Learning Unbounded Stress Systems via Local Inference

Jeff Heinz

University of California, Los Angeles

October 14, 2006

Introduction

- I will present a tractable unsupervised batch learning algorithm which successfully learns the class of attested unbounded stress systems (Stowell 1979, Hayes 1981, Halle and Vergnaud 1987, Hayes 1995, Bailey 1995, Walker 2000, Bakovic 2004).
- The algorithm uses only:
  - a formalized notion of locality
  - and no Optimality-theoretic (OT) constraints (Prince and Smolensky 1993, 2004).

Overview

1. Learning in Phonology
2. Unbounded Stress Systems
3. Representations of Grammars
4. The Learner
5. Predictions
6. Conclusions

Learning in phonology

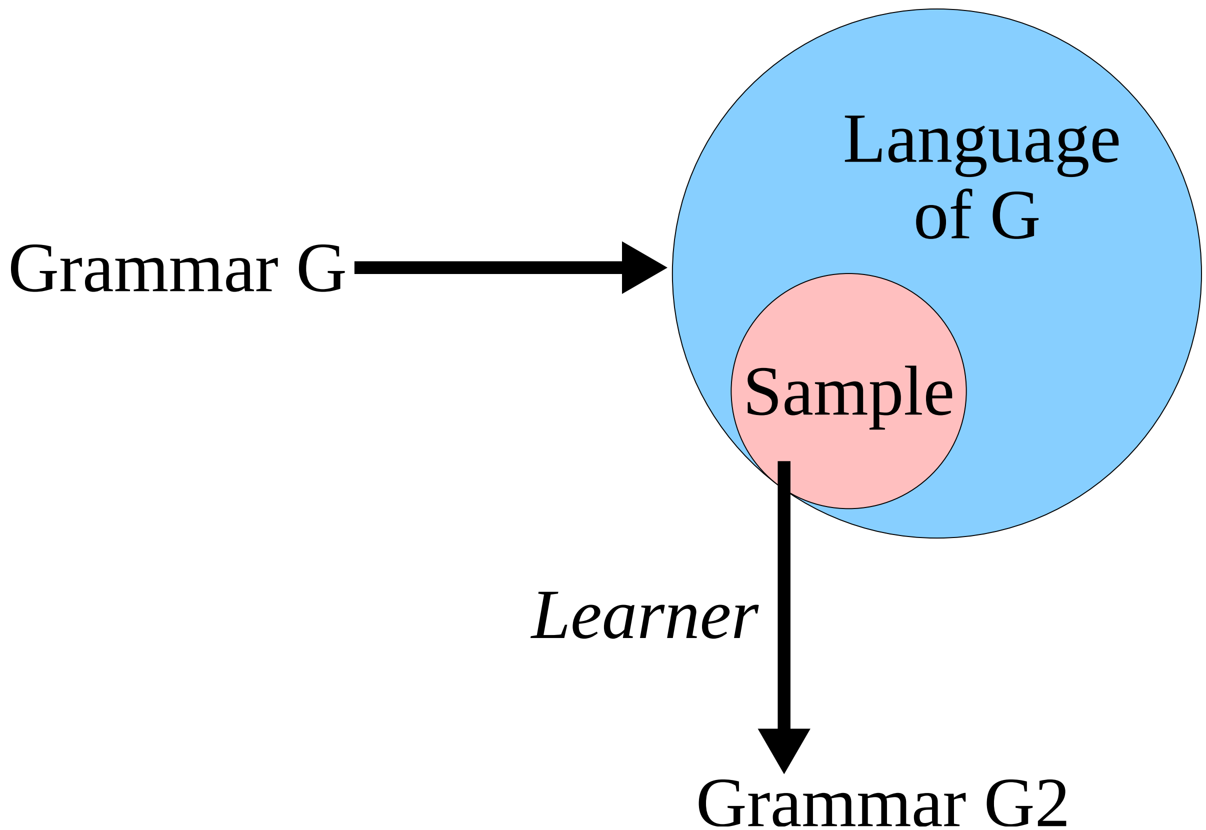
Learning in Optimality Theory (Tesar 1995, Boersma 1997, Tesar 1998, Tesar and Smolensky 1998, Hayes 1999, Boersma and Hayes 2001, Lin 2002, Pater and Tessier 2003, Pater 2004, Prince and Tesar 2004, Hayes 2004, Riggle 2004, Alderete et al. 2005, Merchant and Tesar to appear, Wilson 2006, Riggle 2006, Tessier 2006)

Learning in Principles and Parameters (Wexler and Culicover 1980, Dresher and Kaye 1990)

Learning Phonological Rules (Gildea and Jurafsky 1996, Albright and Hayes 2002, 2003)

Learning Phonotactics (Ellison 1992, Goldsmith 1994, Frisch 1996, Coleman and Pierrehumbert 1997, Frisch et al. 2004, Albright 2006, Goldsmith 2006, Heinz 2006a,b, Hayes and Wilson 2006)

The Learning Model



- What is *Learner* such that G = G2?

Premise

- We can study how learning or generalization occurs by *isolating* factors which play a role in the learning process.
- What are some of the relevant factors for phonotactic learning?
  1. Social factors: ‘the charismatic child’, . . .
  2. Phonetic factors: Articulatory, perceptual processes, . . .
  3. Similarity, locality, . . .
- We should ask: How can any one particular factor benefit learning (in some domain)?

Locality in Phonology

- "Consider first the role of counting in grammar. How long may a count run? General considerations of locality, ...suggest that the answer is probably 'up to two': a rule may fix on one specified element and examine a structurally adjacent element and no other." (McCarthy and Prince 1986:1)
- "...the well-established generalization that linguistic rules do not count beyond two ..." (Kenstowicz 1994:597)
- "...it was felt that phonological processes are essentially local and that all cases of nonlocality should derive from universal properties of rule application" (Halle and Vergnaud 1987:ix)

Locality and Learning

- How can this “well-established generalization” be formalized to benefit learning?

Unbounded Stress Systems

- Unbounded stress systems are sensitive to syllable weight and place no limits on the distances between stress and the word boundary.
- Hayes (1995) describes four basic types of attested unbounded systems.
  - Leftmost Heavy otherwise Leftmost (LHOL)
  - Leftmost Heavy otherwise Rightmost (LHOR)
  - Rightmost Heavy otherwise Leftmost (RHOL)
  - Rightmost Heavy otherwise Rightmos (RHOR)

Unbounded Stress Systems

- Bailey's (1995) database gives 22 variations of these basic types.

		Name	Stress Priority Code	Notes
LHOL	1.	Amele	12..89/1L	
	2.	Murik	12..89/1L	max 1 hvy/word
	3.	Serbo, Croatian	12..89/1L	at least 1 hvy/word
	4.	Maori	12..89/12..89/1L	
	5.	Kashmiri	12..78/12..78/1L	
	6.	Mongolian, Khalkha	12..89/2L	
LHOR	7.	Komi	12..89/9L	
RHOL	8.	Buriat	23..891/9R	
	9.	Cheremis, Eastern	23..89/9R	optional 1R
	10.	Nubian, Dongolese	23..89/9R	
	11.	Chuvash	12..89/9R	
	12.	Arabic, Classical	1/23..89/9R	
RHOR	13.	Golin	12..89/1R	
	14.	Mayan, Aguacatec	12..89/1R	max 1 hvy/word
	15.	Cheremis, Mountain	23..89/2R	words w/no hvys lex
	16.	Cheremis, Western	23..89/2R	
	17.	Seneca	23..89@s@w2/0R	
	18.	Sindhi	23..891/2R	
	19.	Cheremis, Meadow	1/23..891/1R	
	20.	Hindi (per Kelkar)	23..891/23..891/2R	
	21.	Klamath	12..89/23/3R	
	22.	Mam	12..89/12..89/12/2R	

Example: Leftmost Heavy otherwise Rightmost

- Komi (Hayes 1995, Itkonen 1955, Lytkin 1961) is a language with the 'Leftmost Heavy Otherwise Rightmost' pattern.

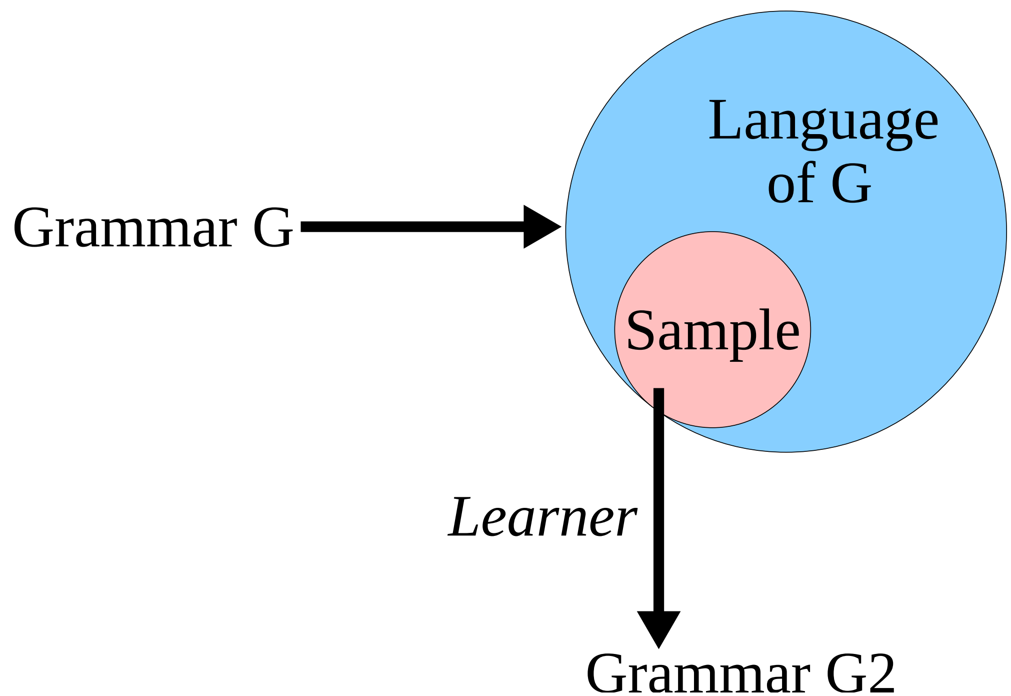
Rule: Stress the heavy syllable closest to the left edge. If there is no heavy syllable, stress the rightmost syllable.

Ex:
1. **H1** H0 H0
2. L0 L0 **H1** L0 L0
3. L0 L0 L0 **H1**
4. L0 L0 L0 **L1**

Key: H-Heavy, L-Light, 0-No stress, 1-Primary stress

Example: Leftmost Heavy otherwise Rightmost

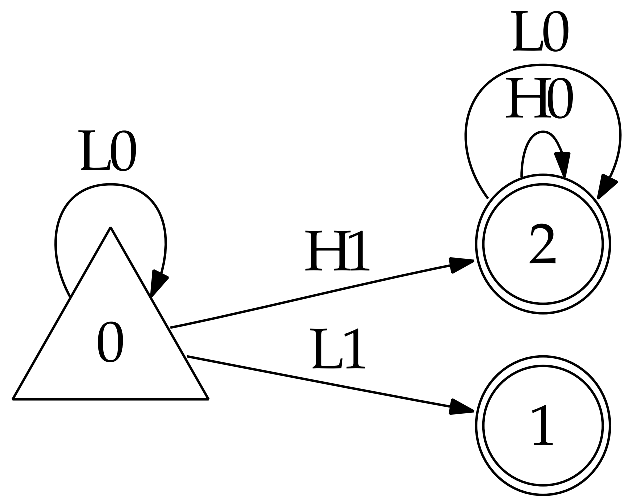
- How can we represent stress rules in the Grammar G?



Finite state acceptors as phonotactic grammars

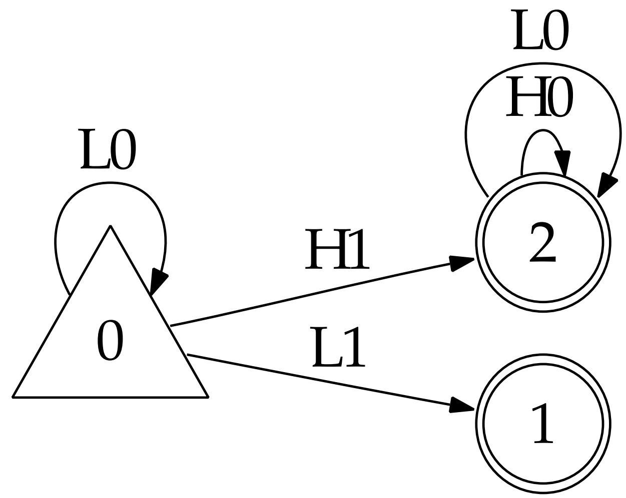
- They accept or reject words. So it meets the minimum requirement for a phonotactic grammar– a device that at least answers Yes or No when asked if some word is possible (Chomsky and Halle 1968, Halle 1978).
- They can be related to finite state OT models, which allow us to compute a phonotactic finite state acceptor (Riggle 2004), which becomes the target grammar for the learner.
- The grammars are well-defined and can be manipulated (Hopcroft et al. 2001). (See also Johnson (1972), Kaplan and Kay (1981, 1994), Ellison (1992), Eisner (1997), Albro (1998, 2005), Karttunen (1998), Riggle (2004), Karttunen (2006) for finite-state approaches to phonology.)

Leftmost Heavy otherwise Rightmost



- Note that the grammar above recognizes an infinite number of legal words, just like the generative grammars of earlier researchers.
- Also note that if the (different) OT analyses of the LHOR pattern given in Walker (2000) and Bakovic (2004) were encoded in finite-state OT, Riggles (2004) algorithm yields the (same) phonotactic acceptor above.

Leftmost Heavy otherwise Rightmost



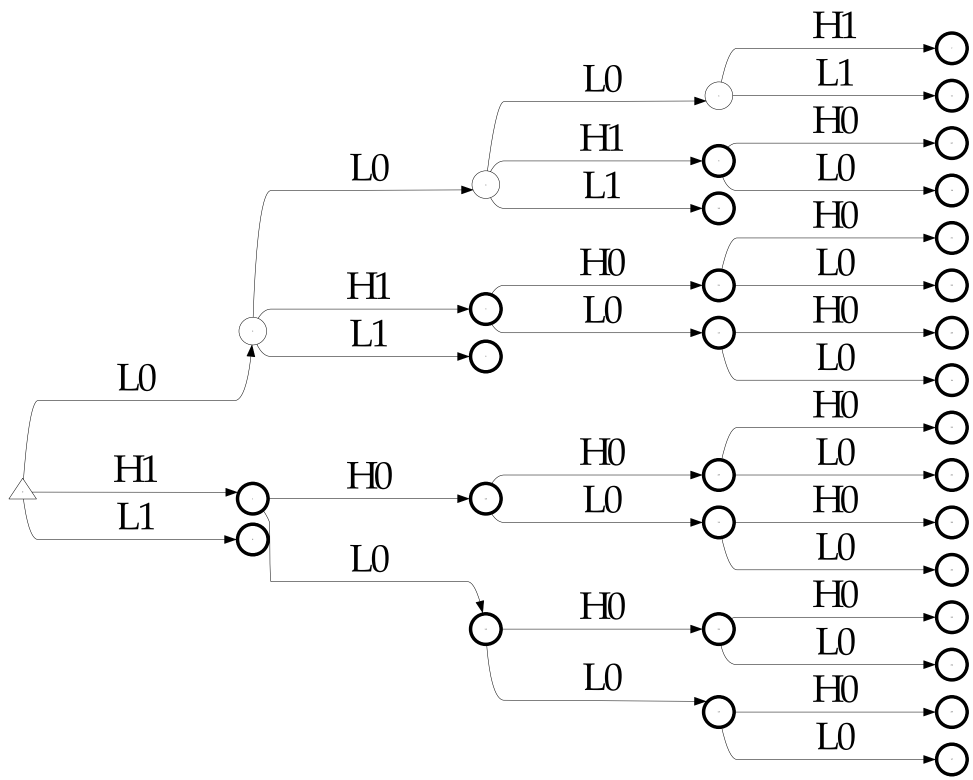
- How can this finite state acceptor be learned from a finite list of LHOR words?

H1	**L1**	**H1** L0	**H1** H0	L0 **H1**
L0 **L1**	**H1** L0 L0	**H1** L0 H0	**H1** H0 L0	**H1** H0 H0
L0 **H1** L0	L0 **H1** H0	L0 L0 **L1**	L0 L0 **H1**	L0 **H1** L0 L0
L0 **H1** L0 H0	**H1** L0 L0 L0	**H1** L0 L0 H0	**H1** H0 L0 L0	**H1** H0 L0 H0
L0 **H1** H0 L0	L0 **H1** H0 H0	**H1** L0 H0 L0	**H1** L0 H0 H0	**H1** H0 H0 L0
H1 H0 H0 H0	L0 L0 **H1** L0	L0 L0 **H1** H0	L0 L0 L0 **L1**	L0 L0 L0 **H1**

Overview of the Learner

- I will describe a simpler version of the learner first, and then describe the actual learner used in this study.
- The learner works in two stages (Cf. Angluin (1982)):
  1. Build a structured representation of the input– construct a 'prefix' tree
  2. Merge states which have the same local phonological environment– 'the neighborhood'

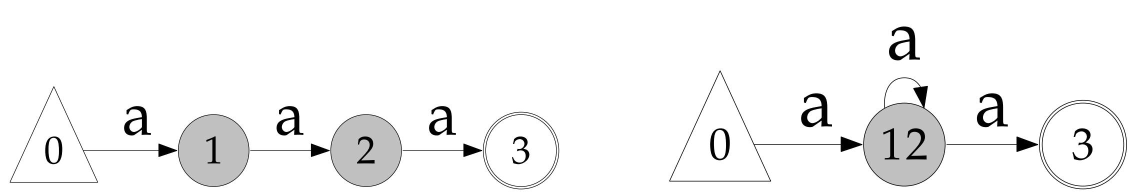
The prefix tree for LHOR



- A structured representation of the input (all thirty words of length four syllables or less).
- It accepts only the forms that have been observed.

State merging

- Generalize by *state-merging.*
  - a process where two states are identified as equivalent and then *merged* (i.e. combined).
- A key concept behind state merging is that transitions are preserved (Hopcroft et al. 2001, Angluin 1982).
- This is one way in which generalizations may occur—because the post-merged machine accepts everything the pre-merged machine accepts, possibly more.



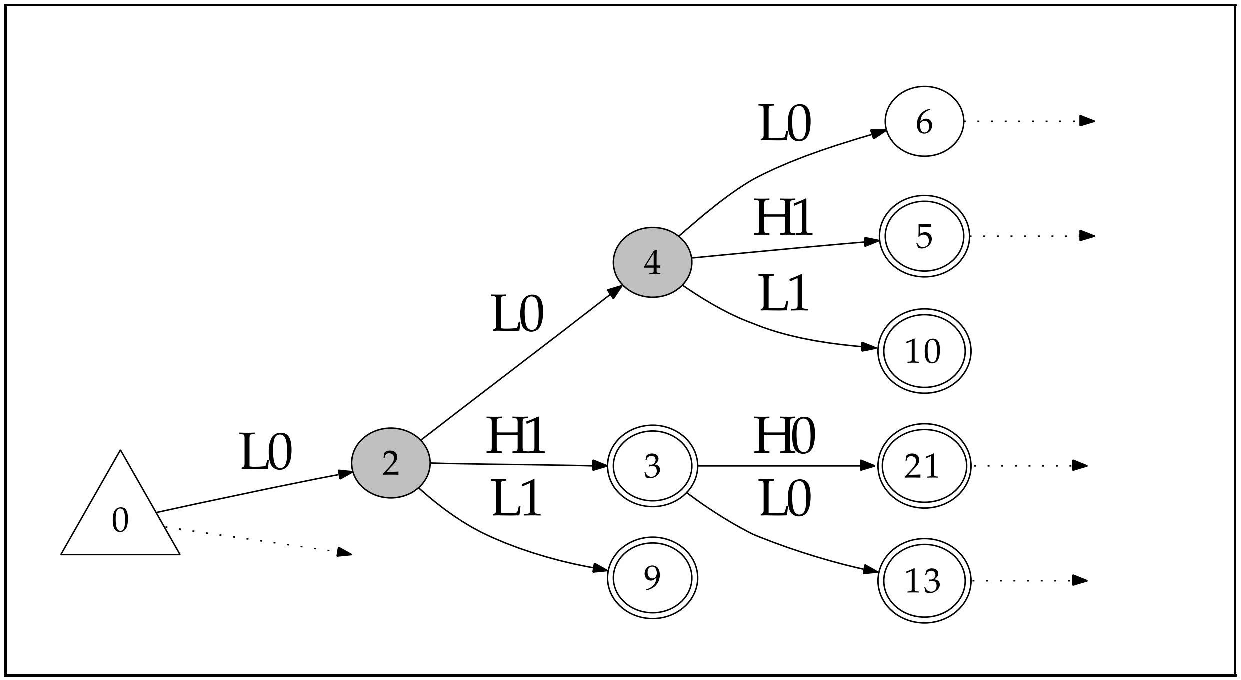
The learner's state merging criteria

- How does the learner decide whether two states are equivalent in the prefix tree?
- Merge states if their local environment is the same.
- I call this environment the *neighborhood*. It is:
  1. the set of incoming symbols to the state.
  2. the set of outgoing symbols to the state.
  3. whether it is a final state or not.
  4. whether it is a start state or not.
- The learner merges states in the prefix tree with the same neighborhood.

Example of neighborhoods

- States p and q have the same neighborhood.

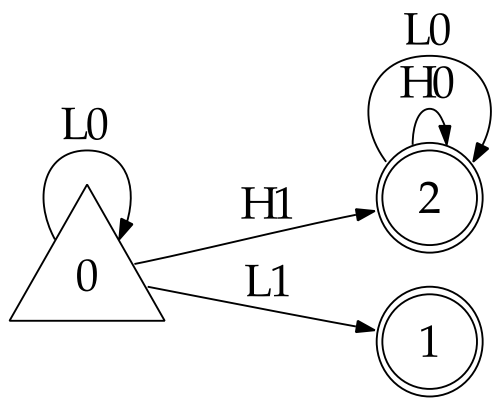
A section of the prefix tree enlarged



- States 2 and 4 have the same neighborhood.
- So these states are merged.

The result of merging states with the same neighborhood

(after minimization)



- The machine above accepts

. . . **H1** H0 H0, L0 **H1** L0 L0, L0 L0 **H1**, L0 L0 **L1**

- The learner has acquired the unbounded stress pattern LHOR, i.e. it has generalized exactly as desired.

Summary of the Forward Learner

1. Builds a prefix tree of the observed words.
2. Merges states in this machine that have the same neighborhood.

Summary of the Forward Learner

- This learner successfully learns 17 of the 22 systems.

		Name	Stress Priority Code	Notes	FL
LHOL	1.	Amele	12..89/1L		✓(5)
	2.	Murik	12..89/1L	max 1 hvy/word	✓(4)
	3.	Serbo, Croatian	12..89/1L	at least 1 hvy/word	✓(4)
	4.	Maori	12..89/12..89/1L		✓(5)
	5.	Kashmiri	12..78/12..78/1L		×
	6.	Mongolian, Khalkha	12..89/2L		✓(5)
LHOR	7.	Komi	12..89/9L		✓(4)
RHOL	8.	Buriat	23..891/9R		×
	9.	Cheremis, Eastern	23..89/9R	optional 1R	✓(4)
	10.	Nubian, Dongolese	23..89/9R		✓(5)
	11.	Chuvash	12..89/9R		✓(4)
	12.	Arabic, Classical	1/23..89/9R		✓(4)
RHOR	13.	Golin	12..89/1R		✓(5)
	14.	Mayan, Aguacatec	12..89/1R	max 1 hvy/word	✓(4)
	15.	Cheremis, Mountain	23..89/2R	words w/no hvys lex	✓(6)
	16.	Cheremis, Western	23..89/2R		✓(6)
	17.	Seneca	23..89@s@w2/0R		✓(7)
	18.	Sindhi	23..891/2R		×
	19.	Cheremis, Meadow	1/23..891/1R		✓(5)
	20.	Hindi (per Kelkar)	23..891/23..891/2R		×
	21.	Klamath	12..89/23/3R		×
	22.	Mam	12..89/12..89/12/2R		✓(5)

Directionality and the Prefix Tree

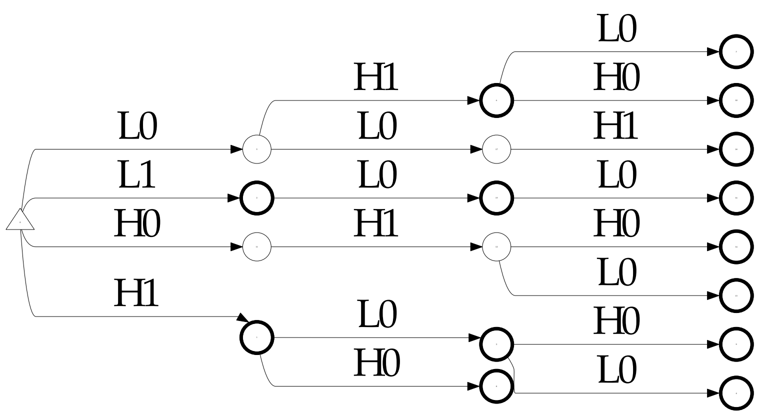
- The five patterns the Forward Learner fails to learn— Buriat, Hindi (per Kelkar), Kashmiri, Klamath, and Sindhi— are typically analyzed as having a metrical unit at the right word edge.
- In each case the learner overgeneralized (i.e. accepted a language strictly larger than the target language).

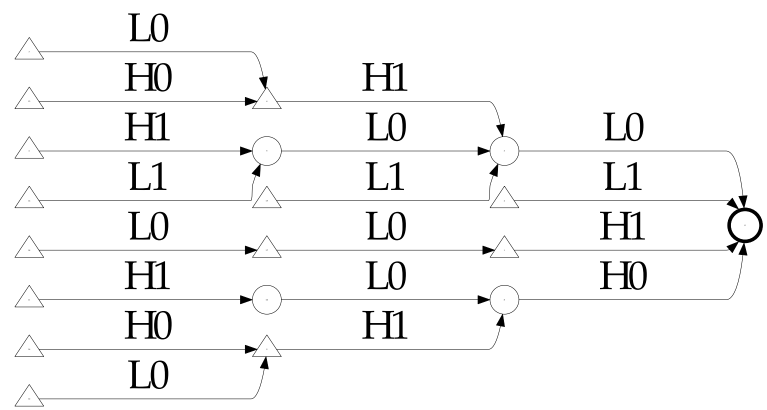
Elaborating the Forward Learner

- The learner in this study is more elaborate than the Forward Learner.
  - The generalization strategy is the same.
  - But it addresses the inherent left-to-right bias in prefix trees.
  - This must be addressed since stress patterns are sensitive to *both* word edges.
- Thus the few failures of the Forward Learner are attributed *not* to the generalization strategy but rather to an inherent bias of the (independent) choice of how the input is represented.

Suffix Trees

- If the input were represented with a *suffix tree*, then the structure obtained has the reverse bias, a right-to-left bias.





Prefix Tree for Buriat Stress Suffix Tree for Buriat Stress

(all words three syllables or less)

- Notice that these two representations are not mirror images of each other, they have different structures, though both accept exactly the same (finite) set of words.

The Forward-Backward Neighborhood Learner

- The Forward-Backward Neighborhood Learner
  1. Build a forward prefix tree and merge states with the same neighborhood.
  2. Build a suffix tree and merge states with the same neighborhood.
  3. Intersect these two machines to get the final grammar.
     – Intersection of two acceptors A and B results in an acceptor which only accepts words accepted by both A and B (Hopcroft et al. 2001).

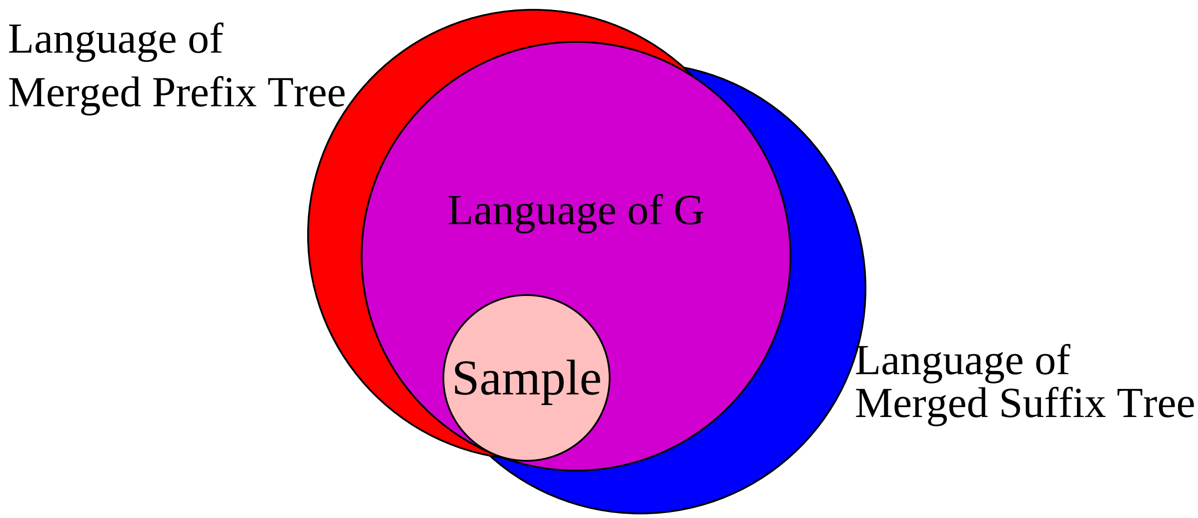
Summary of the Forward Backward Learner

- This learner successfully learns every system.

		Name	Stress Priority Code	Notes	FBL
LHOL	1.	Amele	12..89/1L		✓(5)
	2.	Murik	12..89/1L	max 1 hvy/word	✓(4)
	3.	Serbo, Croatian	12..89/1L	at least 1 hvy/word	✓(4)
	4.	Maori	12..89/12..89/1L		✓(5)
	5.	Kashmiri	12..78/12..78/1L		✓(6)
	6.	Mongolian, Khalkha	12..89/2L		✓(5)
LHOR	7.	Komi	12..89/9L		✓(4)
RHOL	8.	Buriat	23..891/9R		✓(5)
	9.	Cheremis, Eastern	23..89/9R	optional 1R	✓(4)
	10.	Nubian, Dongolese	23..89/9R		✓(5)
	11.	Chuvash	12..89/9R		✓(4)
	12.	Arabic, Classical	1/23..89/9R		✓(4)
RHOR	13.	Golin	12..89/1R		✓(5)
	14.	Mayan, Aguacatec	12..89/1R	max 1 hvy/word	✓(4)
	15.	Cheremis, Mountain	23..89/2R	words w/no hvys lex	✓(6)
	16.	Cheremis, Western	23..89/2R		✓(6)
	17.	Seneca	23..89@s@w2/0R		✓(7)
	18.	Sindhi	23..891/2R		✓(6)
	19.	Cheremis, Meadow	1/23..891/1R		✓(5)
	20.	Hindi (per Kelkar)	23..891/23..891/2R		✓(6)
	21.	Klamath	12..89/23/3R		✓(6)
	22.	Mam	12..89/12..89/12/2R		✓(5)

Why it works: Intersection keeps robust generalizations

- In only a prefix (suffix) tree is used then sometimes the state merging procedure *overgeneralizes*.
- The Forward-Backward Learner works because it is conservative—it keeps only the robust generalizations—those made in both the prefix and suffix trees (see appendix).



Unbounded and Quantity-Insensitive Systems

- Quantity-insensitive (QI) stress systems as described by Gordon (2002) are also learned by this learner (Heinz 2006b).
- QI stress systems are typically considered to be much simpler in character than unbounded stress systems.
- Thus, it is striking that the learner succeeds for both classes, suggesting that these two classes have something in commmon.

Why it works: Neighborhood-distinctness

- A language (regular set) is *neighborhood-distinct* iff there is an acceptor for the language such that each state has its own unique neighborhood.
- Every unbounded stress pattern, like every quantity-insensitive stress pattern, is neighborhood-distinct (this can be verified upon inspection).

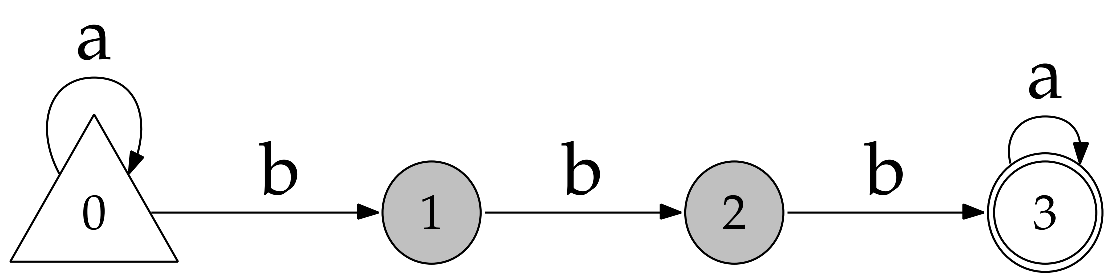
Learning Neighborhood-distinctness

- Because the learner merges states with the same neighborhood, it learns neighborhood-distinct patterns.
- Thus, the learner is really taking advantage of a previously unnoticed universal property of these grammars: neighborhood-distinctness.

Neighborhood-Distinct Hypotheses

- The relevant phonological environment in phonotactic learning is the neighborhood.
- All phonotactic patterns are neighborhood-distinct.

Example of a non-neighborhood-distinct language: $\mathtt{a}^*\mathtt{bbb}\mathtt{a}^*$



- It is not possible to build an acceptor for a language requiring words to have exactly three identical adjacent elements because there will always be two states with the same neighborhoods.

Consequences of Neighborhood-distinctness for the typology of stress

- Consequently, binary, ternary, and—as we have seen—unbounded systems can be learned by neighborhood learning, but not higher $\mathfrak{n}$-ary stress systems.
  - Example: 4-ary: 1, 10, 100, 1000, 10002, 100020, 1000200, 10002000, 100020002, . . .
- The learner fails here because it cannot in some sense "count beyond two."

N-gram models

- In this respect, this learner compares favorably to n-gram models:

  1. N-gram models cannot learn unbounded stress patterns unless they operate on tiers distinguished by syllable weight (e.g. a heavy syllable tier).
  2. A 4-gram model is needed to learn antepenultimate stress patterns, but 4-gram models also admit patterns with 4-syllable sized feet.

Comparisons to other theories

- Some ways the prohibition of three adjacent unstressed syllables in bounded systems is explained:
  1. Only one 'stray' syllable may occur between binary feet (Hayes 1995).
  2. *ExtendedLapse (Gordon 2002).
- Why binary feet and 'stray' syllables, or why just one 'stray' syllable? And why not *EvenMoreExtendedLapse?
  - The answers to these questions fall out from neighborhood-distinctness.

Neighborhood-distinctness

- It is an abstract notion of locality.
- It is novel.
- It serves as a strategy for learning by limiting the kinds of generalizations that can be made (e.g. cannot distinguish ‘three’ from ‘more than two’)
- It places real limits on typology: only finitely many languages are neighborhood-distinct (since there are only finitely many neighborhoods given some alphabet).

Unlearnable stress patterns

- It was discovered that if secondary stress is excluded from the grammars of Klamath (Barker 1963, 1964, Hammond 1986, Hayes 1995) and Seneca (Chafe 1977, Stowell 1979, Prince 1983, Hayes 1995), then the Forward Backward Neighborhood Learner fails to learn these grammars.
- It fails because, in the actual grammars of Klamath and Seneca, the presence of secondary stress distinguishes the neighborhoods of certain states.
- Removing secondary stress causes the patterns to no longer be neighborhood-distinct and hence unlearnable.

Open Questions

- Do human learners behave similarly?

Learnable unnatural patterns

- There are stress patterns that can be learned by neighborhood learning which are not considered natural by phonotactics.
  1. Leftmost Light otherwise Rightmost.
  2. A stress pattern requiring both lapses and clashes.
  3. A stress pattern where all syllables have primary stress.

Locality is but one factor in learning

- It is restrictive: it approximates the attested stress systems in an interesting way.
- This work belongs to a larger research program which is to identify and isolate properties of natural language which are helpful to learning.
- We should ask: What other properties exist
  1. which better approximate the class of possible stress systems?
  2. and which might assist learning?

Conclusions I

1. Every attested unbounded stress pattern as described by Bailey (1995), like the attested QI stress systems as described by Gordon (2002), can be learned by the above algorithm above because these patterns have something in common.
   - They are neighborhood-distinct.
2. The learner succeeds because it generalizes by identifying environments as the same if they are locally the same (i.e. merging same-neighborhood states).

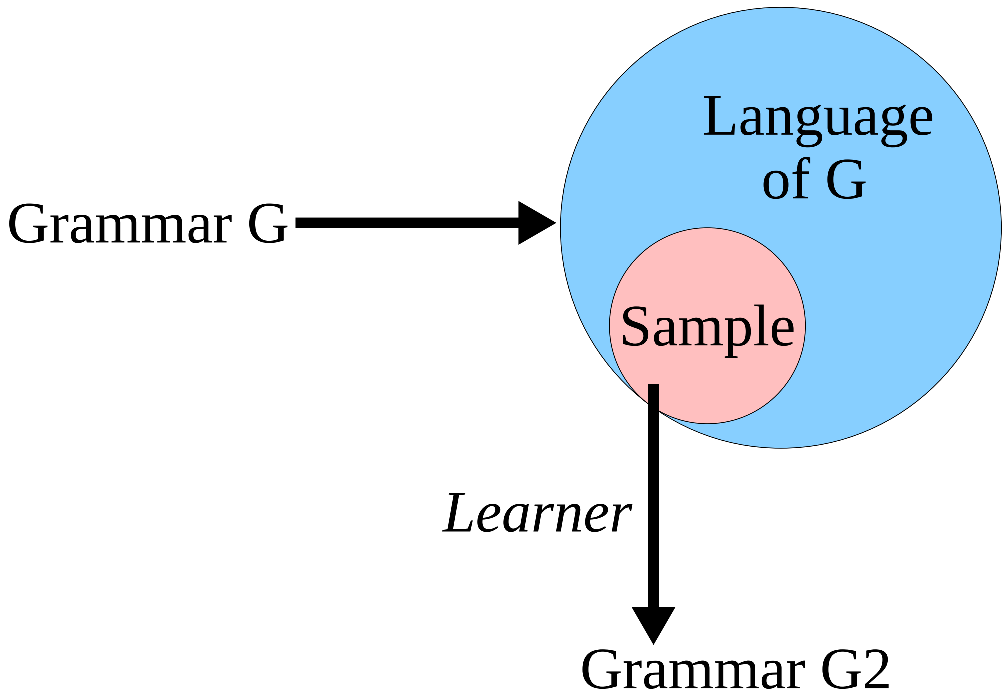
Conclusions II

1. We can approach the learning problem by developing models that isolate specific factors to study how they benefit learning.

Further Questions

- How does the learner perform on quantity-sensitive bounded systems? (in progress)
- How does the learner perform with segmental phonotactics? (for one approach learner see Heinz (2006a)).
- How can the learner be modified to handle noise or gradient phonotactics?

Thank You.



- Special thanks to Bruce Hayes, Ed Stabler, Colin Wilson and Kie Zuraw for insightful comments and suggestions related to this material. I also thank Greg Kobele, Andy Martin, Katya Pertsova, Shabnam Schademan, and Sarah VanWagnenen for helpful discussion.

Summary of the Backward Learner

1. Builds a suffix tree of the observed words.
2. Merges states in this machine that have the same neighborhood.

Summary of the Backward Learner

- This learner successfully learns 21 of the 22 systems.

		Name	Stress Priority Code	Notes	FL
LHOL	1.	Amele	12..89/1L		✓(5)
	2.	Murik	12..89/1L	max 1 hvy/word	✓(4)
	3.	Serbo, Croatian	12..89/1L	at least 1 hvy/word	✓(4)
	4.	Maori	12..89/12..89/1L		✓(5)
	5.	Kashmiri	12..78/12..78/1L		×
	6.	Mongolian, Khalkha	12..89/2L		✓(5)
LHOR	7.	Komi	12..89/9L		✓(4)
RHOL	8.	Buriat	23..891/9R		×
	9.	Cheremis, Eastern	23..89/9R	optional 1R	✓(4)
	10.	Nubian, Dongolese	23..89/9R		✓(5)
	11.	Chuvash	12..89/9R		✓(4)
	12.	Arabic, Classical	1/23..89/9R		✓(4)
RHOR	13.	Golin	12..89/1R		✓(5)
	14.	Mayan, Aguacatec	12..89/1R	max 1 hvy/word	✓(4)
	15.	Cheremis, Mountain	23..89/2R	words w/no hvys lex	✓(6)
	16.	Cheremis, Western	23..89/2R		✓(6)
	17.	Seneca	23..89@s@w2/0R		✓(7)
	18.	Sindhi	23..891/2R		×
	19.	Cheremis, Meadow	1/23..891/1R		✓(5)
	20.	Hindi (per Kelkar)	23..891/23..891/2R		×
	21.	Klamath	12..89/23/3R		×
	22.	Mam	12..89/12..89/12/2R		✓(5)

Directionality and the Suffix Tree

- The one pattern the Backward Learner fails to learn—Seneca—is typically analyzed as having a metrical unit at the left word edge.
- In each case the learner overgeneralized (i.e. accepted a language strictly larger than the target language).

References

Albright, Adam. 2006. Gradient Phonotactic effects: lexical? grammatical? both? neither? Talk handout from the 80th Annual LSA Meeting, Albuquerque, NM.

Albright, Adam and Bruce Hayes. 2002. Modeling English past tense intuitions with minimal generalization. *SIGPHON 6: Proceedings of the Sixth Meeting of the ACL Special Interest Group in Computational Phonology* :58–69.

Albright, Adam and Bruce Hayes. 2003. Rules vs. Analogy in English Past Tenses: A Computational/Experimental Study. *Cognition* 90:119–161.

Albro, Dan. 1998. Evaluation, implementation, and extension of Primitive Optimality Theory. Master's thesis, University of California, Los Angeles.

Albro, Dan. 2005. A Large-Scale, LPM-OT Analysis of Malagasy. Ph.D. thesis, University of California, Los Angeles.

Alderete, John, Adrian Brasoveanua, Nazarre Merchant, Alan Prince, and Bruce Tesar. 2005. Contrast analysis aids in the learning of phonological underlying forms. In *The Proceedings of WCCFL 24*. pages 34–42.

Angluin, Dana. 1982. Inference of Reversible Languages. *Journal for the Association of Computing Machinery* 29(3):741–765.

Bailey, Todd. 1995. Nonmetrical Constraints on Stress. Ph.D. thesis, University of Minnesota. Ann Arbor, Michigan. Stress System Database available at http://www.cf.ac.uk/psych/ssd/index.html.

Bakovic, Eric. 2004. Unbounded stress and factorial typology. In *Optimality Theory in Phonology: A Reader*, edited by John McCarthy. Blackwell, London. ROA-244, Rutgers Optimality Archive, http://roa.rutgers.edu/.

Barker, Muhammad. 1963. *Klamath Dictionary*, volume 31 of *University of California Publications in Linguistics*. University of California Press, Berkeley.

Barker, Muhammad. 1964. *Klamath Grammar*, volume 32 of *University of California Publications in Linguistics*. University of California Press, Berkeley.

Boersma, Paul. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences* 21. University of Amsterdam.

Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32:45–86.

Chafe, Wallace. 1977. Accent and Related Phenomena in the Five Nations Iroquois Languages. In *Studies in Stress and Accent*, edited by Larry Hyman, volume 4 of *Southern California Occasional Papers in Lingustics*. Department of Linguistics, University of Southern California, pages 169–181.

Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English*. Harper & Row.

Coleman, John and Janet Pierrehumbert. 1997. Stochastic Phonological Grammars and Acceptability. In *Computational Phonology*. Somerset, NJ: Association for Computational Linguistics, pages 49–56. Third Meeting of the ACL Special Interest Group in Computational Phonology.

Dresher, Elan and Jonathan Kaye. 1990. A Computational Learning Model for Metrical Phonology. *Cognition* 34:137–195.

Eisner, Jason. 1997. What Constraints Should OT Allow? Talk handout, Linguistic Society of America, Chicago. Available on the Rutgers Optimality Archive, ROA#204-0797, http://roa.rutgers.edu/.

Ellison, M. T. 1992. The Machine Learning of Phonological Structure. Ph.D. thesis, University of Western Australia.

Frisch, S., J. Pierrehumbert, and M. Broe. 2004. Similarity Avoidance and the OCP. *Natural Language and Linguistic Theory* 22:179–228.

Frisch, Stephan. 1996. Similarity and Frequency in Phonology. Ph.D. thesis, Northwestern University.

Gildea, Daniel and Daniel Jurafsky. 1996. Learning Bias and Phonological-rule Induction. *Association for Computational Linguistics* .

Goldsmith, John. 1994. A Dynamic Computational Theory of Accent Systems. In *Perspectives in Phonology*, edited by Jennifer Cole and Charles Kisseberth. Stanford: Center for the Study of Language and Information, pages 1–28.

Goldsmith, John. 2006. Information Thoery and Phonology. Slides presented at the 80th Annual LSA in Alburquerque, New Mexico.

Gordon, Matthew. 2002. A Factorial Typology of Quantity-Insensitive Stress. *Natural Language and Linguistic The-*

ory 20(3):491–552. Additional appendices availables at http://www.linguistics.ucsb.edu/faculty/gordon/pubs.html.

Halle, Morris. 1978. Knowledge Unlearned and Untaught: What Speakers Know about the Sounds of Their Language. In *Linguistic Theory and Psychological Reality*. The MIT Prss.

Halle, Morris and Jean-Roger Vergnaud. 1987. *An Essay on Stress*. The MIT Press.

Hammond, Michael. 1986. The Obligatory Branching Parameter in Metrical Theory. *Natural Language and Linguistic Theory* 4:185–228.

Hayes, Bruce. 1981. A Metrical Theory of Stress Rules. Ph.D. thesis, Massachussetts Institute of Technology. Revised version distribued by Indiana University Linguistics Club, Bloomington, and published by Garland Press, New York 1985.

Hayes, Bruce. 1995. *Metrical Stress Theory*. Chicago University Press.

Hayes, Bruce. 1999. Phonetically-Driven Phonology: The Role of Optimality Theory and Inductive Grounding. In *Functionalism and Formalism in*

Linguistics, Volume I: General Papers. John Benjamins, Amsterdam, pages 243–285.

Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: the early stages. In *Fixing Priorities: Constraints in Phonological Acquisition*, edited by Rene Kager, Joe Pater, and Wim Zonneveld. Cambridge University Press.

Hayes, Bruce and Colin Wilson. 2006. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. Submitted. Available at http://www.linguistics.ucla.edu/people/wilson/papers.html.

Heinz, Jeffrey. 2006a. Learning Phonotactic Grammars without Surface Forms. In *Proceedings of the 25th West Coast Conference of Formal Linguistics*, edited by Donald Baumer, David Montero, and Michael Scanlon. University of Washington, Seattle.

Heinz, Jeffrey. 2006b. Learning Quantity Insensitive Stress Systems via Local Inference. In *Proceedings of the Eighth Meeting of the ACL Special Interest Group in Computational Phonology at HLT-NAACL*. pages 21–30. New York City, USA.

Hopcroft, John, Rajeev Motwani, and Jeffrey Ullman. 2001. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley.

Itkonen, Erkki. 1955. Ueber die Betonungsverhältnisse in den finnisch-ugrischen Sprachen. *Acta Linguistica Academiae Scientiarum Hungaricae* 5:21–23.

Johnson, C. Douglas. 1972. *Formal Aspects of Phonological Description*. The Hague: Mouton.

Kaplan, Ronald and Martin Kay. 1981. Phonological Rules and Finite State Transducers. Paper presented at ACL/LSA Conference, New York.

Kaplan, Ronald and Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics* 20(3):331–378.

Karttunen, Lauri. 1998. The proper treatment of optimality theory in computational phonology. *Finite-state methods in natural language processing* :1–12.

Karttunen, Lauri. 2006. The Insufficiency of Paper-and-Pencil Linguistics: the Case of Finnish Prosody. Rutgers Optimality Archive #818-0406.

Lin, Ying. 2002. Probably Approximately Correct Learning of Constraint Ranking. Master's thesis, University of California, Los Angeles.

Lytkin, I. V. 1961. Komi-iaz'vinskii dialekt. Izdatel'stvo Akademii Nauk SSSR Moscow.

McCarthy, John and Alan Prince. 1986. Prosodic Morphology. Ms., Department of Linguistics, University of Massachusetts, Amherst, and Program in Linguistics, Brandeis University, Waltham, Mass.

Merchant, Nazarre and Bruce Tesar. to appear. Learning underlying forms by searching restricted lexical subspaces. In *The Proceedings of CLS 41*. ROA-811.

Pater, Joe. 2004. Exceptions and Optimality Theory: Typology and Learnability. Conference on Redefining Elicitation: Novel Data in Phonological Theory. New York University.

Pater, Joe and Anne Marie Tessier. 2003. Phonotactic Knowledge and the Acquisition of Alternations. In *Proceedings of the 15th International Congress on Phonetic Sciences, Barcelona*, edited by M.J. Solé, D. Recasens, and J. Romero. pages 1777–1180.

Prince, Alan. 1983. Relating to the Grid. *Linguistic Inquiry* 14(1).

Prince, Alan and Paul Smolensky. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. Technical Report 2, Rutgers University Center for Cognitive Science.

Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell Publishing.

Prince, Alan and Bruce Tesar. 2004. Fixing priorities: constraints in phonological acquisition. In *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge: Cambridge University Cambridge: Cambride University Press.

Riggle, Jason. 2004. Generation, Recognition, and Learning in Finite State Optimality Theory. Ph.D. thesis, University of California, Los Angeles.

Riggle, Jason. 2006. Using Entropy to Learn OT Grammars From Surface Forms Alone. In *Proceedings of the 25th West Coast Conference of Formal Linguistics*, edited by Donald Baumer, David Montero, and Michael Scanlon. Cascadilla Proceedings Project.

Stowell, T. 1979. Stress Patterns of the World, Unite! In *MIT Working Papers in Linguistics*. Department of Linguistics, MIT, Cambridge, MA.

Tesar, Bruce. 1995. Computational Optimality Theory. Ph.D. thesis, University of Colorado at Boulder.

Tesar, Bruce. 1998. An Interative Strategy for Language Learning. *Lingua* 104:131–145.

Tesar, Bruce and Paul Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29:229–268.

Tessier, Anne-Michelle. 2006. Stages of OT Phonological Acquisition and Error-Selective Learning. In *Proceedings of the 25th West Coast Conference of Formal Linguistics*, edited by Donald Baumer, David Montero, and Michael Scanlon. Cascadilla Proceedings Project.

Walker, Rachel. 2000. Mongolian Stress, Licensing, and Factorial Typology. ROA-172, Rutgers Optimality Archive, http://roa.rutgers.edu/.

Wexler, Kenneth and Peter Culicover. 1980. *Formal Principles of Language Acquisition*. MIT Press.

Wilson, Colin. 2006. The Luce Choice Ranker. Talk handout, UCLA Phonology Seminar.