Learning Long Distance Phonotactics

Jeffrey Heinz
heinz@udel.edu

University of Delaware

University of Chicago Workshop
June 12, 2008

Introduction

- I present a learner which learns the attested long distance phonotactic patterns in the world's languages
- This learner
  1. keeps track of the order of sounds—but not the distance between them (precedence relations)
  2. fails to learn logically possible—but unattested—long distance phonotactics
- The conclusion is if humans generalize in the way suggested by the model, it can explain features of the typology of long distance phonotactics (cf. Moreton 2008, *analytic bias*)

Outline

1. Introduction
   - Long Distance Phonotactics
   - Representing Long Distance Phonotactics
2. Precedence-based Learning
   - Learning in Phonology
   - Precedence Grammars
3. Conclusion
   - Issues
   - Summary

Outline

What is long distance phonotactics (LDP)?

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).

Symmetric LDA: Navajo (Athabaskan)

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

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

Examples (Sapir and Hojier 1967):

1.	ʃiːteːʒ	‘we (dual) are lying’
2.	dasdoːlis	‘he (4th) has his foot raised’
3.	*ʃiːteːz	(hypothetical)
4.	*dasdoːliʃ	(hypothetical)

Asymmetric LDA: Sarcee (Athabaskan)

In well-formed words, sibilants agree in the feature [anterior], but only the [-anterior] sibilants are ‘active’.

1.	[s,z,ts,dz] never precedes [ʃ,ʒ,tʃ,dʒ].

Examples (Hansson 2001, citing Cook 1979,1984):

1.	ʃítʃídzàʔ	‘my duck’
2.	nāʃɣátʃ	‘I killed them again’
3.	*zítʃídzàʔ	(hypothetical)
4.	*snāʃɣátʃ	(hypothetical)

Examples of long distance phonotactics

- Consonantal Harmony (Hansson 2001, Rose and Walker 2004)
  - Sibilant, liquid, dorsal, voicing, ... harmony and disharmony
  - Symmetric/Asymmetric LDA
  - $\sim$120 languages documented with consonantal harmony (Hansson 2001).
- possibly Vowel Harmony with 'transparent' vowels
  - Finnish, Hungarian, Nez Perce (see Baković 2000 and references therein)
  - Some controversy over how transparent: see Gordon (1999), Gafos and Benus (2003), and Gick et. al. (2006).

One debate, two puzzles

- Debate: Is it really non-local?
- Puzzles
  - How do we explain the absence of blocking in the typology?
  - (if it non-local) How such non-local patterns learned?

Debate: Is LDP really spreading?

- Spreading means the intervening segments are affected.
- Nasal spreading in Malay (Johore dialect, Walker 1999, citing Onn 1980)

  1. mə̃nãw̃ãn ‘to capture’ (active)
  2. pəŋãw̃ãsan ‘supervision’

- Navajo’ as spreading (+/- indicates [anterior])

  3. ʃi̠ːte̠ːʒ ‘we (dual) are lying’
  4. dasdoːlis ‘he (4th) has his foot raised’
     + + ++

- Gafos (1999) argues that Navajo=Navajo’ (see Hansson 2001 for counterarguments).

Debate: Is LDP really spreading?

- Spreading means the intervening segments are affected.
- Nasal spreading in Malay (Johore dialect, Walker 1999, citing Onn 1980)

1. mə̃nã w̃ãn ‘to capture’ (active)
2. pəŋã w̃ãsan ‘supervision’

- Navajo’ as spreading (+/- indicates [anterior])

3. ʃiːteːʒ ‘we (dual) are lying’
4. dasdoːlis ‘he (4th) has his foot raised’

- Gafos (1999) argues that Navajo=Navajo’ (see Hansson 2001 for counterarguments).

Puzzle # 1: Explaining the typology of LDP

- The typology of LDA is notable in two respects (Hansson 2001, Rose and Walker 2004):
  1. (1) LDA holds between similar segments.
  2. (2) Blocking patterns are absent.
- The latter helps distinguish LDA from spreading.

LDP with Blocking: Hypothetical

In well formed words, voiceless sibilants agree in the feature [anterior] unless, between two voiceless sibilants which disagree in [anterior], there is a voiced sibilant (and no other voiceless sibilants).

1. [ʃ] never precedes [s] unless, for each [ʃ], a [z] or [ʒ] occurs between [ʃ] and its nearest following [s]
2. [s] never precedes [ʃ] unless, for each [s], a [z] or [ʒ] occurs between [s] and its nearest following [ʃ]

Examples (all hypothetical since no language example exists!):

1.	ʃotoʃ	3.	ʃozos	5.	*ʃotos	7.	*ʃosozos
2.	sotos	4.	sosozoʃ	6.	*sotoʃ	8.	*soʃosozos

LDP with Blocking is Unattested

- The absence of this type of LDP is robust!
- Consenus has formed in the few proposed counterexamples (Sanskrit, Kinyarwanda) that they are better analyzed as spreading (Schein and Steriade 1986, Mpiranya and Walker 2005).

Current Proposal Explaining The Typology of LDP

- Rose and Walker (2004) take both gaps as systematic.
- Their Agreement By Correspondence (ABC) analysis of LDA in OT uses:
    - CC-Correspondance constraints: two consonants are in correspondence if they are sufficiently similar (agnostic about similarity metric)
    - ID-CC(FEATURE) constraints which enforce agreement of FEATURE for corresponding consonants.
- This is intended to capture both the similarity and blocking effects.

But it fails. . . hence Puzzle #1

- Hansson (2007) studies the predicted typology of ABC and shows the ABC approach does predict non-local blocking effects of certain types.
- . . . reluctantly suggests that the absence of blocking patterns is accidental.

But it fails. . . hence Puzzle #1

- Hansson (2007) studies the predicted typology of ABC and shows the ABC approach does predict non-local blocking effects of certain types.
- . . . reluctantly suggests that the absence of blocking patterns is accidental.

Current theory doesn't explain the absence of blocking in the typology of LDP

Puzzle # 2: Learning LDP

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 LDP patterns.

The Meaning of “arbitrarily many”

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

Outline

1. Introduction
   - Long Distance Phonotactics
   - Representing Long Distance Phonotactics
2. Precedence-based Learning
   - Learning in Phonology
   - Precedence Grammars
3. Conclusion
   - Issues
   - Summary

Phonotactics as sets

- The possible words of English can be thought of a set which includes:

  { slam, fist, blick, flump, . . . }

- and which excludes:

  { sram, fizt, bnick, flumk, . . . }

Phonotactics as sets

- The binary, categorical distinction between 'well-formed' and 'ill-formed' is a convenient abstraction.

kɪp > θwiːks > bzɑrʃk

(Coleman and Pierrehumbert 1997, Frisch, Pierrehumbert and Cole 2004, Albright and Hayes 2003, Kirby and Yu 2007, Hayes and Wilson 2008)

What kind of sets are long distance phonotactic sets?

word	Navajo	Sarcee	Hypothetical
to	✓	✓	✓
sotos	✓	✓	✓
ʃotoʃ	✓	✓	✓
ʃotos	×	✓	×
sotoʃ	×	×	×
ʃozos	×	✓	✓
sozoʃ	×	×	✓
soʃozoʃ	×	×	×
…			

What kind of sets are long distance phonotactics?

- Long distance phonotactic patterns are *regular*.

  [Johnson(1972), Kaplan and Kay(1981), Kaplan and Kay(1994), Ellison(1992), Eisner(1997), Albro(1998), Albro(2005), Karttunen(1998b), Frank and Satta(1998), Riggle(2004), Karttunen(2006)]
- Regular sets have many characterizations (see e.g. Kracht 2003). They are those sets describable with:
  - finite state acceptors
  - right-branching rewrite grammars
  - regular expressions
  - monadic second order logic

Finite State Acceptors

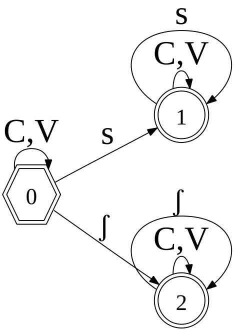
FSAs

(1) can be related to finite state OT and rule-based models, which allow us to compute a phonotactic finite-state acceptor (Johnson 1972, Kaplan and Kay 1994, Karttunnen 1998, Riggle 2004), which becomes the target grammar for the learner.

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

Symmetric LDP: Navajo

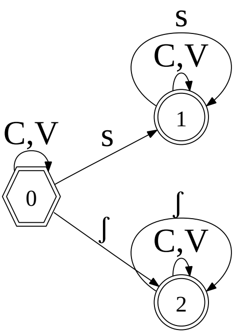
1. [s,z,ts,ts',dz] never precedes [ʃ,ʒ,tʃ,tʃ',dʒ].
2. [ʃ,ʒ,tʃ,tʃ',dʒ] never precedes [s,z,ts,ts',dz].



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

Accepts	Rejects
sos	soʃ
ʃoʃ	ʃos
sots	ʃtos
ʃotoʃ	...
...	

The FSA Representation of Navajo Sibilant Harmony

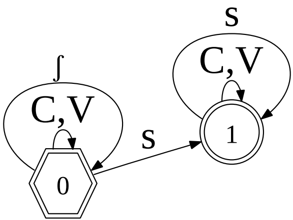


- This grammar recognizes an infinite number of legal words, just like the generative grammars of earlier researchers.
- It does accept words like [tnʃʃʃʃtttttʃiiii]—but this violates other constraints on well-formedness (e.g. syllable structure constraints).

- 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 Karttunen's (1998) and Riggle's (2004) algorithms.

Asymmetric LDP: Sarcee

1. [s,z,ts,dz] never precedes [ʃ,ʒ,tʃ,dʒ].

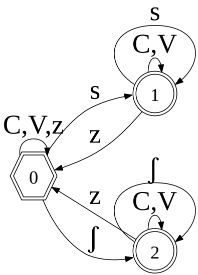


C = any consonant except sibilants
s = [+anterior] sibilants
V = any vowel
ʃ = [-anterior] sibilants

Accepts	Rejects
sos	soʃ
ʃoʃ	ʃosoʃ
ʃots	stoʃ
ʃoʃos	. . .
. . .	

LDP with Blocking: Hypothetical

1. [ʃ] never precedes [s] unless, for each [ʃ], a [z] or [ʒ] occurs between [ʃ] and its nearest following [s]
2. [s] never precedes [ʃ] unless, for each [s], a [z] or [ʒ] occurs between [s] and its nearest following [ʃ]



C = any consonant except sibilants
s = [+anterior] voiceless sibilants
V = any vowel
ʃ = [-anterior] voiceless sibilants
z = any voiced sibilant

Accepts	Rejects
sos	soʃ
ʃoʃ	ʃos
ʃotozotos	ʃtozoʃos
. . .	. . .

Outline

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)Alderete, Brasoveanua, Merchant, Prince, and Tesar, Merchant and Tesar(2006), Wilson(2006), Riggle(2006), Tessier(2006)]

- Learning in Principles and Parameters

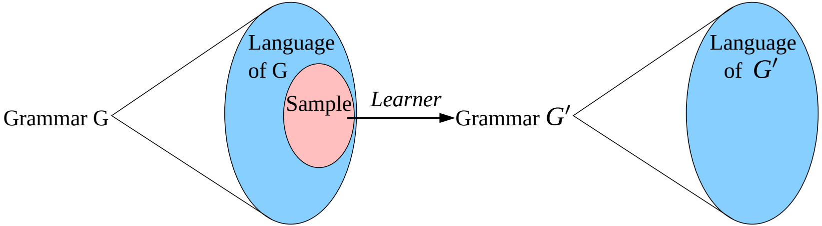
  [Wexler and Culicover(1980), Dresher and Kaye(1990), Niyogi(2006), Pearl(2007)]

- Learning Phonological Rules

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

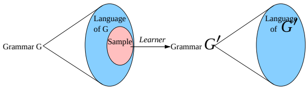
- Learning Phonotactics [Ellison(1992), Goldsmith(1994), Frisch(1996), Coleman and Pierrehumbert(1997), Frisch et al.(2004)Frisch, Pierrehumbert, and Broe, Albright(2006), Goldsmith and Xanthos(2006), Heinz(2007), Hayes and Wilson(2008), Goldsmith and Riggle(submitted)]

Identification in the Limit from Positive Data (Gold 1967)



- What is *Learner* so that Language of G' = 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.
- G' 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 LDP patterns are different from patterns which restrict the distribution of adjacent, contiguous segments.

Factoring the Phonotactic Learning Problem

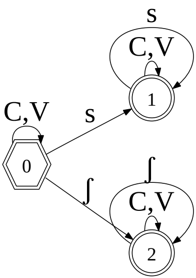
- Different kinds of phonotactic constraints can be learned by different learning algorithms.
- A complete phonotactic learner is a combination of these different learning algorithms
- Here, I am only showing how one part of the whole learner—the part that learns long-distance constraints—can work.

Some concerns regarding identification in the limit from positive data

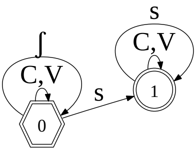
- No noise in input
- No requirement for learner to be efficient
- No requirement on ‘small’ sample to succeed
- Exact identification is too strict a criterion

The Learning Question in Context

Symmetric LDP

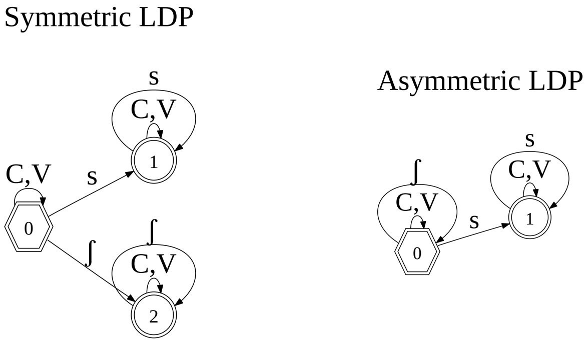


Asymmetric LDP



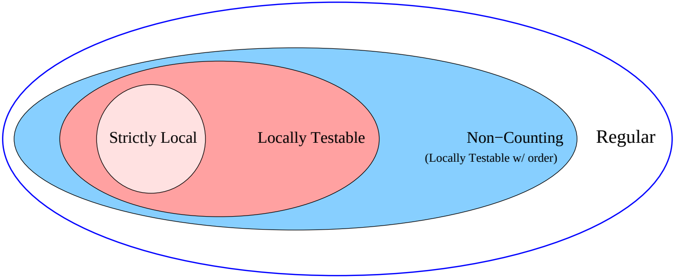
- What learner can acquire the machines above from finite samples of Navajo, Sarcee, respectively?
- This question is not easy. There is no simple ‘fix’.
- The class of regular sets is known to be insufficiently restrictive for learning to occur!
  (Gold 1967, Osherson et. al. 1986, Jain et. al. 1999).

The Learning Question in Context



- What learner can acquire the machines above from finite samples of Navajo, Sarcee, respectively?
- This question is not easy. There is no simple 'fix'.
- The class of regular sets is known to be insufficiently restrictive for learning to occur!
  (Gold 1967, Osherson et. al. 1986, Jain et. al. 1999).

The Sub-regular Hierarchy
(McNaughton and Papert 1971, Pullum and Rogers 2007)



- Some subclasses of the regular languages are sufficiently restrictive for learning to occur
  - Locally *k*-testable in the strict sense (Strictly Local)
  - Locally *k*-testable
  - Many others from grammatical inference community
    Angluin(1982), Garcia et al.(1990), Muggleton(1990), Denis et al.(2002), Fernau(2003)

The Sub-regular Hierarchy (McNaughton and Papert 1971, Pullum and Rogers 2007)

Locally 2-testable in the strict sense (Strictly Local)

- $sotos \in L$ iff $\{so, ot, to, os\} \subseteq G_L$
- E.g. bigrams

The Sub-regular Hierarchy (McNaughton and Papert 1971, Pullum and Rogers 2007)

Locally 2-testable in the strict sense (Strictly Local)

- $sotos \in L$ iff $\{so, ot, to, os\} \subseteq G_L$
- E.g. bigrams

Locally 2-testable

- $sotos \in L$ iff $\{so, ot, to, os\} \in G_L$
- E.g sets of bigrams

The Sub-regular Hierarchy (McNaughton and Papert 1971, Pullum and Rogers 2007)

Locally 2-testable in the strict sense (Strictly Local)

- $sotos \in L$ iff $\{so, ot, to, os\} \subseteq G_L$
- E.g. bigrams

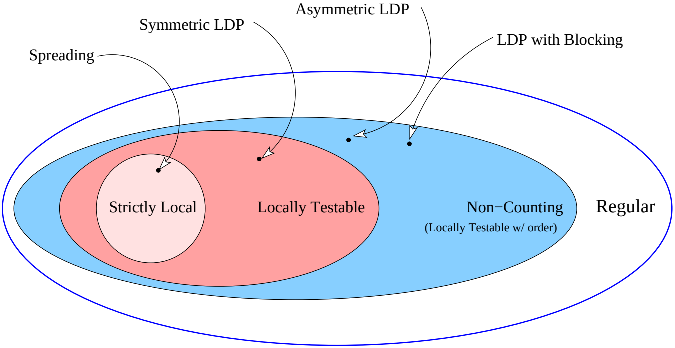
Locally 2-testable

- $sotos \in L$ iff $\{so, ot, to, os\} \in G_L$
- E.g sets of bigrams

Noncounting

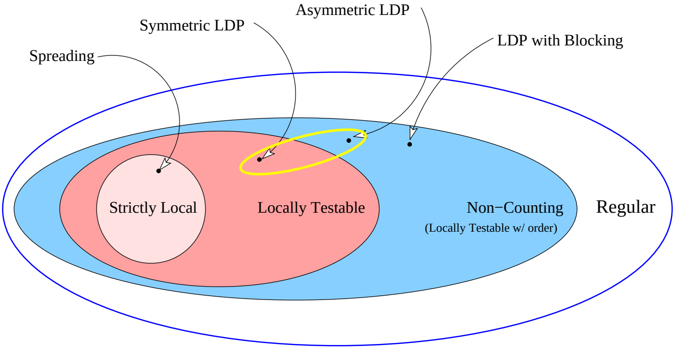
- there is some $n > 0$, for all $uv^nw \in L$ iff $uv^{n+1}w \in L$ for all strings $u, v, w \in \Sigma^*$.
- E.g. closure of Locally Testable class under concatenation and boolean operations.

The Sub-regular Hierarchy
(McNaughton and Papert 1971, Pullum and Rogers 2007)



- Spreading is locally 2-testable in the strict sense
- Symmetric LDP is locally 1-testable
- Asymmetric LDP and Hypothetical are noncounting

The Sub-regular Hierarchy (McNaughton and Papert 1971, Pullum and Rogers 2007)



- The goal!

Outline

Discontiguous Ordered Strings

Order matters, but not distance.

- Whitney and Berndt(1999), Whitney(2001), Schoonbaert and Grainger(2004), and Grainger and Whitney(2004) use discontiguous ordered strings of length two in a model for reading comprehension
- Shawe-Taylor and Christianini (2005, chap. 11) also discuss kernels defined over discontigouous, ordered strings for use in text classification

Recalling How We Can Describe Symmetric LDP: Navajo

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

Recalling How We Can Describe Symmetric LDP: Navajo

1.	[s,z,ts,ts’,dz] never precedes [ʃ,ʒ,tʃ,tʃ’,dʒ].
2.	[ʃ,ʒ,tʃ,tʃ’,dʒ] never precedes [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 $\Sigma = \{\text{s,ʃ,t,o}\}$.)

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

(1) The Language of G includes *sotos*.

Languages Recognized by Precedence Grammars

- Words recognized by a precedence grammar are those for which every precedence relation is in the grammar.
- Example. (Assume $\Sigma = \{s, ʃ, t, o\}$.)

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

(1) The Language of G includes *sotos*.

Languages Recognized by Precedence Grammars

- Words recognized by a precedence grammar are those for which every precedence relation is in the grammar.
- Example. (Assume $\Sigma = \{\text{s,ʃ,t,o}\}$.)

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

(1) The Language of G includes *sotos*.

Languages Recognized by Precedence Grammars

- Words recognized by a precedence grammar are those for which every precedence relation is in the grammar.
- Example. (Assume $\Sigma = \{\text{s,ʃ,t,o}\}$.)

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

(1) The Language of G includes *sotos*.

Languages Recognized by Precedence Grammars

- Words recognized by a precedence grammar are those for which every precedence relation is in the grammar.
- Example. (Assume $\Sigma = \{\text{s,ʃ,t,o}\}$.)

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

(1) The Language of G includes *sotos*.

Languages Recognized by Precedence Grammars

- Words recognized by a precedence grammar are those for which every precedence relation is in the grammar.
- Example. (Assume $\Sigma = \{s, ʃ, t, o\}$.)

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

(1) The Language of G includes *sotos*.

Languages Recognized by Precedence Grammars

- Words recognized by a precedence grammar are those for which every precedence relation is in the grammar.
- Example. (Assume $\Sigma = \{\text{s},\int,\text{t},\text{o}\}$.)

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

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

Languages Recognized by Precedence Grammars

- Words recognized by a precedence grammar are those for which every precedence relation is in the grammar.
- Example. (Assume $\Sigma = \{s,\int,t,o\}$.)

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

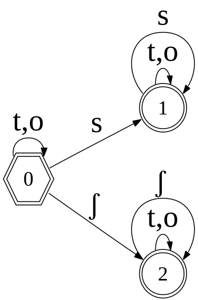
(1) The Language of G includes *sotos*.

(2) The Language of G excludes *soto∫*.

Precedence Languages are Regular.

These grammars are notational variants.

Symmetric LDP (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 Heinz (2007) on how to write a finite-state acceptor given a precedence grammar.

Learning Precedence Grammars: Navajo Fragment

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

1.	[s] never precedes [ʃ].
2.	[ʃ] never precedes [s].

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

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

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

1.	[s] never precedes [ʃ].
2.	[ʃ] never precedes [s].

Learning

Precedence $G = \left\{ \quad \right\}$.

Sample = { }

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

Learning Precedence Grammars: Navajo Fragment

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

1.	[s] never precedes [ʃ].
2.	[ʃ] never precedes [s].

Learning

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

Sample = { tosos }

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

Learning Precedence Grammars: Navajo Fragment

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

1.	[s] never precedes [ʃ].
2.	[ʃ] never precedes [s].

Learning

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

Sample = { tosos , ʃotoʃ }

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

Learning Precedence Grammars: Navajo Fragment

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

1.	[s] never precedes [ʃ].
2.	[ʃ] never precedes [s].

Learning

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

Sample = { tosos , ʃotoʃ , stot }

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

Learning Precedence Grammars: Navajo Fragment

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

1.	[s] never precedes [ʃ].
2.	[ʃ] never precedes [s].

Learning

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

Sample = { tosos , ʃotoʃ , stot }

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

Learning Precedence Grammars: Navajo Fragment

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

1.	[s] never precedes [ʃ].
2.	[ʃ] never precedes [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]

Learning Precedence Grammars: Sarcee Fragment

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

1.	[s] never precedes [ʃ].

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

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

Learning Precedence Grammars: Sarcee Fragment

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

1. [s] never precedes [ʃ].

Learning

Precedence $G = \left\{ \quad \right\}$.

Sample = { }

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

Learning Precedence Grammars: Sarcee Fragment

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

1.	[s] never precedes [ʃ].

Learning

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

Sample = { tosos }

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

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

1. [s] never precedes [ʃ].

Learning

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

Sample = { tosos , ʃotoʃ }

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

Learning Precedence Grammars: Sarcee Fragment

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

1.	[s] never precedes [ʃ].

Learning

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

Sample = { tosos , ʃotoʃ , ʃtots }

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

Learning Precedence Grammars: Sarcee Fragment

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

1.	[s] never precedes [ʃ].

Learning

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

Sample = { tosos , ʃotoʃ , ʃtots }

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

Learning Precedence Grammars: Sarcee Fragment

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

1.	[s] never precedes [ʃ].

Learning

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

Sample = { tosos , ʃotoʃ , ʃtots }

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

Local Summary

- Any symmetric or asymmetric LDP pattern (e.g. Navajo and Sarcee) can be described with a precedence grammar.
- Any symmetric or asymmetric LDP pattern can be learned efficiently in the manner described above.

Why Learning LDP 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

Locality and LDP

- Precedence-based learners do not distinguish on the basis of distance at all.
- In one sense, every segment is adjacent to every preceding segment.
- The notion of “arbitrarily many may intervene”—not being able to count distance, while keeping track of order—is sufficiently restrictive for learning to occur.

Locality and LDP

- Precedence-based learners do not distinguish on the basis of distance at all.
- In one sense, every segment is adjacent to every preceding segment.
- The notion of "arbitrarily many may intervene"—not being able to count distance, while keeping track of order—is sufficiently restrictive for learning to occur.

The Precedence Learner cannot learn LDP with blocking

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

1.	[ʃ] never precedes [s] unless, for each [ʃ], a [z] or [ʒ] occurs between [ʃ] and its nearest following [s]
2.	[s] never precedes [ʃ] unless, for each [s], a [z] or [ʒ] occurs between [s] and its nearest following [ʃ]

Learning

Precedence $G = \left\{ \quad \right\}$.

Sample = { }

- The learner has failed to learn Hypothetical! E.g. it accepts [ʃos].

The Precedence Learner cannot learn LDP with blocking

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

1. [ʃ] never precedes [s] unless, for each [ʃ], a [z] or [ʒ] occurs between [ʃ] and its nearest following [s]
2. [s] never precedes [ʃ] unless, for each [s], a [z] or [ʒ] occurs between [s] and its nearest following [ʃ]

Learning

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

Sample = { tosos }

- The learner has failed to learn Hypothetical! E.g. it accepts [ʃos].

The Precedence Learner cannot learn LDP with blocking

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

1. [ʃ] never precedes [s] unless, for each [ʃ], a [z] or [ʒ] occurs between [ʃ] and its nearest following [s]
2. [s] never precedes [ʃ] unless, for each [s], a [z] or [ʒ] occurs between [s] and its nearest following [ʃ]

Learning

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

Sample = { tosos , ʃotoʃ }

- The learner has failed to learn Hypothetical! E.g. it accepts [ʃos].

The Precedence Learner cannot learn LDP with blocking

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

1. [ʃ] never precedes [s] unless, for each [ʃ], a [z] or [ʒ] occurs between [ʃ] and its nearest following [s]
2. [s] never precedes [ʃ] unless, for each [s], a [z] or [ʒ] occurs between [s] and its nearest following [ʃ]

Learning

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

Sample = { tosos , ʃotoʃ , ʃozos }

- The learner has failed to learn Hypothetical! E.g. it accepts [ʃos].

The Precedence Learner cannot learn LDP with blocking

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

1. [ʃ] never precedes [s] unless, for each [ʃ], a [z] or [ʒ] occurs between [ʃ] and its nearest following [s]
2. [s] never precedes [ʃ] unless, for each [s], a [z] or [ʒ] occurs between [s] and its nearest following [ʃ]

Learning

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

Sample = { tosos , ʃotoʃ , ʃozos }

- The learner has failed to learn Hypothetical! E.g. it accepts [ʃos].

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

1. [ʃ] never precedes [s] unless, for each [ʃ], a [z] or [ʒ] occurs between [ʃ] and its nearest following [s]
2. [s] never precedes [ʃ] unless, for each [s], a [z] or [ʒ] occurs between [s] and its nearest following [ʃ]

Learning

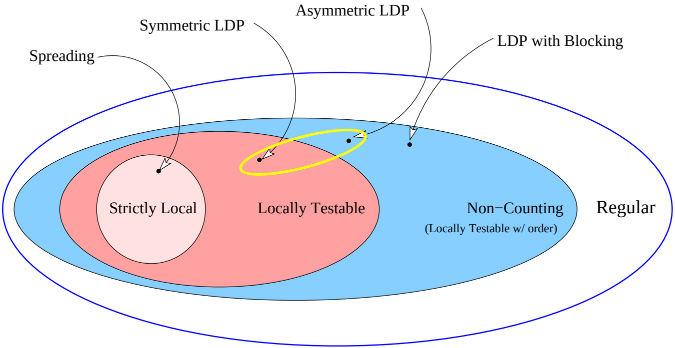
Precedence $G =$

(s,s)		(s,t)	(s,o)	
(ʃ,s)	(ʃ,ʃ)	(ʃ,t)	(ʃ,o)	(ʃ,z)
(t,s)	(t,ʃ)	(t,t)	(t,o)	
(o,s)	(o,ʃ)	(o,t)	(o,o)	(o,z)
(z,s)			(z,o)	

Sample = { tosos , ʃotoʃ , ʃozos }

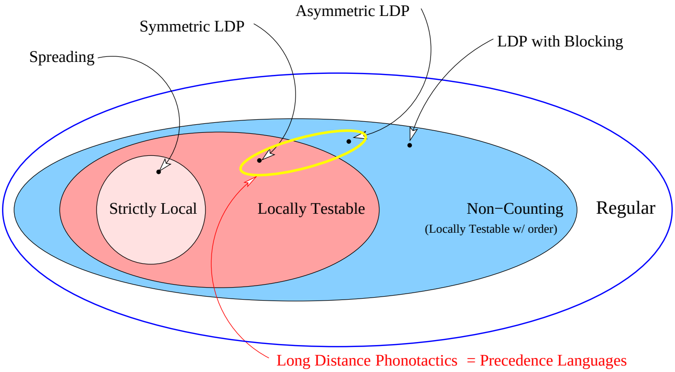
- The learner has failed to learn Hypothetical! E.g. it accepts [ʃos].

Main Conclusion



- If humans generalize in the way suggested by the precedence learner, it explains why
  (1) there are long-distance phonotactic patterns
  (2) there are no long-distance phonotactic with blocking patterns

Main Conclusion



- If humans generalize in the way suggested by the precedence learner, it explains why
  (1) there are long-distance phonotactic patterns
  (2) there are no long-distance phonotactic with blocking patterns

Outline

Why not just use *n*-grams over tiers?

(1) Phonologists often employ tiers, also called projections, in their analyses of long distance phenomenon (Goldsmith 1976, 1990, Prince 1984, Hayes and Wilson 2008, Goldsmith and Riggle, under review).

- E.g. vowel tiers, consonant tiers, sibilant tiers

ʃ				ʒ
↑				↑
ʃ	iː	t	eː	ʒ

‘we (dual) are lying’

Bigram learning over tiers learns LDP with Blocking

Consider a word from Hypothetical.

s				z		ʃ	
↑				↑		↑	
s	o	t	o	z	o	ʃ	(hypothetical)

- Maybe only project voiceless sibilants in this case?
- What is the theory of tiers? Cf.
  - Rose and Walker's agnosticism about what is appropriate similarity metric
  - Hayes and Wilson's antecedently given tiers
- but see also Goldsmith and Xanthos (2006)

Learning Gradient Phonotactics

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

- Nothing in the design on the model depends on its categorical nature.
- There are many ways to make the model gradient:
  - minimum distance length (Ellison 1994), Bayes law (Tenenbaum 1999, Goldwater 2006), maximum entropy (Goldwater and Johnson 2003, Hayes and Wilson 2008), kernel methods (Shawe-Taylor and Christianini 2005), and approaches inspired by Darwinian-like processes (Clark 1992, Yang 2000)

Learning Gradient Phonotactics

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

precedes	s	ʃ	t	o
s	0.01		...	
ʃ		0.01		
t	⋮		⋱	
o				

  - Compute cells by calculating the joint probability over precedence relations
  - Compute cells by calculating conditional probability of a segment (given all preceding segments)
- evaluate utility of precedence model with MDL (Goldsmith and Riggle, under review)

Learning with a Noisy Sample

(3) 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?

- Angluin and Laird (1988) show that there are classes of languages which, under certain noisy conditions, which can be "probably approximately correctly" learned (Valiant 1984, Kearns and Vazirani 1994).
- Precedence languages are such a class.
- It remains to be seen exactly what the precedence learner which handles noise looks like.

Learning Phonetically Unmotivated LDP Patterns.

(4) Precedence Learning can learn ‘unmotivated’ LDP patterns.
E.g. “[b] never precedes [ʒ].”

- What do people do?
- 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).
- Use the independently motivated theory of similarity to set Bayesian priors over the precedence-based hypothesis space

This independence is a good thing

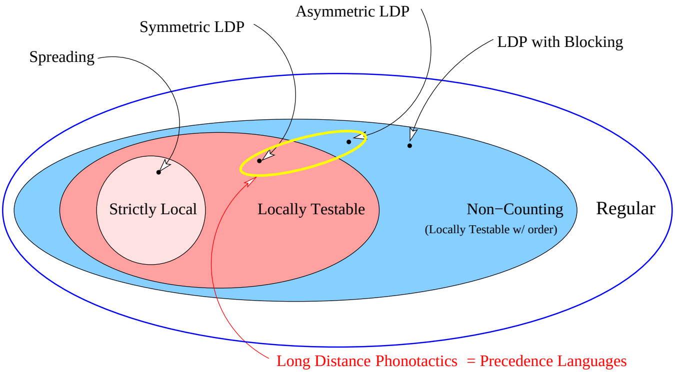
- Other models require independently motivated theory of similarity (OT-CC, tiers)
- Here, such a theory is not needed for learning
- Allows us to study these factors independently
- What is the contribution of sound similarity to learning phonological patterns?

Outline

Summary

- A learner which keeps track of order—and not distance— (i.e. precedence relations) learns attested long distance phonotactics, and explains a key feature of the typology—absence of blocking.
- This helps explain why LDP is distinct from spreading.
- We ought to investigate
  - How successful as grammars w.r.t MDL
  - How to integrate similarity
  - Whether predictions are confirmed by language acquisition studies

Acknowledgements



Thank You.

Thanks to Jim Rogers and Ed Stabler for helpful discussion. I also thank audiences of the U. of Delaware's Lingustics Spring 2008 Colloquium series and the U. of Delaware's Cognitive Science Brown-Bag series.

LDP 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.	ksunonus	'I obey him'	5.	ʃtijepus	'he tells him'
2.	kʃunotʃ	'I am obedient'	6.	*sustimeʃ	(hypothetical)
3.	*ksunonuʃ	(hypothetical)	7.	ʃiʃlusisin	'they (dual) are gone awry'
4.	kʃunots	(hypothetical)			

LDP 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.

*kʃinots	(hypothetical)
ʃtijepus	'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.

*kʃinots	(hypothetical)
ʃtijepus	'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).

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.

$$\text{Precedence } G = \left\{ \quad \right\}$$

Sample = { }

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

Learning Relativized Precedence Bigram Grammars

The learner simply records the relativized precedence bigram relations observed.

Precedence G =

(ʃi,ʃ)					
	(iʃ,l)	(iʃ,u)	(iʃ,s)	(iʃ,i)	(iʃ,n)
		(ʃl,u)	(ʃl,s)	(ʃl,i)	(ʃl,n)
			(lu,s)	(lu,i)	(lu,n)
			(us,s)	(us,i)	(us,n)
			(si,s)		(si,n)
				(is,i)	

Sample = { ʃiʃlusisin }

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

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].

Learning Relativized Precedence Bigram Grammars

The learner simply records the relativized precedence bigram relations observed.

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

Sample = { ʃiʃlusisin }

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

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.