Learning Long-Distance Agreement Phonotactics

Jeffrey Heinz
jheinz@humnet.ucla.edu
University of California, Los Angeles

The 81st meeting of the LSA
Anaheim, California, 2007

1 Introduction

Introduction

- I will present a learning algorithm that learns long-distance agreement phonotactic patterns without *a priori* Optimality-theoretic constraints (Prince and Smolensky 1993, 2004).
- The proposed algorithm simply keeps track of *precedence* relations.
- This approach demonstrates the utility of factoring the learning problem.

Outline

Contents

1.1 Long-Distance Agreement

What is Long-Distance Agreement?

- Long Distance Agreement (LDA) patterns are those within which particular segments, separated by at least one other segment, must (dis)agree in some feature (Hansson 2001, Rose and Walker 2004).

- Hansson (2001) adds that the intervening segments are not audibly affected by the agreeing feature.
- This is in order to clearly distinguish LDA from spreading (see also Gafos 1999 and Walker 1998).

Examples of Long-Distance Agreement

- Consonantal Harmony (Hansson 2001, Rose and Walker 2004)
  - Sibilant Harmony
  - Liquid Harmony
  - Dorsal Harmony
  - . . .
- Vowel Harmony with transparent vowels
  - Finnish, Hungarian, Nez Perce (see Baković 2000 and references therein)
  - But see also Gordon (1999), Gafos and Benus (2003), and Gick et. al. (2006).

LDA with No Blocking: Navajo

In well formed words, sibilants agree in the feature [anterior].

1.	[s,z,ts,ts’,dz] are never preceded by [ʃ,ʒ,tʃ,tʃ’,dʒ].
2.	[ʃ,ʒ,tʃ,tʃ’,dʒ] are never preceded by [s,z,ts,ts’,dz].

Examples (Sapir and Hojier 1967):

1.	**ʃ**iːteː**ʒ**	‘we (dual) are lying’
2.	da**s**doːli**s**	‘he (4th) has his foot raised’
3.	***ʃ**iːteː**z**	(hypothetical)
4.	*da**s**doːli**ʃ**	(hypothetical)

LDA with Local Blocking: Ineseño Chumash

In well formed words:

1.	[ʃ] is never preceded by [s].
2.	[s] is never preceded by [ʃ] unless the nearest preceding [ʃ] is immediately followed by [n,t,l].

Examples (Applegate 1972, Poser 1982):

1.	k**s**unonu**s**	‘I obey him’	5.	**ʃt**ijepu**s**	‘he tells him’
2.	k**ʃ**unot**ʃ**	‘I am obedient’	6.	*su**st**ime**ʃ**	(hypothetical)
3.	*k**s**unonu**ʃ**	(hypothetical)	7.	**ʃ**i**ʃl**u**s**i**s**in	‘they (dual) are gone awry’
4.	*k**ʃ**unot**s**	(hypothetical)			

Why LDA Patterns are Thought to be a Challenge to Learn

Arbitrarily many segments may intervene between agree-ers.

- Albright and Hayes (2003a) observe that "the number of logically possible environments... rises exponentially with the length of the string."
- Thus there are potentially too many environments for a learner to consider in discovering LDA patterns.

The Meaning of "arbitrarily many"

- However, does "arbitrarily many" really require a learner to consider every logically possible nonlocal environment?

1.2 Learning in Phonology

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), Niyogi (2006)

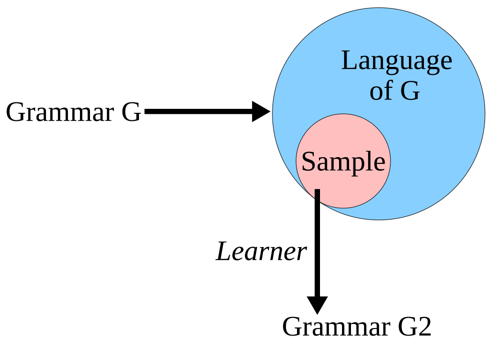
- *Learning Phonological Rules*

  Gildea and Jurafsky (1996), Albright and Hayes (2002, 2003a,b)

- *Learning Phonotactics*

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

The Learning Framework



- What is *Learner* so that *Language of G2 = Language of G?*
- See Nowak et. al. (2002) and Niyogi (2006) for overviews.

Inductive Learning and the Hypothesis Space

- Learning cannot take place unless the hypothesis space is restricted.
- G2 is not drawn from an unrestricted set of possible grammars.
- The hypotheses available to the learner ultimately determine:
  (1) the kinds of generalizations made
  (2) the range of possible natural language patterns
- Under this perspective, Universal Grammar (UG) is the set of available hypotheses.

Different Kinds of Hypothesis Spaces are Learned Differently.

- The set of syntactic hypotheses available to children is not the same as the set of phonological hypotheses available to children.
  - The two domains do not have the same kind of patterns and so we expect them to have different kinds of learners.
- Likewise, the set of Long Distance Agreement patterns are different from patterns which restrict the distribution of adjacent segments.

Factoring the Phonotactic Learning Problem

- Different kinds of phonotactic constraints can be learned by different learning algorithms.
- A complete phonotactic learner is a combination of these different learning algorithms.
- Here, I am only showing how one part of the whole learner—the part that learns LDA constraints—can work.

2 Learning Long Distance Agreement

2.1 Representing LDA Patterns

Representing LDA Patterns with Finite-state Machines

- *LDA patterns are regular*—that is, describable by a finite-state acceptor (Johnson 1972, Kaplan and Kay 1981, 1994, Ellison 1992, Eisner 1997, Albro 1998, 2005, Karttunen 1998, Frank and Satta 1998, Riggle 2004, Karttunen 2006)
- Finite-state acceptors

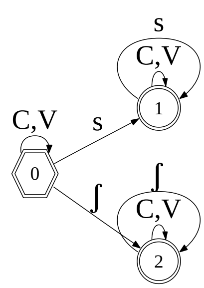
(1) accept or reject words. So it meets the minimum requirement for a phonotactic grammar–a device that at least answers Yes or No when asked if some word is possible. (Chomsky and Halle 1968, Halle 1978)

(2) 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.

(3) are well-defined and can be manipulated. (Hopcroft et. al. 2001).

LDA with No Blocking: Navajo

1.	[s,z,ts,ts',dz] are never preceded by [ʃ,ʒ,tʃ,tʃ',dʒ].
2.	[ʃ,ʒ,tʃ,tʃ',dʒ] are never preceded by [s,z,ts,ts',dz].



C = any consonant except sibilants
s = [+anterior] sibilants
V = any vowel
ʃ = [-anterior] sibilants

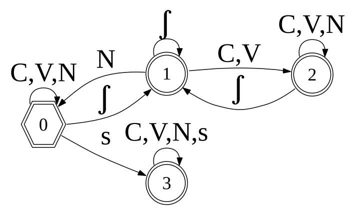
Accepts { s, ʃ, si, ʃi, ss, ʃʃ, sis, ʃiʃ, sns, ʃnʃ, . . . }.

The Finite-State Representation of the LDA Pattern in Navajo

- This grammar recognizes an infinite number of legal words, just like the generative grammars of earlier researchers.
- It does accept words like [tnʃʃʃʃttttttʃiiii]—but this violates other constraints on well-formedness (e.g. syllable structure constraints).
- If the OT analyses of LDA given in Hansson (2001) or Rose and Walker (2004) were written in finite-state terms, this acceptor is exactly the one returned by Riggle's (2004) algorithm.

LDA with Local Blocking: Chumash

1. [ʃ] is never preceded by [s].
2. [s] is never preceded by [ʃ] unless the nearest preceding [ʃ] is immediately followed by [n,t,l].



C = any consonant except [s,ʃ,n,t,l]
V = any vowel
N = [n,t,l]

Accepts { s, ʃ, si, ʃi, ss,
ʃʃ, sis, ʃiʃ, sns, ʃnʃ,
ʃns, ʃnis, ʃniis, . . . } .

The Learning Question in Context

- How can the acceptors above be acquired from finite samples of Navajo and Chumash, respectively?
- The class of patterns describable by finite state acceptors is known to be insufficiently restrictive for learning to occur (Gold 1967, Osherson et. al. 1986).

2.2 Precedence Grammars

Recalling How We Can Describe LDA with No Blocking: Navajo

1. [s,z,ts,ts',dz] are never preceded by [ʃ,ʒ,tʃ,tʃ',dʒ].
2. [ʃ,ʒ,tʃ,tʃ',dʒ] are never preceded by [s,z,ts,ts',dz].

=

[s] can be preceded by [s].
[s] can be preceded by [t].
. . .
[t] can be preceded by [s].
. . .
[ʃ] can be preceded by [ʃ].
[ʃ] can be preceded by [t].
. . .

Precedence Grammars

- A **precedence grammar** is a list of the allowable *precedence* relations in a language.

Languages Recognized by Precedence Grammars

- Words recognized by a precedence grammar are those for which every *precedence relation* is in the grammar.

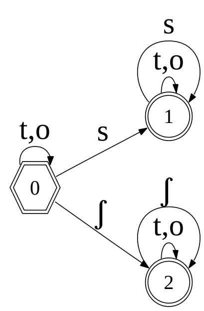
- Example. (Assume $\Sigma = \{\text{s,ʃ,t,o}\}$.)

$$\text{Precedence } G = \left\{ \begin{array}{cccc} \textit{(s,s)} & & \textit{(s,t)} & \textit{(s,o)} \\ & \text{(ʃ,ʃ)} & \text{(ʃ,t)} & \text{(ʃ,o)} \\ \text{(t,s)} & \text{(t,ʃ)} & \text{(t,t)} & \text{(t,o)} \\ \text{(o,s)} & \text{(o,ʃ)} & \textit{(o,t)} & \text{(o,o)} \end{array} \right\}.$$

(1) The Language of G **includes** *sotos*.
(2) The Language of G **excludes** *sotoʃ*.

Precedence Languages are Regular.
These grammars are notational variants.

LDA with No Blocking (e.g. Navajo)



Precedence Grammar

$$G = \left\{ \begin{array}{cccc} \text{(s,s)} & & \text{(s,t)} & \text{(s,o)} \\ & \text{(ʃ,ʃ)} & \text{(ʃ,t)} & \text{(ʃ,o)} \\ \text{(t,s)} & \text{(t,ʃ)} & \text{(t,t)} & \text{(t,o)} \\ \text{(o,s)} & \text{(o,ʃ)} & \text{(o,t)} & \text{(o,o)} \end{array} \right\}.$$

See appendix on how to write a finite-state acceptor given a precedence grammar.

Learning Precedence Grammars

Navajo Fragment. (Assume $\Sigma = \{$s,ʃ,t,o$\}$.)

1.	[s] is never preceded by [ʃ].
2.	[ʃ] is never preceded by [s].

Learning

$$\text{Precedence } G = \left\{ \begin{array}{cccc} \text{(s,s)} & & \text{(s,t)} & \text{(s,o)} \\ & \text{(ʃ,ʃ)} & \text{(ʃ,t)} & \text{(ʃ,o)} \\ \text{(t,s)} & \text{(t,ʃ)} & \text{(t,t)} & \text{(t,o)} \\ \text{(o,s)} & \text{(o,ʃ)} & \text{(o,t)} & \text{(o,o)} \end{array} \right\}.$$

Sample = { tosos , ʃotoʃ , stot }

- The learner has already generalized; it accepts [ʃoʃ], [ʃtot], [sototos]
- but not words like [ʃtos] or [sosoʃ]

Local Summary

- Any LDA with no blocking pattern (e.g. Navajo) can be described with a precedence grammar.
- Any LDA with no blocking pattern can be learned efficiently in the manner described above.

LDA with Local Blocking and Precedence Grammars: Chumash

1.	[ʃ] is never preceded by [s].
2.	[s] is never preceded by [ʃ] unless the nearest preceding [ʃ] is immediately followed by [n,t,l].

- Precedence Grammars as given cannot describe the pattern in Chumash.

*k**ʃ**inot**s**	(hypothetical)
ʃtijepu**s**	'he tells him'

- Next I will show how to extend precedence grammars to capture patterns like those found in Chumash.
  - Bigram Precedence
  - Relative Precedence

Bigram Precedence

- The grammar contains elements of the form (ab,c): "[c] can be preceded by [ab]".
- The idea is that in Chumash (ʃt,s) is in the grammar, but (ʃi,s) is not.

*k**ʃi**not**s**	(hypothetical)
ʃtijepu**s**	'he tells him'

Relative Precedence

- [ab] relatively precedes [c] iff
  (1) [ab] precedes [c] *and*
  (2) no [a] intervenes between [ab] and [c]
- The second conjunct captures the "nearest-preceding" aspect of the Chumash description above.

ʃiʃlusisin 'they (dual) are gone awry'

- [ʃi] *precedes* [s]
- but [ʃi] *does not relatively precede* [s]
- Thus local blocking is achieved by not including (ʃi,s) in the grammar but including (ʃt,s).

Learning Relativized Precedence Bigram Grammars

The learner simply records the relativized precedence bigram relations observed.

$$\text{Precedence } G = \left\{ \begin{array}{cccccc} (\text{ʃi},\text{ʃ}) & & & & & \\ & (\text{iʃ},\text{l}) & (\text{iʃ},\text{u}) & (\text{iʃ},\text{s}) & (\text{iʃ},\text{i}) & (\text{iʃ},\text{n}) \\ & & (\text{ʃl},\text{u}) & (\text{ʃl},\text{s}) & (\text{ʃl},\text{i}) & (\text{ʃl},\text{n}) \\ & & & (\text{lu},\text{s}) & (\text{lu},\text{i}) & (\text{lu},\text{n}) \\ & & & (\text{us},\text{s}) & (\text{us},\text{i}) & (\text{us},\text{n}) \\ & & & (\text{si},\text{s}) & & (\text{si},\text{n}) \\ & & & & (\text{is},\text{i}) & \end{array} \right\}$$

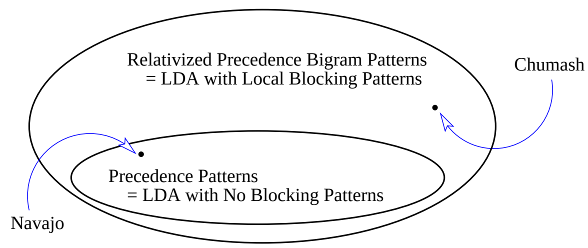
Sample = { ʃiʃlusisin }

- The learner has already generalized: it accepts [ʃiʃ, ʃin, ʃlun, ʃlis, sisisin]
- but not to words like [ʃis, ʃilus].

Local Summary

- Any LDA with local blocking pattern such as the one in Chumash can be described with a Relativized Precedence Bigram Grammar.
- Any pattern describable by a Relativized Precedence Bigram Grammar can be learned efficiently by the algorithm described above.

Relativized Bigram Precedence Patterns include Precedence Patterns

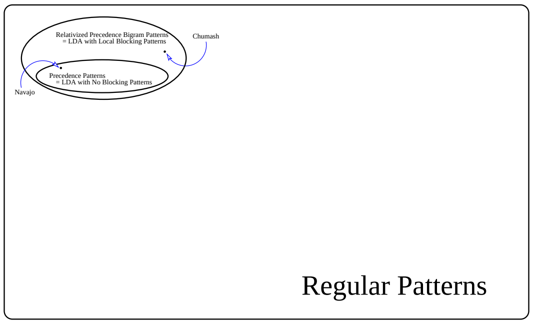


- Any pattern that can be described with a Precedence Grammar can be described with a Relativized Precedence Bigram Grammar.

3 Conclusions

3.1 Summary

Regular Patterns include Relativized Bigram Precedence Patterns



- The class of relativized precedence bigram patterns shown here:
  (1) is a small subset of regular patterns
  (2) includes LDA patterns attested in natural language phonotactics
  (3) is learned simply in the manner described.

Why Learning LDA is Simple

The number of logically possible nonlocal environments increases exponentially with the length of the word.

- Precedence-based learners do not consider every logically possible nonlocal environment. They cannot learn logically possible nonlocal patterns like:

  (1) If the third segment after a sibilant is a sibilant, they must agree in [anterior].

  (2) If the second, third, or fifth segments after a sibilant is a sibilant, they must agree in [anterior].

  (3) and so on

- These learners do not distinguish on the basis of distance at all.

- The notion of "arbitrarily many"—not being able to count—*is sufficiently restrictive* for learning to occur.

Summary

- A learner can keep track of *precedence* relations to learn attested Long Distance Agreement patterns.

- This algorithm is properly thought of as one part of a complete phonotactic learner—the part which returns LDA-type constraints.

- Factoring the learning problem is a useful way to address how phonological learning can occur.

3.2 Remaining Questions

Learning Gradient Phonotactics

(1) Phonotactic patterns are gradient; this is categorical.

- Categorical patterns are a special case of gradient ones.

- Nothing in the design on the model depends on its categorical nature.

- There are many ways to make the model gradient.

Learning with a Noisy Sample

(2) Can Precedence Learning occur in the presence of noise?

a. What if certain precedence relations are not in the sample?

b. What if there are just a few exceptions to the constraint?

- We know what the categorical model does. These are really empirical questions. What do people do in these cases?
- *Yes*, it can be shown that under certain noisy conditions, with high confidence, the patterns describable with precedence grammars can be learned allowing for a certain amount of error (Angluin and Laird 1988, Valiant 1984, Kearns and Vazirani 1994) .

Learning Phonetically Unmotivated LDA Patterns.

(3) Precedence Learning can learn 'unmotivated' LDA patterns. E.g. "[b] never precedes [ʒ]."

- Independently motivated restrictions can be built into this grammar to further restrict the hypothesis space.
  - Similarity restrictions on potential agree-ers (Hansson 2001, Rose and Walker 2004) (See also Frisch et. al. 2004)
  - Relevency Conditions on interveners (Jensen 1974) (See also Odden 1994).

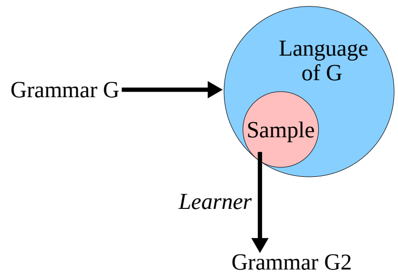
Some LDA patterns do measure distance

(4) Some LDA patterns do measure distance.

  - In Ndonga, nasal agreement fails if the nasals are not within adjacent syllables (Viljoen 1973, Rose and Walker 2004).
  - See also Martin (2004) regarding the domain of sibilant harmony in Navajo compounds.

- The distance appears to be measured in syllables, not segments.
- How this can be incorporated into the present learner I leave unresolved for future investigation.

Acknowledgements



Thank You.

Thanks to Bruce Hayes, Ed Stabler, Colin Wilson, and Kie Zuraw for invaluable discussion related to this material. I also thank Greg Kobele, Andy Martin, Katya Pertsova, Shabnam Shademan, and Sarah VanWagnenen for their insightful and helpful comments.

Appendix. FSA Construction from Precedence Grammars: Basic Definitions

This appendix shows how a finite state acceptor can be written which is equivalent to any precedence grammar. The relativized precedence bigram grammars work similarly.

- Σ denotes a fixed finite set of symbols, the *alphabet*. Σ^* denotes all finite strings formed over this alphabet.
- A *language* is a subset of Σ^*.
- Let uv denote the concatenation of two strings u and v.
- The prefixes of a language L are defined below.

$$Pr(L) = \{u \in \Sigma^* : \exists v \in \Sigma^* \text{ such that } uv \in L\}$$

Appendix. FSA Construction from Precedence Grammars: Definitions

- To facilitate FSA construction, we augment the symbols in these grammars with the word boundary symbol #. We write $V = \Sigma \cup \{\#\}$.
- A *precedence grammar* G is a subset of V^2.
- For some precedence grammar G, the Language of G is given below

$$L(G) = \{w \in \Sigma^* : PS(\#w\#) \subseteq G\}$$

  where $PS(x)$ is the set of precedence relations of string x.

Appendix. FSA Construction of Precedence Grammars

- A finite state acceptor is a tuple (Q, I, F, δ) where Q is the set of states, I is the set of initial states (a subset of Q), F is the set of final states (a subset of Q), and δ is the transition function, which maps a (state,symbol) pair to a set of states.
- For any precedence grammar G, an FSA which accepts exactly $L(G)$ may be written as follows:

$$\begin{array}{rcl} Q & = & \{PS(\#u\#) : u \in Pr(L(G))\} \\ I & = & \{\{(\#, \#)\}\}\text{iff } L(G) \neq \emptyset\text{, otherwise } \emptyset \\ F & = & Q \\ \delta(PS(\#u\#), a) & = & PS(\#ua\#) \text{ whenever } u, ua \in Pr(L(G)) \end{array}$$

- This acceptor is deterministic so it is possible to compute the canonical acceptor using standard minimization algorithms (Hopcroft et. al. 2001).

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. 2003a. Learning NonLocal Environments. Talk Handout of 77th Annual Meeting of LSA, Atlanta Georgia.

Albright, Adam and Bruce Hayes. 2003b. 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. 1980. Inductive inference of formal languages from positive data. *Information Control* 45:117–135.

Angluin, Dana and Philip Laird. 1988. Learning from Noisy Examples. *Machine Learning* 2:343–370.

Applegate, R.B. 1972. Ineseño Chumash Grammar. Ph.D. thesis, University of California, Berkeley.

Baković, Eric. 2000. Harmony, Dominance and Control. Ph.D. thesis, Rutgers University.

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.

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.

Fountain, Amy. 1998. An Optimality Theoretic Account of Navajo Prefixal Syllables. Ph.D. thesis, University of Arizona.

Frank, Robert and Giorgo Satta. 1998. Optimality Theory and the Generative Complexity of Constraint Violability. *Computational Linguistics* 24(2):307–315.

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.

Gafos, Adamantios. 1999. *The Articulatory Basis of Locality in Phonology*. New York, Garland.

Gick, Bryan, Douglas Pulleyblank, Fiona Campbell, and Ngessimo Mutaka. 2006. Kinande Vowel Harmony. *Phonology* 23(1):1–20.

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

Gold, E.M. 1967. Language Identification in the Limit. *Information and Control* 10:447–474.

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. 1999. The "neutral" vowels of Finnish: How neutral are they? *Linguistica Uralica* 35:17–21.

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.

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. To appear. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. 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.

Heinz, Jeffrey. To appear. Learning Unbounded Stress Systems via Local Inference. In *Proceedings of the 37th Meeting of the Northeast Linguistics Society*. University of Illionois, Urbana-Champaign.

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

Jensen, John. 1974. Variables in Phonolgy. *Language* 50(4):675–686.

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.

Kearns, Michael and Umesh Vazirani. 1994. *An Introduction to Computational Learning Theory*. MIT Press.

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

Martin, Andrew. 2004. The effects of distance on lexical bias: sibilant harmony in Navajo compounds. Masters Thesis. University of California, Los Angeles.

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

Niyogi, Partha. 2006. *The Computational Nature of Language Learning and Evolution*. The MIT Press.

Nowak, Martin A., Natalia L. Komarova, and Partha Niyogi. 2002. Computational and evolutionary aspects of language. *Nature* 417:611–617.

Odden, David. 1994. Adjacency Parameters in Phonology. *Language* 70(2):289–330.

Osherson, Daniel, Scott Weinstein, and Michael Stob. 1986. *Systems that Learn*. MIT Press, Cambridge, Massachusetts.

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.

Piattelli-Palmarini, Massimo, ed. 1980. *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. Harvard University Press.

Poser, William. 1982. Phonological Representation and Action-at-a-Distance. In *The Structure of Phonological Representations*, edited by H. van der Hulst and N.R. Smith. Dordrecht: Foris, pages 121–158.

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.

Recasens, D., J.M. Sole, and J. Romero, eds. 2003. *On Neutral Vowels in Hungarian*. Universitat Autonoma de Barcelona, Spain.

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.

Sapir, Edward and Harry Hojier. 1967. The Phonology and Morphology of the Navaho Language. *University of California Publications* 50.

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.

Valiant, L.G. 1984. A theory of the learnable. *Communications of the ACM* 27:1134–1142.

Viljoen, J.J. 1973. *Manual for Ndonga*. Pretoria: University of South Africa. Part 1.

Walker, Rachel. 1998. Nasalization, Neutral Segments, and Opacity Effects. Ph.D. thesis, University of California, Santa Cruz.

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.