa.paa

ˈpaa.ta.ta

ta.ˈpaa.ta

ˈta.ta.paa

Variable forms:

		Latin ˈpaa.ta.paa		*Irregular* paa.ˈta.paa	
		Latin ˈta.ta.ta	*Irregular* ta.ˈta.ta	*Latin* ˈta.ta.ta	*Irregular* ta.ˈta.ta
Latin paa.ˈpaa.ta	*Latin* ta.ˈpaa.paa	*Latin stress*	*1 exc.*	*1 exc.*	*2 exc.*
	Irregular ˈta.paa.paa	*1 exc.*	*2 exc.*	*2 exc.*	*3 exc.*
Irregular ˈpaa.paa.ta	*Latin* ta.ˈpaa.paa	*1 exc.*	*2 exc.*	*2 exc.*	*3 exc.*
	Irregular ˈta.paa.paa	*2 exc.*	*3 exc.*	*3 exc.*	*Lexical stress*

Results

Results

- Latin stress language and lexical stress language (@ regularization prior variance = 100,000 & Mark>Faith = 0.035; range of other settings produces similar effects):
  - Latin stress language: wug-testing reveals no effects of Faithfulness
  - Lexical stress language: wug-testing reveals no effects of quantity-sensitivity

Results

- All intermediate languages preserve underlying stress, but only in non-final position (wugs with underlying final stress are made regular)
  - Some wug-tested as having default **QS** stress (with Faithfulness effects in non-final position)
  - Some wug-tested as having default **QI** stress (with Faithfulness effects in non-final position)

Results

Red:
QS language with Faith effects

Blue:
QI language with Faith effects

all constraints start out at w = 10

		Latin 'paa.ta.paa		*Irregular* paa.'ta.paa	
		Latin 'ta.ta.ta	*Irregular* ta.'ta.ta	*Latin* 'ta.ta.ta	*Irregular* ta.'ta.ta
Latin paa.'paa.ta	*Latin* ta.'paa.paa	*Latin stress*	*1 exc.*	*1 exc.*	*2 exc.*
	Irregular 'ta.paa.paa	*1 exc.*	*2 exc.*	*2 exc.*	*3 exc.*
Irregular 'paa.paa.ta	*Latin* ta.'paa.paa	*1 exc.*	*2 exc.*	*2 exc.*	*3 exc.*
	Irregular 'ta.paa.paa	*2 exc.*	*3 exc.*	*3 exc.*	*Lexical stress*

Results

- All languages with 1 exception learned as default Latin (QS) stress and Faith effects
- Languages with 2 and 3 exceptions learned variably as default QS or default QI stress, based on the particular pattern of exceptions

Results

Invariant forms:

paa.ˈpaa.paa

ˈpaa.ta.ta

ta.ˈpaa.ta

ˈta.ta.paa

Variable forms:

		Latin 'paa.ta.paa		*Irregular* paa.'ta.paa	
		Latin 'ta.ta.ta	*Irregular* ta.'ta.ta	*Latin* 'ta.ta.ta	*Irregular* ta.'ta.ta
Latin paa.'paa.ta	*Latin* ta.'paa.paa	*Latin stress*	*1 exc.*	*1 exc.*	*2 exc.*
	Irregular 'ta.paa.paa	*1 exc.*	*2 exc.*	*2 exc.*	*3 exc.*
Irregular 'paa.paa.ta	*Latin* ta.'paa.paa	*1 exc.*	*2 exc.*	*2 exc.*	*3 exc.*
	Irregular 'ta.paa.paa	*2 exc.*	*3 exc.*	*3 exc.*	*Lexical stress*

Results

- Number of exceptions not all-powerful
- However, more exceptions in input language leads to greater chance of quantity-sensitivity not being learned
  - 1 exception: 0/4 grammars are QI
  - 2 exceptions: 3/6 grammars are QI
  - 3 exceptions: 3/4 grammars are QI

Results

- Thus, my model predicts:
  - Sophisticated (quantity-sensitive) default pattern more difficult to discover as number of exceptions increases
  - But whether default pattern is part of grammar depends on the particular kind of exceptions that contradict it

Concluding remarks

Concluding remarks

- Problem: how to assess whether a stress pattern is regular or lexical when in doubt?
  - Example: West Germanic quantity-sensitivity – effect of Weight-to-Stress or Faithfulness?

- Modeled by constructed languages:
  - QS Latin stress with various amounts and kinds of exceptions – will quantity-sensitivity still be learned?

Concluding remarks

- Computational approach – complements experimental approach
- Proposal: competition of stress constraints and Faithfulness in learning model predicts regular vs. lexical stress
  - This competition follows from standard assumptions in OT and learning theory (Hayes 2004, Prince & Smolensky 2004)

Concluding remarks

- Results: type of exceptions predicts type of grammar learned
  - More exceptions means less chance of QS grammar
  - But whether QS grammar is picked depends on the particular kind of exceptions

Concluding remarks

- For future research:
  - Scale model up to more realistic data (e.g., some simplified form of West Germanic stress)
  - Make the connection to existing experimental results (e.g., Domahs et al. 2014)

- Another issue:
  - Strength of Markedness-over-Faithfulness bias: parameter without clear interpretation; seek possibility to eliminate it (Jarosz 2006, Wilson 2011)

Many thanks to:

John Kingston

John McCarthy

Claire Moore-Cantwell

Joe Pater

David Smith

Robert Staubs

Kristine Yu

All participants of the UMass Sound Seminar

Thank you!

References

Albright, A. & Bruce Hayes. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, 90, 119-161.

Berko, J. 1958. The child's learning of English morphology. *Word*, 14, 150–177.

Boersma, P. & J. Pater. 2008. Convergence properties of a Gradual Learning Algorithm for Harmonic Grammar. Ms, University of Amsterdam & University of Massachusetts Amherst. Available as ROA-970 from the Rutgers Optimality Archive.

Booij, G. & J. Rubach. 1992. Lexical Phonology. In: W. Bright (ed.), *International Encyclopedia of Linguistics*, 293-296. Oxford: OUP.

Domahs, U., I. Plag & R. Carroll. 2014. Word stress assignment in German, English and Dutch: Quantity-sensitivity and extrametricality revisited. *The Journal of Comparative Germanic Linguistics*, 17, 1, 59-96.

Giegerich, Heinz. 1985. Metrical phonology and phonological structure: German and English. Cambridge: Cambridge University Press.

References

Goldwater, S. & M. Johnson. 2003. Learning OT constraint rankings using a Maximum Entropy model. In: J. Spenader, A. Eriksson & Ö. Dahl (eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, Stockholm: Stockholm University, Department of Linguistics, pp. 111–120.

Hayes, B. 2004. Phonological acquisition in Optimality Theory: The early stages. In: R. Kager, J. Pater & W. Zonneveld, *Constraints in phonological acquisition*, Cambridge University Press, pp. 158-203.

Hayes, B. & C. Wilson. 2008. A Maximum Entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39, 3, 379-440.

van der Hulst, H.G. 1984. *Syllable structure and stress in Dutch*. Dordrecht: Foris.

Jarosz, G. 2006. Rich lexicons and restrictive grammars – Maximum likelihood learning in Optimality Theory. PhD dissertation, Rutgers University.

Jarosz, G. 2013. Learning with hidden structure in Optimality Theory and Harmonic Grammar: beyond Robust Interpretive Parsing. *Phonology*, 30, 1, 27-72.

Kager, R. 1989. *A metrical theory of stress and distressing in English and Dutch*. Doctoral dissertation. (= Linguistic Models 14). Dordrecht: Foris Publications.

McCarthy, J.J. 1981. The role of the evaluation metric in the acquisition of phonology. In: C.L. Baker & J.J. McCarthy (eds.), *The logical problem of language acquisition*, Cambridge, MA: MIT Press, pp. 218-248.

References

Nouveau, D. 1994. Language Acquisition, Metrical Theory and Optimality. A study of Dutch Word Stress. PhD dissertation, Universiteit Utrecht.

Oostendorp, M. van. 1997. Lexicale variatie in de Optimaliteitstheorie. [Lexical variation in Optimality Theory.] *Nederlandse Taalkunde*, 2, 133-154.

Oostendorp, M. van. 2012. Quantity and the three-syllable window in Dutch word stress. *Language and Linguistics Compass*, 6, 6, 343–358.

Pater, J., R. Staubs, K. Jesney & B. Smith. 2012. Learning probabilities over underlying representations. In: *Proceedings of the Twelfth Meeting of the ACL-SIGMORPHON: Computational Research in Phonetics, Phonology, and Morphology*, pp. 62-71.

Prince, A. & B. Tesar. 2004. Learning phonotactic distributions. In: R. Kager, J. Pater & W. Zonneveld, *Constraints in phonological acquisition*, Cambridge University Press, pp. 245-291.

Prince, A. & P. Smolensky. 1993/2004. Optimality Theory: Constraint interaction in generative grammar. Technical report, Rutgers University and University of Colorado at Boulder, 1993. ROA 537, 2002. Revised version published by Blackwell, 2004.

Smolensky, Paul. 1996. The Initial State and 'Richness of the Base'. Technical Report

JHU-CogSci-96-4.

References

Staubs, R. 2011. Harmonic Grammar in R (hgR). Software package.

Amherst, MA: University of Massachusetts Amherst.

http://blogs.umass.edu/hgr/

Tesar, B. 2006. Faithful contrastive features in learning. *Cognitive Science*, 30, 863–903.

Tesar, B. 2013. *Output-driven phonology: Theory and learning*. Cambridge Series in Linguistics 139. Cambridge: Cambridge University Press.

Tesar, B., J. Alderete, G. Horwood, N. Merchant, K. Nishitani and A. Prince. 2003. Surgery in language learning. In: G. Garding & M. Tsujimura (eds.), *WCCFL 22 Proceedings*, Somerville, MA: Cascadilla Press, pp. 477-490.

Tesar, B. & P. Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Wiese, Richard. 1996/2000. The phonology of German. Oxford: Oxford University Press.

Wilson, C. 2011. Bayesian inference for constraint-based phonology. Talk given at the University of Massachusetts Amherst on December 2 2011.

Appendix

Competing model

- Competing model: Yang (2005, 2011)
  - If pattern (e.g., Quantity Sensitivity) is to be represented in the grammar:
    - number of exceptions to pattern cannot exceed a certain threshold (derived from Zipf's law)
  - Type of exceptions does not matter

Competing model

- Current model:
  - grounded in assumptions already necessary for OT
  - while Yang's model requires adopting extra assumptions on top of those necessary for OT

- My model is more sensitive to the data (type of exceptions matters), while requiring fewer assumptions

Hidden structure

- Hidden structure for one input/output mapping

input	*output*	*hidden structure*	
/taˈtata/	[taˈtata]	L L L	ta(ˈtata)
/taˈtata/	[taˈtata]	L L L	ta(ˈta)ta
/taˈtata/	[taˈtata]	L L L	(taˈta)ta
/taˈtata/	[taˈtata]	L L H	ta(ˈtata)
…			
/taˈtata/	[taˈtata]	L H L	ta(ˈtata)
…			
/taˈtata/	[taˈtata]	H H H	ta(ˈtata)
/taˈtata/	[taˈtata]	H H H	ta(ˈta)ta
/taˈtata/	[taˈtata]	H H H	(taˈta)ta

Biases in learning

- Two biases added to the objective function:
  - Regularization bias: L2 regularization
    - Objective function penalized proportional to the sum of all constraint weights
      - Prevents constraint weights for being too high
    - Various values considered – see discussion below

  - Markedness-over-Faithfulness bias:
    - Objective function penalized proportional to the sum of all Faithfulness constraint weights
      - Prevents Faithfulness from being too powerful
    - Appropriate values tied to value of regularization bias