An Unambiguous Strategy For Learning Complex Linguistic Systems
Lisa Pearl, University of California, Irvine
Nov 26, 2007
University of Southern California

The Learning Problem
There is often a non-transparent relationship between the observable form of the data and the underlying system that produced it. Moreover, data are often ambiguous.

Syntactic System
Observable form: word order
Difficulty: interactive structural pieces

Subject → Verb → Object

Metrical Phonology System
Observable form: stress contour
Difficulty: interactive structural pieces

af ter noon

One Solution: Constraints on Hypothesis Space
Premise: learner considers finite range of hypotheses (parameters) (Halle & Vergnaud, 1987; Chomsky, 1981)

But this doesn’t solve the learning problem…

“Assuming that there are $n$ binary parameters, there will be $2^n$ possible core grammars.” - Clark (1994)
How do learners choose among these hypotheses? I think so, Brain - but do we really need two tongues?

Real learning seems to be gradual.

Probabilistic Learning with Parameters


The Naïve Parameter (NPar) Learner

Probabilistic learning strategy explicitly compatible with parameterized grammars: learning is gradual & variable

“grammars that succeed in analyzing [a data point] are rewarded and those that fail are punished”

Complex System Woes

But this may not always work when we have complex systems with multiple parameters.

Are there other strategies a learner might adopt as well?

Learning Framework: 3 Components

(1) Hypothesis space

(2) Data intake

(3) Update procedure

A \( P_A = 0.5 \) B \( P_B = 0.5 \)

A \( P_A = ?? \) B \( P_B = ?? \)

Input

Investigating Data Intake Filtering

Intuition 1: Use all available data to uncover a full range of systematicity, and allow probabilistic model enough data to converge.

Intuition 2: Use more "informative" data or more "accessible" data only.

Ambiguous Data Woes:
Feasibility of an Unambiguous Data Filter

It is unlikely that any example … would show the effect of only a single parameter value; rather, each example is the result of the interaction of several different principles and parameters" - Clark (1994)

(1) Is something beyond probabilistic learning necessary? (Necessity)

(2) Is there a data sparseness problem for an unambiguous data filter? (Feasibility)

(3) Does learning from unambiguous data yield correct behavior? (Sufficiency)

Today’s Plan
Given a realistic complex system to learn and realistic data to learn from, we ask...

Useful Tool: Modeling
Why? Can easily and ethically manipulate some part of the learning mechanism and observe the effect on learning.

Recent computational modeling surge for language learning mechanisms: Niyogi & Berwick, 1996; Saffran, 1997; Yang, 2000; Berwick & Levelt, 2000; Boersma & Hayes, 2001; Sakas & Fodor, 2001; Yang, 2002; Sakas & Nishimoto, 2002; Sakas, 2003; Mintz, 2003; Apoussidou & Boersma, 2004; Fodor & Sakas, 2004; Pearl, 2005; Fodor, Rots, & Brent, 2006; Mintz, 2006; Pearl & Weinberg, 2007; Hayes & Wilson, 2007; Wang & Mintz, 2007

Important: Empirically grounded in realistic data & psychologically plausible learning constraints

Road Map
I. The System
Parameterized Metrical Phonology

II. The Input

III. Learning Without Filters

IV. The Filter

V. Learning With Filters

VI. Good Ideas

Metrical Phonology
What tells you to put the Emphasis on a particular Syllable sample metrical phonology structure from parametric system

Metrical Phonology Parameters

Foot Headedness

Foot Boundedness

Quantity Sensitivity

Extrametricality

Foot Directionality

Syllable type (Light, Heavy)
Metrical Phonology Parameters

Quantity Sensitivity

- Feet Headedness
- Feet Boundedness
- Extrametricality
- Feet Directionality

Feet Headedness

Extrametricality

Feet Boundedness

Feet Directionality

Extrametricality

Quantity Sensitivity: QI

Quantity Insensitive (QI): All syllables are treated the same (S)

- S
- S
- S
- VV
- V
- VC
- CVV
- CV
- CCVC
- lu
di
crous

Quantity Sensitivity: QS

Quantity Sensitive (QS):
Syllables are separated into Light and Heavy
V are always L, VV are always H

- VC-Light (QSVCL) = VC syllable is L
- VC-Heavy (QSVCH) = VC syllable is H

- H
- L
- L/H
- VV
- V
- VC
- CVV
- CV
- CCVC
- lu
di
crous

Quantity Sensitivity: Stress

Rule of Stress: If a syllable is Heavy, it should get stressed - unless some other parameter interacts with it

Extrametricality, Metrical Feet, and Stress

Rule of Stress: If a syllable is extrametrical, it is not included in a metrical foot. If a syllable is not in a metrical foot, it cannot have stress.
Extrametricality: None

Extrametricality: Some

Extrametricality: Some

Metrical Phonology Parameters

Feet Directionality

Feet Directionality
Feet Directionality

**Feet Direction** What edge of the word **metrical foot** construction begins at

**Feet Direction Left**: start from **left** edge

\[(H \quad L) (H)\]

**Feet Direction Right**: start from **right** edge

\[(H) (L \quad H)\]

---

Metrical Phonology Parameters

- Quantity Sensitivity
- Extrametricality
- Feet Headedness
- Feet Boundedness
- Feet Directionality

---

Boundedness: Unbounded Feet

Unbounded: a metrical foot **extends until a heavy syllable** is encountered

- start from **left**

\[L \quad L \quad L \quad H \quad L\]
Boundedness: Unbounded Feet

Unbounded: a metrical foot extends until a heavy syllable is encountered

\begin{align*}
\text{start from left} & \rightarrow (L \ L \ L)(H \ L) \\
L & \ L & \ L & \ H & \ L \quad \text{start from right}
\end{align*}

Boundedness: Unbounded Feet

Unbounded: a metrical foot extends until a heavy syllable is encountered

- Start from left: (L L L)(H L)
- Start from right: (L L L H)(L)

Boundedness: Bounded Feet

Bounded: a metrical foot only extends a certain amount (cannot be longer)

- Bounded-2: a metrical foot only extends 2 units
  - Start from left: (x x)(x x)(x)
  - Start from right: x x x x x

- Bounded-3: a metrical foot only extends 3 units
  - Start from left: (x x)(x x)(x)
  - Start from right: x x x x x
Boundedness: Bounded Feet

**Bounded**: a metrical foot only extends a certain amount (cannot be longer)

**Bounded-2**: a metrical foot only extends 2 units

\[
\text{start from left} \Rightarrow (x \ x) (x \ x) (x)
\]

**Bounded-3**: a metrical foot only extends 3 units

\[
\text{start from left} \Rightarrow (x \ x \ x) (x \ x)
\]

**Bounded-Syllabic**: counting unit is syllable

\[
\text{start from left} \Rightarrow \text{L H L L H}
\]

\[
\text{bounded-2} \Rightarrow \text{L H L L H}
\]

**Bounded-Moramic**: counting unit is mora

\[
H = 2 \text{ moras, } L = 1 \text{ mora}
\]

\[
\text{start from left} \Rightarrow (L \ H) (L \ L) (H)
\]

\[
\text{bounded-2} \Rightarrow (L \ H) (L \ L) (H)
\]

\[
H = 2 \text{ moras, } L = 1 \text{ mora}
\]
Boundedness: Bounded Feet

**Bounded-Syllabic:** counting unit is syllable
- start from left: $$(L\ L)(L\ L)(H)$$
- bounded-2: $$(H\ H)(L\ L)(H)$$

**Bounded-Moramic:** counting unit is mora
- $H = 2$ moras, $L = 1$ mora
- start from left: $$(X\ X)(X\ X)(X\ X)(X\ X)$$
- bounded-2: $$(H\ H)(L\ L)(H)$$

Metrical Phonology Parameters

- Quantity Sensitivity
- Extrametricality
- Feet Headedness
- Feet Directionality
- Feet Boundedness

**Metrical Feet and Stress**

Rule of Stress: Exactly one syllable per metrical foot must have stress.
Feet Headedness

Feet Headedness: which syllable of metrical foot gets stress

Feet Head Left: leftmost syllable in foot gets stress

(\textbf{H}) (L \textbf{H})

Feet Head Right: rightmost syllable in foot gets stress

(\textbf{H}) (L \textbf{H})

Metrical Phonology Parameters

- Quantity Sensitivity
- Extrametricality
- Feet Headedness
- Feet Boundedness
- Feet Directionality

English Metrical Phonology

Non-trivial language: English (full of exceptions)

Input: data unambiguous for the incorrect value in the adult system

Adult English system values:

- QS, QSVCH, Em-Some, Em-Right, Ft Dir Right, Bounded, B-2, B-Syllabic, Ft Hd Left

Exceptions:

- Qi, QSVCL, Em-None, Ft Dir Left, Unbounded, B-3, B-Moraic, Ft Hd Right
Empirical Grounding in Realistic Data: Estimating English Data Distributions

Caretaker speech to children between the ages of 6 months and 2 years (CHILDES: MacWhinney, 2000)

Total Words: 540505
Mean Length of Utterance: 3.5

Words parsed into syllables and assigned stress using the American English CALLHOME database of telephone conversation (Canavan et al., 1997) & the MRC Psycholinguistic database (Wilson, 1988)

Road Map

I. The System
II. The Input
III. Learning Without Filters
   The Naïve Parameter Learner
IV. The Filter
V. Learning With Filters
VI. Good Ideas

Probabilistic Learning with Parameters

Naïve Parameter (NPar) Learner

For any data point encountered, a parameter value combination (grammar) is generated using the current probabilities. The learner attempts to parse the current data point with this combination.

Probabilistic Learning with Parameters

For each parameter, the learner associates a probability with each of the competing parameter values

QI = 0.7
Em-Some = 0.4
Em-None = 0.6
Ft Dir Left = 0.8
Ft Dir Rt = 0.2
Bounded = 0.6
Unbounded = 0.4
Ft Hd Left = 0.5
Ft Hd Rt = 0.5
...
The NPar Learner on English Metrical Phonology

Learning Period Length: 1,160,000 words (based on estimates of words heard in a 6 month period, using Akhtar et al. (2004)).

Learning rate: (0.01 ≤ x ≤ 0.05)

Results using distributions in English child-directed speech:

Learners never converge on English.

If learners ignore monosyllabic words (since such words don’t have a stress contour per se), less than 1.2% converge on English.

Examples of incorrect target languages NPar learners converged on:

- Em-None, Ft Dir Left, Unb, Ft Dir Left
- GS, Em-None, QSVCH, Ft Dir Left, Ft Dir Left, B-Mor, Bounded, Bounded
- Em-None, Ft Dir Left, Unb, Ft Dir Left
- Em-None, Ft Dir Left, B-2, Unb, Ft Dir Left
A More Conservative Learner: NPar Learner + Batch

For any data point encountered, a parameter value combination (grammar) is generated using the current probabilities. The learner attempts to parse the current data point with this combination.

If the data point can be parsed, then all participating parameter values have their batch counter incremented by 1.

If the data point cannot be parsed, then all participating parameter values have their batch counters decremented by 1.

"...it slows down the learning rate when [the parameter] is bad and speeds it up when [the parameter] gets better" - Yang (2002)

For each parameter, the learner associates a probability with each of the competing parameter values.

- Qi = 0.7  QS = 0.3
- Em-None = 0.4  Em-None = 0.6
- Ft Dir Left = 0.8  Ft Dir Rt = 0.2
- Bounded = 0.6  Unbounded = 0.4
- Ft Hd Left = 0.5  Ft Hd Rt = 0.5

...
Examples of incorrect target languages per se), Learners never converge on English.

If learners ignore monosyllabic words (since such words don’t have a stress contour per se), less than 0.8% converge on English. (Always when b = 2)

Examples of incorrect target languages NPar + B learners converged on:

- Em-Right, Em-None, Unbounded, Ft Dir Right, QS, Ft Dir Left
- Em-Right, F1HdLe, Em-None, Unbounded, Ft Dir Left, QS, QSVC
- Em-Right, Em-None, QSVC
- QI

Results using distributions in English child-directed speech:

- Lower limit b = 0
- Upper limit b = 10

The learner may already have knowledge of F1HdLe and QI. Perhaps this knowledge of the system will allow the NPar + B learner to converge on the rest of the English values more reliably.
The NPar + B Learner with Prior Knowledge on English Metrical Phonology

Learning Period Length: 1,160,000 words (based on estimates of words heard in a 6 month period, using Akhtar et al. (2004)).

0.01 ≤ learning rate γ ≤ 0.05
2 ≤ batch size b ≤ 10

Results using distributions in English child-directed speech:
Learners never converge on English.

If learners ignore monosyllabic words (since such words don’t have a stress contour per se), 5% or less learners converge on English.

Examples of incorrect target languages NPar + B learners with prior knowledge converged on:
Fi Hi Lat, QS, Em-Right, Em-some, QS/CH, Fl Di Lat, Bounded, B-Mer, B-2
Fi Hi Lat, QS, Em-Right, QS/Ch, Em- some, Unbounded, Fl Di Lat
Fl Hi Lat, QS, Em-Right, Em- some, QS/CH, Fl Di Lat, B-3, Bounded, B-Mor
Fl Hi Lat, QS, Em-Right, Em- some, QS/CH, Fl Di R, Unbounded

---

Big picture

For the complex linguistic system under consideration, knowledge of the parameters and a probabilistic learning strategy doesn’t seem to yield the correct behavior. These are insufficient for learning.

Something else seems necessary for successful acquisition.

Can data filtering help?

---

Road Map

I. The System
II. The Input
III. Learning Without Filters
IV. The Filter
  Unambiguous Data
V. Learning With Filters
VI. Good Ideas

---

Filter Feasibility

How feasible is an unambiguous data filter for a complex system?

Data sparseness: are there unambiguous data? (Clark 1992)

How could a learner identify such data?

---

Identifying Unambiguous Data

Identifying unambiguous data:

Cues (Dresher, 1999; Lightfoot, 1999)

Parsing (Fodor, 1998; Sakas & Fodor, 2001)

---

Changing Knowledge States:
Unambiguous Data is a Moving Target

Current knowledge of system influences perception of unambiguous data. The informativity of a data point changes over time.

Data initially ambiguous may later be perceived as unambiguous.
Data initially unambiguous may later be perceived as exceptional.

Point: The order in which parameters are set may determine if they are set correctly (Dresher, 1999).
Cues: Overview

A cue is a local "specific configuration in the input" that corresponds to a specific parameter value. A cue matches an unambiguous data point. (Dresher, 1999)

Cues for Metrical Phonology Parameters

Recall: Cues match local surface structure (sample cues below)

QS: 2 syllable word with 2 stresses

<table>
<thead>
<tr>
<th>2 syllable word</th>
<th>2 stresses</th>
</tr>
</thead>
<tbody>
<tr>
<td>VV</td>
<td>VV</td>
</tr>
</tbody>
</table>

Em-Right: Rightmost syllable is Heavy and unstressed

<table>
<thead>
<tr>
<th>Rightmost syllable</th>
<th>Heavy and unstressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>... H</td>
<td>... H</td>
</tr>
</tbody>
</table>

Unb: 3+ unstressed S/L syllables in a row

<table>
<thead>
<tr>
<th>3+ unstressed S/L syllables</th>
<th>in a row</th>
</tr>
</thead>
<tbody>
<tr>
<td>... S S S...</td>
<td>... L L L</td>
</tr>
</tbody>
</table>

Ft Dir Right: Leftmost foot has stress on leftmost syllable

<table>
<thead>
<tr>
<th>Leftmost foot</th>
<th>Stress on leftmost syllable</th>
</tr>
</thead>
<tbody>
<tr>
<td>S S S...</td>
<td>H L L ...</td>
</tr>
</tbody>
</table>

 Parsing: Overview

Parsing tries to analyze a data point with "all possible parameter value combinations", conducting an "exhaustive search of all parametric possibilities", and then discovering what is common to them. (Dresher, 1996)

Parsing with Metrical Phonology Parameters

Sample Datum: VC VC VV (afternoon)

 Parsing with Metrical Phonology Parameters

Sample Datum: VC VC VV (afternoon)

Parsing with Metrical Phonology Parameters

Sample Datum: VC VC VV (afternoon)

Parsing with Metrical Phonology Parameters

Sample Datum: VC VC VV (afternoon)
### Parsing with Metrical Phonology Parameters

**Sample Datum:** VC VC VV (‘afternoon’)

<table>
<thead>
<tr>
<th>(x)</th>
<th>(x)</th>
<th>(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>L</td>
<td>H</td>
</tr>
<tr>
<td>VC</td>
<td>VC</td>
<td>VV</td>
</tr>
</tbody>
</table>

Values leading to successful parses of data:

- (QS, QSVC, Em-None, Fl Dr Right, B, B-2, B-Syl, Fl Hd Left)
- (QS, QSVC, Em-None, Fl Dr Right, B, B-2, B-Syl, Fl Hd Left)

Data point is unambiguous for `Em-None`.

Perception of unambiguous data changes over time:

- If QI already set, data point is unambiguous for `Em-None`, B, B-2, and B-Syl.

### Cues vs. Parsing: Comparison

**Cues:**

- Easy identification of unambiguous data
- Can find information in sub-part of data point
- Can tolerate exceptions

**Parsing:**

- Not heuristic (exhaustive search)
- No additional knowledge beyond parameter values
- No default values used

---

**Cues:**

- Are heuristic
- Require additional knowledge
- May rely on default values
- Bounded-Syl unless data indicate Bounded-Moraic

**Fl Hd Lt**

**Fl Hd Rt**

---

**Fl Hd Lt**

**Fl Hd Rt**
Cues vs. Parsing: Comparison

**Parsing:**
- Resource-intensive identification of unambiguous data
- Needs complete parse of data point to get any information:
  - Cannot find information in sub-part of data point
  - Cannot tolerate exceptions

**Cues:**
- Psychological plausibility: does not require entire data set at once to learn from
  - Does not use default values
  - Does not require additional knowledge
  - Is not heuristic
  - Can find information in datum sub-part
  - Can tolerate exceptions
  - Is not heuristic
  - Does not require additional knowledge
  - Does not use default values
  - Psychological plausibility: does not require entire data set at once to learn from

Road Map

I. The System
II. The Input
III. Learning Without Filters
IV. The Filter
V. Learning With Filters
  - Simulating What Children Do
VI. Good Ideas

The Learning Process

Data Filtering: The learner encounters a data point and decides if it's unambiguous for any parameter values.

Updating Hypotheses: If so, the learner shifts probability to those parameter values.

Probabilistic Learning Intuition: The parameter value whose unambiguous data have a higher probability of being encountered by the learner will win when the learner is setting that parameter value.
Moving Targets & Unambiguous Data; What Happens After Parameter Setting

<table>
<thead>
<tr>
<th>Quantity Sensitivity</th>
<th>Extrametricality</th>
</tr>
</thead>
<tbody>
<tr>
<td>QI: 0.00398</td>
<td>QS: 0.0205</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feet Directionality</th>
<th>Boundedness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left: 0.000</td>
<td>Right: 0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feet Headedness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left: 0.00148</td>
</tr>
</tbody>
</table>

Moving Targets & Unambiguous Data

<table>
<thead>
<tr>
<th>Quantity Sensitivity</th>
<th>Extrametricality</th>
</tr>
</thead>
<tbody>
<tr>
<td>QI: 0.00398</td>
<td>QS: 0.0205</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feet Directionality</th>
<th>Boundedness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left: 0.000</td>
<td>Right: 0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feet Headedness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left: 0.00148</td>
</tr>
</tbody>
</table>

Setting a Parameter Value

Moving Targets & Unambiguous Data

<table>
<thead>
<tr>
<th>QS-VC-Heavy/Light</th>
<th>Extrametricality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heavy: 0.00265</td>
<td>Light: 0.00398</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feet Directionality</th>
<th>Boundedness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left: 0.000</td>
<td>Right: 0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feet Headedness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left: 0.000588</td>
</tr>
</tbody>
</table>
Getting to English

The learner must set all the parameter values in order to converge on a language system.

Current knowledge of the system influences the perception of unambiguous data. So, the order in which parameters are set influences the probability of encountering unambiguous data for unset parameters.

To get to English, the learner must converge on QS, QSVCH, Em-Some, Em-Right, Ft Dir Rt, Bounded, Bounded-2, Bounded-Syl, Ft Hd Left.

Getting to English: Exhaustive Search of All Parameter Setting Orders

(a) Try one parameter-setting order...

(b) For all currently unset parameters, determine the unambiguous data distribution in the corpus.

(c) Choose a currently unset parameter to set. The value chosen for this parameter is the value that has a higher probability in the data the learner perceives as unambiguous.

(d) Repeat steps (a-b) until all parameters are set.

Is it English?

Getting to English: Exhaustive Search of All Parameter Setting Orders

Is it English?

(d) Compare final set of values to English set of values. If they match, this is a viable parameter setting order. If they don’t, it isn’t.

Getting to English: Exhaustive Search of All Parameter Setting Orders

Repeat for all possible orders...

Try one parameter-setting order...

Is it English?

Results: Set of viable orders that lead to English (we hope)

Viable Parameter Setting Orders

Worst Case: learning with unambiguous data produces insufficient behavior.
No orders lead to English.

Better Case: learning with unambiguous data produces sufficient behavior.
Viable orders exist, even if some orders don’t lead to English.

Best Case: learning with unambiguous data is a brilliant plan!
All orders lead to English.

Road Map

I. The System
II. The Input
III. Learning Without Filters
IV. The Filter
V. Learning With Filters
VI. Good Ideas
Filters, Predictions, & Future Directions
Parameter Setting Orders: Knowledge Necessary for Acquisition Success

"Viable parameter setting order" means...

If the learner manages to set the parameters in this order, the learner will converge on English.

But wouldn’t it be better if the viable orders could be captured more compactly, instead of being explicitly listed in the learner’s mind?

Order #23 looks good!

Feasibility & Sufficiency of the Unambiguous Data Filter

Either method of identifying unambiguous data (cues or parsing) is successful. Given the non-trivial parametric system (9 interactive parameters) and the non-trivial data set (English is full of exceptions), this is no small feat.

“It is unlikely that any example ... would show the effect of only a single parameter value” - Clark (1994)

Feasibility & Sufficiency of the Unambiguous Data Filter

Either method of identifying unambiguous data (cues or parsing) is successful. Given the non-trivial parametric system (9 interactive parameters) and the non-trivial data set (English is full of exceptions), this is no small feat.

"It is unlikely that any example ... would show the effect of only a single parameter value" - Clark (1994)

(1) Unambiguous data exist and can be identified in sufficient relative quantities to extract the correct systematically for a complex parametric system.

(2) The data intake filtering strategy is robust across a realistic (highly ambiguous, exception-filled) data set. It’s feasible to identify such data, and the strategy yields sufficient learning behavior.
Predictions: Links to the Experimental Side

- **Cues**
  - (a) QS-VS-Heavy before Em-Right
  - (b) Em-Right before Bounded-Syl
  - (c) Bounded-2 before Bounded-Syl

Are predicted parameter setting orders observed in real-time learning? E.g. whether cues or parsing is used. Quantity Sensitivity (QS, QSVCH) is predicted to be set before Extrametricality (Em-Some, Em-Right).

---

Future Directions in Modeling

(1) Is the unambiguous data filter successful for other languages besides English? Other instantiations of metrical phonology? Other complex linguistic domains like syntax?

(2a) Are these order constraints reasonable/feasible as knowledge the learner needs for acquisition success? (Ask me!)

(2b) Can we combine the strengths of cues and parsing? (Ask me how!)

(3) Are there other methods of data filtering that might be successful for learning English metrical phonology? (e.g. Yang, 2005)

(4) How necessary is a data filtering strategy for successful learning? Would other probabilistic learning strategies that are not as selective about the data intake succeed? (e.g. Fodor & Hayes, 2001; Bayesian learning strategies)

(5) Can other knowledge implementations, such as constraint satisfaction systems (Tesar & Smolensky, 2000; Boersma & Hayes, 2001), be successfully learned from noisy data sets like English? (Theoretical implications based on learnability of the system)
Take Home Message

(1) Modeling results for a realistic system and realistic data set suggest the necessity of something beyond a simple probabilistic learning strategy, even if the hypothesis space of learners is already constrained.

(2) They also demonstrate the viability of the unambiguous data filter as a learning strategy.

(3) Computational modeling is a very useful tool:
   (a) empirically test learning strategies that would be difficult to investigate with standard techniques
   (b) generate experimentally testable predictions about learning

Thank You

Amy Weinberg
Bill Idsardi

Jeff Lidz
Charles Yang

The audiences at
BUCLD 32
UC Irvine Language Learning Group
UC Irvine Department of Cognitive Sciences
CUNY Psycholinguistics Support Club
UDelaware Linguistics Department
Yale Linguistics Department
UMaryland Cognitive Neuroscience of Language Lab

Why Parameters?

Why posit parameters instead of just associating stress contours with words?
Arguments from stress change over time (Dresher & Lahiri, 2003):

(1) If word-by-word association, expect piece-meal change over time at the individual word level. Instead, historical linguists posit changes to underlying systems to best explain the observed data that change altogether.

(2) If stress contours are not composed of pieces (parameters), expect start and end states of change to be near each other. However, examples exist where start & end states are not closely linked from perspective of observable stress contours.
Calculating Unambiguous Data Probability:
Relativizing Probabilities

Relativize-against-all:
- probability conditioned against entire input set
- relativizing set is constant across methods

<table>
<thead>
<tr>
<th>Cues or Parsing</th>
<th>QI</th>
<th>QS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unambiguous Data Points</td>
<td>2140</td>
<td>11213</td>
</tr>
<tr>
<td>Relativizing Set</td>
<td>540505</td>
<td>540505</td>
</tr>
<tr>
<td>Relativized Probability</td>
<td>0.00396</td>
<td>0.0207</td>
</tr>
</tbody>
</table>

Relativize-against-potential:
- probability conditioned against set of data points that meet preconditions of being an unambiguous data point
- relativizing set is not constant across methods

Cues: have correct syllable structure (e.g. 2 syllables if cue is 2 syllable word with both syllables stressed)

<table>
<thead>
<tr>
<th>Unambiguous Data Points</th>
<th>QI</th>
<th>QS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relativizing Set</td>
<td>2755</td>
<td>85268</td>
</tr>
<tr>
<td>Relativized Probability</td>
<td>0.777</td>
<td>0.132</td>
</tr>
</tbody>
</table>

Cues vs. Parsing:
Success Across Relativization Methods
(Getting to English)

<table>
<thead>
<tr>
<th>Cues</th>
<th>Parsing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative-Against-All</td>
<td>Successful</td>
</tr>
<tr>
<td>Relative-Against-Potential</td>
<td>Unsuccessful</td>
</tr>
</tbody>
</table>

...so parsing seems more robust across relativization methods.
Order Constraints

Good: Order constraints exist that will allow the learner to converge on the adult system, provided the learner knows these constraints.

Better: These order constraints can be derived from properties of the learning system, rather than being stipulated, or they're already known through other means.

Deriving Constraints from Properties of the Learning System

Data saliency: presence of stress is more easily noticed than absence of stress, and indicates a likely parametric cause

Data quantity: more unambiguous data available

Default values (cues only): if a value is set by default, order constraints involving it may disappear

Note: data quantity and default values would be applicable to any system. Data saliency is more system-dependent.

Deriving Constraints: Cues

(a) QS-VC-Heavy before Em-Right

(b) Em-Right before Bounded-Syl

(c) Bounded-2 before Bounded-Syl

Knowing Through Other Means

Infant research has shown that infants are sensitive to some of the rhythmic properties of their language.

Jusczyk, Cutler, & Redanz (1993): English 6-month olds prefer strong-weak stress bisyllables (trochaic) to weak-strong ones (iambic).

Turk, Jusczyk, & Gerken (1995): English infants are sensitive to the difference between long vowels and short vowels in syllables.

The learner may already have knowledge of Ft Hd Left and QS, so these are set early.
Deriving Constraints: Cues

(a) QS-VC-Heavy before Em-Right
   - Em-Right: absence of stress is less salient (data saliency); prior knowledge

(b) Em-Right before Bounded-Syl
   - Bounded-Syl as default (default values)
   - Em-Right: more unambiguous data than Bounded-Syl (data quantity)

(c) Bounded-2 before Bounded-Syl
   - Bounded-Syl as default (default values)
   - Bounded-2 has more unambiguous data than Bounded-2 or Bounded-Syl (data quantity)

Deriving Constraints: Parsing

Group 1:
- QS, Ft, Håf, Left, Bounded

Group 2:
- Ft Dir Right, QS-VS-Heavy

Group 3:
- Em-Right, Bounded-2, Bounded-Syl

Em-Some, Em-Right: absence of stress is less salient (data saliency)
Combining Cues and Parsing

Cues and parsing have a complementary array of strengths and weaknesses.

Problem with cues: requires prior knowledge
Problem with parsing: requires parse of entire data point

Viable combination of cues & parsing:
  parsing of data point subpart = derivation of cues?

Non-derivable Constraints: Predictions Across Languages?

Parsing Constraints

Group 1:
  QS, Ft Ht Left, Bounded

Group 2:
  Ft Dir Right, QS-VS-Heavy

Group 3:
  Em-Some, Em-Right, Bounded-2, Bounded-Syl

Combining Cues and Parsing

Em-Right: Rightmost syllable is Heavy and unstressed

If a syllable is Heavy, it should be stressed.
If an edge syllable is Heavy and unstressed, an immediate solution (given the available parametric system) is that the syllable is extrametrical.
Combining Cues and Parsing

Viable combination of cues & parsing:

\[ \text{parsing of data point subpart} = \text{derivation of cues?} \]

Would \textit{partial parsing}:

(a) derive cues that lead to successful acquisition?
(b) retain the strengths that cues & parsing have separately?
(c) be a more psychologically plausible implementation of the unambiguous data filter?