Putting the Emphasis on Unambiguous: The Feasibility of Data Filtering for Learning English Metrical Phonology

Lisa Pearl
University of California, Irvine
BUCLD 32
Nov 3, 2007

Human Language Learning

Theoretical work:
object of acquisition

x
(x x) x
H L H
em pha sis

Experimental work:
time course of acquisition

mechanism of acquisition
given the boundary conditions provided by
(a) linguistic representation
(b) the trajectory of learning

The Learning Problem

There is often a non-transparent relationship between the observable form of the data and the underlying system that produced it.

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

af ter noon

(x x) (x)
S S S
af ter noon

(x x) (x)
L L H
af ter noon

(x) (x x)
L L H
af ter noon

Learner Bias: Parameters

Premise: learner considers finite range of hypotheses (parameters) (Halle & Vergnaud, 1987)

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)

The Mechanism of Language Learning: Extracting Systematicity

Data is often ambiguous

"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)

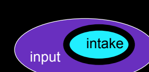
(x x) (x)
S S S
af ter noon

(x x) (x)
L L H
af ter noon

(x) (x x)
L L H
af ter noon

Learner Bias: Data Filtering

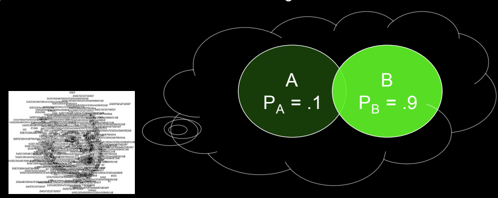
Potential solution: the learner is biased to focus in on an informative subset of the data.



feasibility issue: data sparseness

Useful Tool: Modeling

Why? Can easily and ethically manipulate some part of the learning process and observe the effect on learning.



Recent computational modeling surge: Niyogi & Berwick, 1996; Boersma, 1997; Yang, 2000; Boersma & Levelt, 2000; Boersma & Hayes, 2001; Sakas & Fodor, 2001; Yang, 2002; Sakas & Nishimoto, 2002; Sakas, 2003; Apoussidou & Boersma, 2004; Fodor & Sakas, 2004; Pearl, 2005; Pater, Potts, & Bhatt, 2006; Pearl & Weinberg, 2007; Hayes & Wilson, 2007

Questions

How viable are these kind of biases in a realistic environment?

Is a complex parametric system really learnable?
Are there enough data to learn from if the learner filters the input set and learns only from a select subset?

Questions

How viable are these kind of biases in a realistic environment?

Is a complex parametric system really learnable?
Are there enough data to learn from if the learner filters the input set and learns only from a select subset?

Feasibility: Is there a data sparseness problem?
Sufficiency: Can the learner filter and still display correct learning behavior?

Questions

How viable are these kind of biases in a realistic environment?

Is a complex parametric system really learnable?
Are there enough data to learn from if the learner filters the input set and learns only from a select subset?

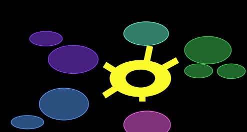
Feasibility: Is there a data sparseness problem?
Sufficiency: Can the learner filter and still display correct learning behavior?

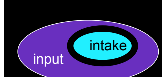
Key: Learning from a realistic data set (CHILDES: MacWhinney, 2000)

Today’s Plan: Demonstrate Viability

Learning a complex parametric system from a noisy data set by filtering the data intake is both feasible and sufficient

System: metrical phonology,
9 interactive parameters





Filter: Learn only from unambiguous data

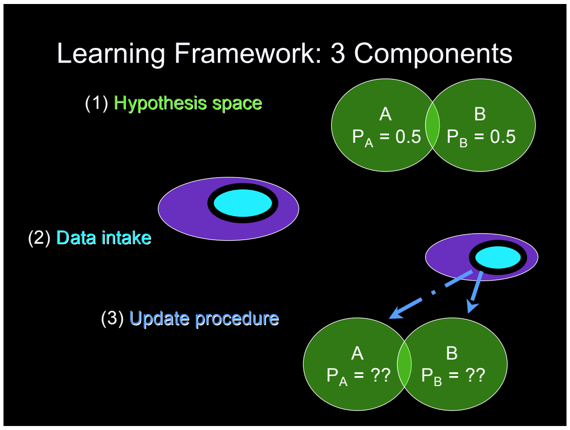
Data Set: highly noisy English child-directed speech (540505 words)

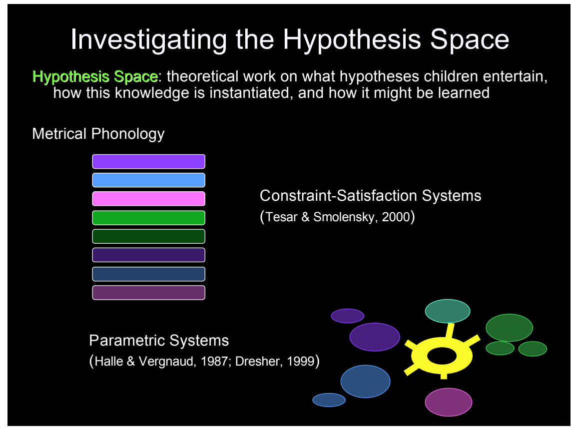
Road Map

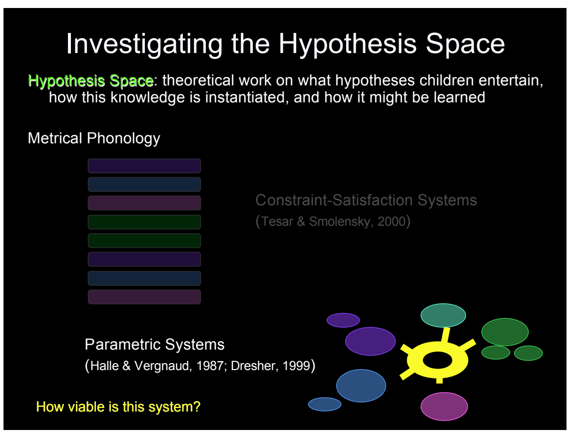
Learning Framework Overview

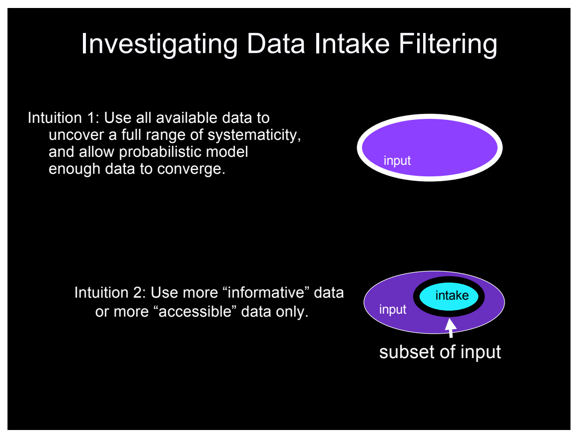
Computational Modeling: Learning Metrical Phonology
Data intake filtering and learning a complex parametric system for metrical phonology

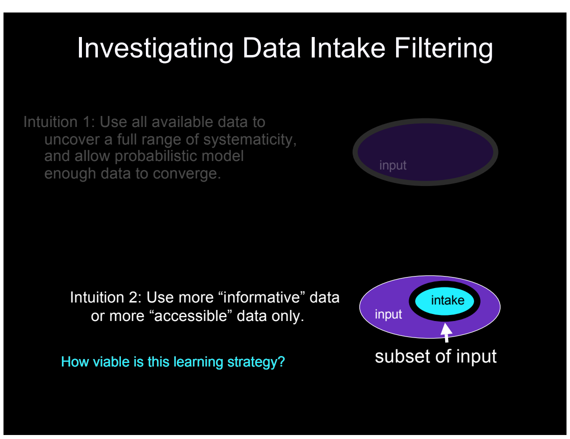
Important Features: empirical grounding
- searching realistic data space for evidence of underlying system
- considering psychological plausibility of learning methods

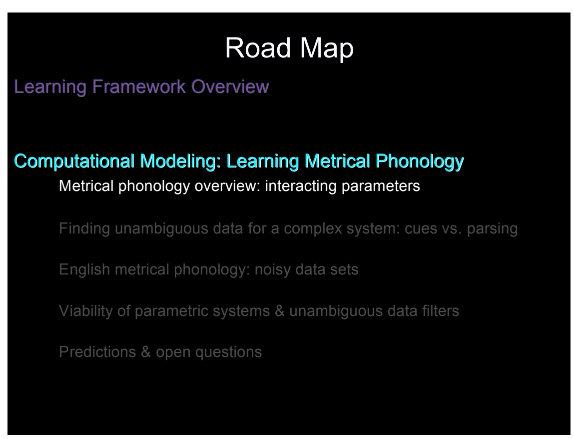

Learning Framework: 3 Components
(1) Hypothesis space
A
$P_A = 0.5$
B
$P_B = 0.5$
(2) Data intake
(3) Update procedure
A
P_A = ??
B
P_B = ??


Investigating the Hypothesis Space
Hypothesis Space: theoretical work on what hypotheses children entertain, how this knowledge is instantiated, and how it might be learned
Metrical Phonology
Constraint-Satisfaction Systems
(Tesar & Smolensky, 2000)
Parametric Systems
(Halle & Vergnaud, 1987; Dresher, 1999)


Investigating the Hypothesis Space
Hypothesis Space: theoretical work on what hypotheses children entertain, how this knowledge is instantiated, and how it might be learned
Metrical Phonology
Constraint-Satisfaction Systems
(Tesar & Smolensky, 2000)
Parametric Systems
(Halle & Vergnaud, 1987; Dresher, 1999)
How viable is this system?


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.
input
Intuition 2: Use more "informative" data or more "accessible" data only.
intake
input
subset of input

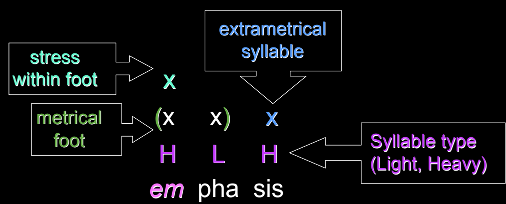

Investigating Data Intake Filtering
Intuition 1: Use all available data to uncover a full range of systematicity, and allow probabilistic model enough data to converge.
input
Intuition 2: Use more "informative" data or more "accessible" data only.
intake
input
How viable is this learning strategy?
subset of input


Road Map
Learning Framework Overview
Computational Modeling: Learning Metrical Phonology
Metrical phonology overview: interacting parameters
Finding unambiguous data for a complex system: cues vs. parsing
English metrical phonology: noisy data sets
Viability of parametric systems & unambiguous data filters
Predictions & open questions

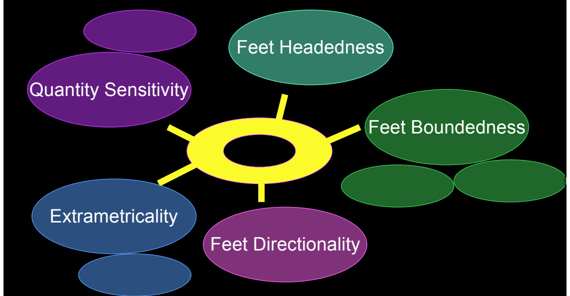
Metrical Phonology

What tells you to put the *EM*phasis on a particular *SYL*lable

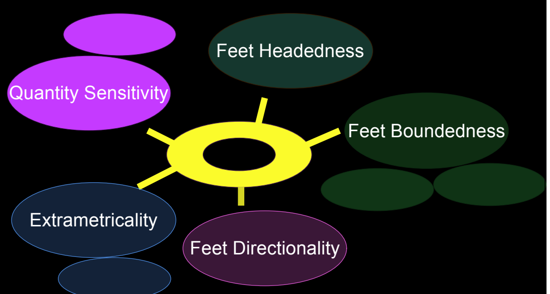
sample metrical phonology structure from parametric system



Metrical Phonology Parameters



Metrical Phonology Parameters



Quantity Sensitivity: QI

Quantity-**I**nsensitive (**QI**): All syllables are treated the same (S)

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

Quantity Sensitivity: QS

Quantity-**S**ensitive (**QS**):
Syllables are separated into **L**ight and **H**eavy
V are always L, VV are always H

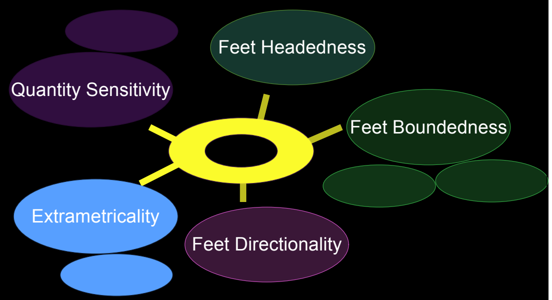
VC-**L**ight (**QSVCL**) = **VC** syllable is **L**
VC-**H**eavy (**QSVCH**) = **VC** syllable is **H**

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

Quantity Sensitivity: Stress

Rule of Stress: If a syllable is **H**eavy, it **should have stress** - unless some other parameter interacts with it

Metrical Phonology Parameters



Extrametricality, Metrical Feet, and Stress

Rule of Stress: If a syllable is **extrametrical**, it ***cannot* have stress** because it is not included in a metrical foot.

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

Extrametricality: None

Extrametricality-None (Em-None): All syllables are in metrical feet

metrical foot → (L L) (H)

VC VC VV

af ter **noon**

Extrametricality: Some

Extrametricality-Some (Em-Some): One edge syllable not in foot

Extrametricality-Left (Em-Left): Leftmost syllable not in foot - ***cannot* have stress**

extrametrical syllable → L (H L) ← metrical foot

V VC V

a **gen** da

Extrametricality: Some

Extrametricality-Some (Em-Some): One edge syllable not in foot

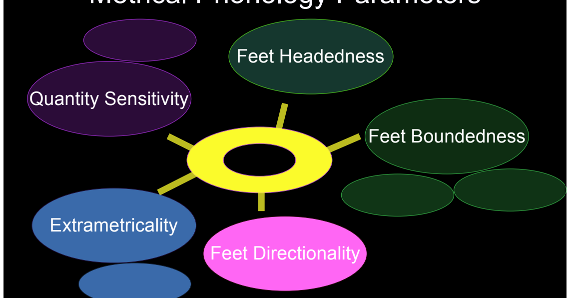
Extrametricality-Right (Em-Right): Rightmost syllable not in foot - ***cannot* have stress**

metrical foot → (H L) H ← extrametrical syllable

VV V VC

lu di crous

Metrical Phonology Parameters



Feet Directionality

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

Feet Direction Left: start from left edge

H L H

Feet Direction Right: start from right edge

H L H

Feet Directionality

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

Feet Direction Left: start from left edge

(H L) H

Feet Direction Right: start from right edge

H L H

Feet Directionality

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

Feet Direction Left: start from left edge

(H L) (H)

Feet Direction Right: start from right edge

H L H

Feet Directionality

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

Feet Direction Left: start from left edge

(H L) (H)

Feet Direction Right: start from right edge

H (L H)

Feet Directionality

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

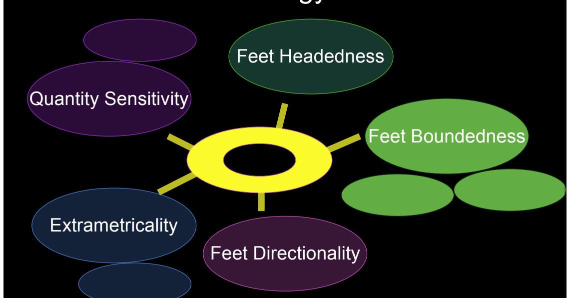
Feet Direction Left: start from left edge

(H L) (H)

Feet Direction Right: start from right edge

(H) (L H)

Metrical Phonology Parameters



Boundedness: Unbounded Feet

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

Boundedness: Unbounded Feet

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

start from left ⇨ L L L H L

Boundedness: Unbounded Feet

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

start from left ⇨ (L L L) H L

Boundedness: Unbounded Feet

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

start from left ⇨ (L L L)(H L)

Boundedness: Unbounded Feet

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

start from left ⇨ (L L L)(H L)

L L L H L ⇦ start from right

Boundedness: Unbounded Feet

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

start from left ⇨ (L L L)(H L)

L L L H (L) ⇦ start from right

Boundedness: Unbounded Feet

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

start from left → (L L L)(H L)

(L L L H)(L) ← start from right

Boundedness: Unbounded Feet

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

start from left → (L L L)(H L)

(L L L H)(L) ← start from right

start from left → L L L L L

Boundedness: Unbounded Feet

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

start from left → (L L L)(H L)

(L L L H)(L) ← start from right

start from left → (L L L L L)

Boundedness: Unbounded Feet

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

start from left → (L L L)(H L)

(L L L H)(L) ← start from right

start from left → (L L L L L)

(L L L L L) ← start from right

Boundedness: Bounded Feet

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

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

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

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

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

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)

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

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)

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

start from left → 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**

start from left → (x x)(x x)(x)

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

start from left → (x x x)(x x)

Boundedness: Bounded Feet

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

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

H = **2** moras, **L** = **1** mora

Boundedness: Bounded Feet

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

start from left → L H L L H

bounded-2 →

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

H = **2** moras, **L** = **1** mora

Boundedness: Bounded Feet

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

start from left → (L H)(L L)(H)

bounded-2 →

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

H = **2** moras, **L** = **1** mora

Boundedness: Bounded Feet

Bounded-Syllabic: counting unit is syllable

start from left → (L H)(L L)(H)

bounded-2 → H H L L H

Bounded-Moraic: counting unit is mora

H = 2 moras, L = 1 mora

Boundedness: Bounded Feet

Bounded-Syllabic: counting unit is syllable

start from left → (L H)(L L)(H)

bounded-2 → (H H)(L L)(H)

Bounded-Moraic: counting unit is mora

H = 2 moras, L = 1 mora

Boundedness: Bounded Feet

Bounded-Syllabic: counting unit is syllable

start from left → (L H)(L L)(H)

bounded-2 → (H H)(L L)(H)

S S S S S

Bounded-Moraic: counting unit is mora

H = 2 moras, L = 1 mora

Boundedness: Bounded Feet

Bounded-Syllabic: counting unit is syllable

start from left → (L H)(L L)(H)

bounded-2 → (H H)(L L)(H)

(S S)(S S)(S)

Bounded-Moraic: counting unit is mora

H = 2 moras, L = 1 mora

Boundedness: Bounded Feet

Bounded-Syllabic: counting unit is syllable

start from left → (L H)(L L)(H)

bounded-2 → (H H)(L L)(H)

(S S)(S S)(S)

Bounded-Moraic: 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

Boundedness: Bounded Feet

Bounded-Syllabic: counting unit is syllable

start from left → (L H)(L L)(H)

bounded-2 → (H H)(L L)(H)

(S S)(S S)(S)

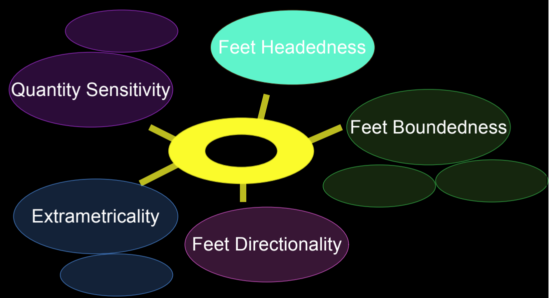
Bounded-Moraic: counting unit is mora

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



Feet Headedness

Feet Headedness: which syllable of metrical foot gets stress

Feet Head Left: leftmost syllable in foot gets stress

(H) (L H)

Feet Head Right: rightmost syllable in foot gets stress

(H) (L H)

Feet Headedness

Feet Headedness: which syllable of metrical foot gets stress

Feet Head Left: leftmost syllable in foot gets stress

(H) (L H)

Feet Head Right: rightmost syllable in foot gets stress

(H) (L H)

Feet Headedness

Feet Headedness: which syllable of metrical foot gets stress

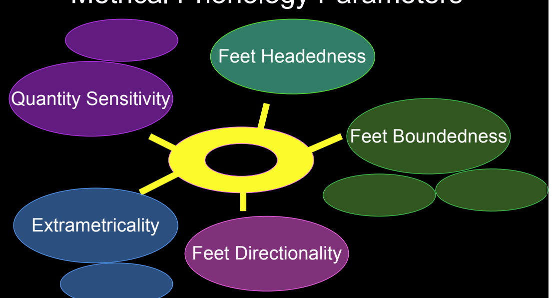
Feet Head Left: leftmost syllable in foot gets stress

(H) (L H)

Feet Head Right: rightmost syllable in foot gets stress

(H) (L H)

Metrical Phonology Parameters



Road Map

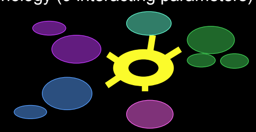
Learning Framework Overview

Computational Modeling: Learning Metrical Phonology

- Metrical phonology overview: interacting parameters
- Finding unambiguous data for a complex system: cues vs. parsing
- English metrical phonology: noisy data sets
- Viability of parametric systems & unambiguous data filters
- Predictions & open questions

Filter Feasibility

Metrical phonology (9 interacting parameters)



How feasible is an unambiguous data filter for a complex system with a noisy data set as input?

Data sparseness: are there unambiguous data? (Clark 1992)
How could a learner identify such data?

Interactive Parameters

Current knowledge of system influences perception of unambiguous data: The order in which parameters are set may determine if they are set correctly (Dresher, 1999).

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

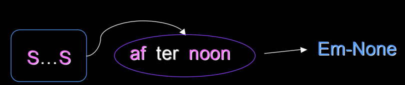
Identifying unambiguous data:

Cues (Dresher, 1999; Lightfoot, 1999)

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

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)



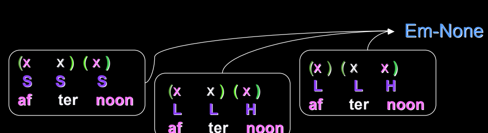
Cues for Metrical Phonology Parameters

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

QS: 2 syllable word with 2 stresses	VV VV
Em-Right: Rightmost syllable is Heavy and unstressed	… H
Unb: 3+ unstressed S/L syllables in a row	…S S S… … L L L L
Ft Hd Left: Leftmost foot has stress on leftmost syllable	S S S… H L L …

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. (Fodor, 1998)



Parsing with Metrical Phonology Parameters

Sample Datum: VC VC VV ('afternoon')

Parsing with Metrical Phonology Parameters

Sample Datum: VC VC VV ('afternoon')

(QS, QSVCL, Em-None, Ft Dir Right,
B, B-2, B-Syl, Ft Hd Right)

```
( x )   ( x     x )
  L       L     H )
  VC      VC    VV
```

Parsing with Metrical Phonology Parameters

Sample Datum: VC VC VV ('afternoon')

(QS, QSVCL, Em-None, Ft Dir Right,
B, B-2, B-Syl, Ft Hd Right)

```
( x )   ( x     x )
  L       L     H )
  VC      VC    VV
```

(QS, QSVCL, Em-None, Ft Dir Left,
B, B-2, B-Syl, Ft Hd Left)

```
( x       x )  ( x )
( L       L      H
  VC      VC     VV
```

Parsing with Metrical Phonology Parameters

Sample Datum: VC VC VV ('afternoon')

(QS, QSVCL, Em-None, Ft Dir Right,
B, B-2, B-Syl, Ft Hd Right)

```
( x )   ( x     x )
  L       L     H )
  VC      VC    VV
```

(QS, QSVCL, Em-None, Ft Dir Left,
B, B-2, B-Syl, Ft Hd Left)

```
( x       x )  ( x )
( L       L      H
  VC      VC     VV
```

(QI, Em-None, Ft Dir Right,
B, B-2, B-Syl, Ft Hd Right)

```
( x       x )  ( x )
  S       S      S )
  VC      VC     VV
```

Parsing with Metrical Phonology Parameters

Values leading to successful parses of datum:

(QI, Em-None, Ft Dir Left, Ft Hd Left, B, B-2, B-Syl)
(QI, Em-None, Ft Dir Right, Ft Hd Right, B, B-2, B-Syl)
(QS, QSVCL, Em-None, Ft Dir Left, Ft Hd Left, UnB)
(QS, QSVCL, Em-None, Ft Dir Left, Ft Hd Left, B, B-2, B-Syl)
(QS, QSVCL, Em-None, Ft Dir Right, Ft Hd Right, B, B-2, B-Syl)

Datum is unambiguous for Em-None.

Parsing with Metrical Phonology Parameters

Values leading to successful parses of datum:

(QI, Em-None, Ft Dir Left, Ft Hd Left, B, B-2, B-Syl)
(QI, Em-None, Ft Dir Right, Ft Hd Right, B, B-2, B-Syl)
(QS, QSVCL, Em-None, Ft Dir Left, Ft Hd Left, UnB)
(QS, QSVCL, Em-None, Ft Dir Left, Ft Hd Left, B, B-2, B-Syl)
(QS, QSVCL, Em-None, Ft Dir Right, Ft Hd Right, B, B-2, B-Syl)

Datum is unambiguous for Em-None.

Perception of unambiguous data changes over time:

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

Cues vs. Parsing: A Note on Psychological Plausibility

Both cues and parsing are learning methods that are incremental. They operate over a single data point at a time, and do not require the learner to conduct analyses across the entire collection of data points encountered.

Road Map

Learning Framework Overview

Computational Modeling: Learning Metrical Phonology

- Metrical phonology overview: interacting parameters
- Finding unambiguous data for a complex system: cues vs. parsing
- English metrical phonology: noisy data sets
- Viability of parametric systems & unambiguous data filters
- Predictions & open questions

Finding Unambiguous Data: English Metrical Phonology

Non-trivial parametric system: metrical phonology

Non-trivial language: English (full of exceptions)
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

Learning Framework Overview

Computational Modeling: Learning Metrical Phonology

- Metrical phonology overview: interacting parameters
- Finding unambiguous data for a complex system: cues vs. parsing
- English metrical phonology: noisy data sets
- Viability of parametric systems & unambiguous data filters
- Predictions & extensions

Sufficient Filters: Viable Parameter-Setting Orders

Can learners using unambiguous data (identified by either cues or parsing) learn the English parametric system? What parameter-setting orders lead to the correct English system?

Viable orders are derived for each method via an exhaustive walk-through of all possible parameter-setting orders.

Viable Parameter-Setting Orders: Encapsulating the Knowledge for Acquisition Success

Worst Case: learning with unambiguous data produces insufficient behavior
No orders lead to correct system - parametric system is unlearnable

Better Cases: learning with unambiguous data produces sufficient behavior
Slightly Better Case: Viable orders available, but fairly random

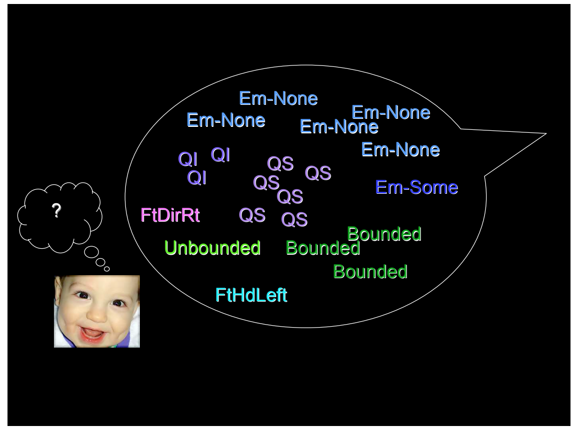
Better Case: Viable orders available, can be captured by small number of *order constraints*

Best Case: All orders lead to correct system

Identifying Viable Parameter-Setting Orders

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

Quantity Sensitivity		Extrametricality	
QI: .00398	QS: 0.0205	None: 0.0294	Some: .0000259
Feet Directionality		**Boundedness**	
Left: 0.000	Right: 0.00000925	Unbounded: 0.00000370	Bounded: 0.00435
Feet Headedness			
Left: 0.00148	Right: 0.000		

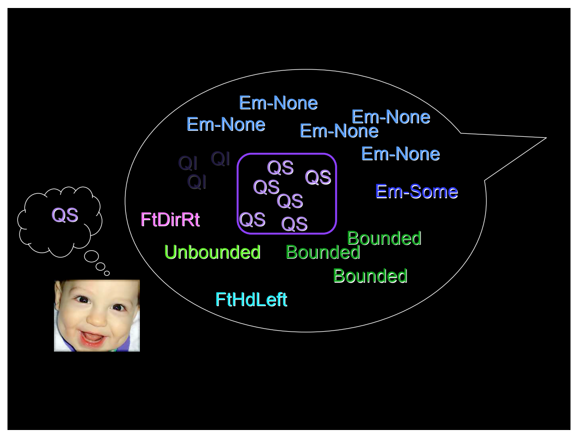


Identifying Viable Parameter-Setting Orders

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

(b) 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.

Quantity Sensitivity		Extrametricality	
QI: .00398	QS: 0.0205	None: 0.0294	Some: .0000259
Feet Directionality		**Boundedness**	
Left: 0.000	Right: 0.00000925	Unbounded: 0.00000370	Bounded: 0.00435
Feet Headedness			
Left: 0.00148	Right: 0.000		



Identifying Viable Parameter-Setting Orders

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

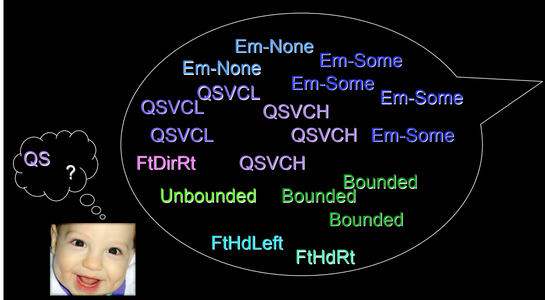
(b) Choose a currently unset parameter to set. The value chosen for this parameter is the value that has a higher probability in the data the learner perceives as unambiguous.

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

Identifying Viable Parameter-Setting Orders

(a) For all currently unset parameters, determine the unambiguous data distribution in the corpus…

QS-VC-Heavy/Light		Extrametricality	
Heavy: .00265	Light: 0.00309	None: 0.0240	Some: .0485
Feet Directionality		**Boundedness**	
Left: 0.000	Right: 0.00000555	Unbounded: 0.00000370	Bounded: 0.00125
Feet Headedness			
Left: 0.000588	Right: 0.0000204		



Identifying Viable Parameter-Setting Orders

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

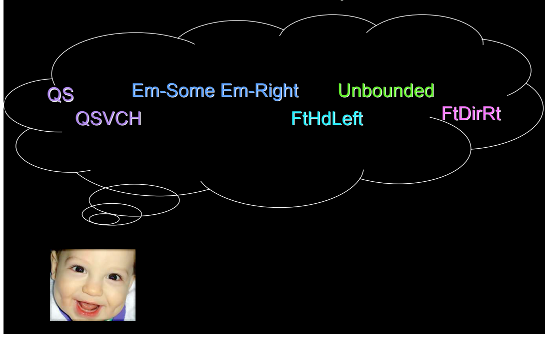
(b) Choose a currently unset parameter to set. The value chosen for this parameter is the value that has a higher probability in the data the learner perceives as unambiguous.

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

(d) Compare final set of values to English set of values. If they match, this is a viable parameter-setting order.

Is it English?

QS, QSVCH, Em-Some, Em-Right, FtDirRt, Bounded, Bounded-2, Bounded-Syl, FtHdLeft



Identifying Viable Parameter-Setting Orders

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

(b) 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.

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

(d) Compare final set of values to English set of values. If they match, this is a viable parameter-setting order.

(e) Repeat (a-d) for all parameter-setting orders.

Sufficiency of an Unambiguous Filter for a Complex Parametric System

Are there any viable parameter-setting orders for a learner using unambiguous data (identified by either cues or parsing)?

Cues: Parameter-Setting Orders

Cues: Sample viable orders

(a) QS, QS-VC-Heavy, Bounded, Bounded-2, Feet Hd Left, Feet Dir Right, Em-Some, Em-Right, Bounded-Syl

(b) Feet Dir Right, QS, Feet Hd Left, Bounded, QS-VC-Heavy, Bounded-2, Em-Some, Em-Right, Bounded-Syl

Cues: Sample failed orders

(a) QS, Bounded, Feet Hd Left, Feet Dir Right, QS-VC-Heavy, Em-Some, Em-Right, Bounded-Syl, Bounded-2

(b) Feet Hd Left, Feet Dir Right, Bounded, Bounded-Syl, Bounded-2, QS, QS-VC-Heavy, Em-Some, Em-Right

Parsing: Parameter-Setting Orders

Parsing: Sample viable orders

(a) Bounded, QS, Feet Hd Left, Feet Dir Right, QS-VC-Heavy, Bounded-Syl, Em-Some, Em-Right, Bounded-2

(b) Feet Hd Left, QS, QS-VC-Heavy, Bounded, Feet Dir Right, Em-Some, Em-Right, Bounded-Syl, Bounded-2

Parsing: Sample failed orders

(a) Feet Dir Right, QS, Feet Hd Left, Bounded, QS-VC-Heavy, Bounded-2, Em-Some, Em-Right, Bounded-Syl

(b) Em-Some, Em-Right, QS, Bounded, Feet Hd Left, Feet Dir Right, QS-VC-Heavy, Bounded-Syl, Bounded-2

Cues vs. Parsing: Order Constraints

Cues

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

(b) Em-Right
before Bounded-Syl

(c) Bounded-2
before Bounded-Syl

The rest of the parameters are freely ordered w.r.t. each other.

Parsing

Group 1:
QS, Ft Head Left, Bounded

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

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

The parameters are freely ordered w.r.t. each other within each group.

Feasibility & Sufficiency of the Unambiguous Data Filter for Learning a Parametric System

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 for Learning a Parametric System

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 can be identified in sufficient quantities to extract the correct systematicity for a complex parametric system.

(2) The data intake filtering strategy is robust across a realistic (highly ambiguous, exception-filled) data set.

Big Questions for Learning a Complex Parametric System and the Data Intake Filtering Strategy: English Metrical Phonology

(1) Feasibility
No data sparseness problem, even for a complex system with multiple interactive parameters.

(2) Sufficiency
Learning from unambiguous data yields the correct learning behavior.

Road Map

Learning Framework Overview

Computational Modeling: Learning Metrical Phonology

- Metrical phonology overview: interacting parameters
- Finding unambiguous data for a complex system: cues vs. parsing
- English metrical phonology: noisy data sets
- Viability of parametric systems & unambiguous data filters
- Predictions & open questions

Predictions

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

Parsing
Group 1:
QS, Ft Head Left, Bounded
Group 2:
Ft Dir Right, QS-VS-Heavy
Group 3:
Em-Some, Em-Right, Bounded-2, Bounded-Syl

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

Open Questions

(1) Is the unambiguous data filter successful for other languages besides English? Other complex linguistic domains?

(2) Can we combine the strengths of cues and parsing?

(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 learning strategies that are not as selective about the data intake succeed? (e.g. Yang, 2002; Fodor & Sakas, 2004)

(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?

Take Home Message

(1) Modeling results support the viability of both the parametric implementation of metrical phonology knowledge and the unambiguous data filter as a learning strategy, even for a noisy data set.

(2) 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 — Jeff Lidz
Bill Idsardi — Charles Yang
the Cognitive Neuroscience of Language Lab
at the University of Maryland
the Department of Cognitive Sciences at UC Irvine

Benefits of Learning Framework

Components:
(1) hypothesis space (2) data intake (3) update procedure

Application to a wide range of learning problems, provided these three components are defined
- Ex: hypothesis space defined in terms of parameter values (Yang, 2002) or in terms of how much structure is posited for the language (Perfors, Tenenbaum, & Regier, 2006)

Can combine discrete representations (hypothesis space) with probabilistic components (update procedure)

Cues vs. Parsing in a Probabilistic Framework

Critique of Learning Behavior:
"Both models ... cannot capture the variation in and the gradualness of language development…when a parameter *is* set, it is set in an all-or-none fashion." - Yang (2002)

Benefit of using learning framework to sidestep this problem - separable components used in combination:
(1) cues/parsing to identify unambiguous data
(2) probabilistic framework of gradual updating based on unambiguous data

Why Parameters?

Why posit parameters instead of just associating stress contours with words?

Arguments from stress change over time (Dresher & Lahiri, 2003):

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.

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.

(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.

Relativizing Probabilities

Relativize-against-all:

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

Cues or Parsing

	QI	QS
Unambiguous Data Points	2140	11213
Relativizing Set	540505	540505
Relativized Probability	0.00396	0.0207

Relativizing Probabilities

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)

	QI	QS
Unambiguous Data Points	2140	11213
Relativizing Set	2755	85268
Relativized Probability	0.777	0.132

Relativizing Probabilities

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

Parsing: able to be parsed

	QI	QS
Unambiguous Data Points	2140	11213
Relativizing Set	*p*	*p*
Relativized Probability	2140/*p*	11213/*p*

Cues vs. Parsing: Preference?

Is there any (additional) reason to prefer one method of identifying unambiguous data over the other?

Cues

VV VV L H H

… L L L L

H L L… …S S S S…

S S S…

Parsing

(QI, Em-None, Ft Dir Left, Ft Hd Left, B, B-2, B-Syl)
(QI, Em-None, Ft Dir Right, Ft Hd Right, B, B-2, B-Syl)
(QS, QSVCL, Em-None, Ft Dir Left, Ft Hd Left, UnB)
(QS, QSVCL, Em-None, Ft Dir Left, Ft Hd Left, B, B-2, B-Syl)
(QS, QSVCL, Em-None, Ft Dir Right, Ft Hd Right, B, B-2, B-Syl)

Cues vs. Parsing: Success Across Relativization Methods

	Cues	Parsing
Relative-Against-All	Successful	Successful
Relative-Against-Potential	*Unsuccessful*	Successful

…so parsing seems more robust across relativization methods.

Another Consideration: Constraint Derivability

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.

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 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

Deriving Constraints: Cues

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

Em-Right: absence of stress is less salient (data saliency)

(b) Em-Right
before Bounded-Syl

(c) Bounded-2
before Bounded-Syl

Deriving Constraints: Cues

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

Em-Right: absence of stress is less salient (data saliency)

(b) Em-Right before Bounded-Syl

Bounded-Syl as default (default values)

(c) Bounded-2 before Bounded-Syl

Deriving Constraints: Cues

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

Em-Right: absence of stress is less salient (data saliency)

(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

Deriving Constraints: Cues

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

Em-Right: absence of stress is less salient (data saliency)

(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)

Deriving Constraints: Cues

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

Em-Right: absence of stress is less salient (data saliency)

(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 once Em-Right is set; Em-Right has much more than Bounded-2 or Bounded-Syl (data quantity)

Deriving Constraints: Parsing

Group 1:
QS, Ft Head Left, Bounded

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

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

Deriving Constraints: Parsing

Group 1:
QS, Ft Head Left, Bounded

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

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

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

Deriving Constraints: Parsing

Group 1:
QS, Ft Head Left, Bounded

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

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

Other groupings cannot be derived from data quantity, however…

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

Cues vs. Parsing: Comparison

	Cues	Parsing
Easy identification of unambiguous data	+	
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	+	+

Combining Cues and Parsing

Cues and parsing have a complementary array of strengths and weaknesses

Problem with cues: require prior knowledge
Problem with parsing: requires parse of entire datum

Viable combination of cues & parsing:
parsing of datum subpart = derivation of cues?

Combining Cues and Parsing

Em-Right: Rightmost syllable is Heavy and unstressed …H(H)

If a syllable is Heavy, it should be stressed.

If an edge syllable is Heavy and unstressed, an immediate solution (given the available parameteric system) is that the syllable is extrametrical.

Combining Cues and Parsing

Viable combination of cues & parsing:
parsing of datum subpart = derivation of cues?

Would partial parsing
(a) derive cues that lead to successful acquisition?
(b) be a more psychologically plausible representation of the learning mechanism?

Non-derivable Constraints: Predictions Across Languages?

Parsing Constraints

Group 1:
QS, Ft Head Left, Bounded

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

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

Do we find these same groupings if we look at other languages?

The Necessity of Data Intake Filtering

Alternate Strategy: learn from all data (no filters)

Yang (2002): Naïve Parameter Learner (NP Learner)

- Learner has probabilities associated with each parameter value
- For each data point
  - learner randomly chooses a parameter value combination, based on the associated probabilities
  - learner tries to parse data point with this random parameter value combination
  - if parse succeeds, all participating values rewarded
  - if parse fails, all participating values punished

Idea: unambiguous data will only be parseable by correct parameter value; incorrect value eventually punished into zero probability

Preliminary results: not successful for English data set (possibly due to numerous exceptions in data set); Batch Learner version also not successful.