Cognitive Modeling:
How Humans Learn Complex Linguistic Systems

Lisa Pearl, UC Irvine
Mar 10, 2008
AIML Seminar Series
Center for Machine Learning & Intelligent Systems
UC Irvine
Cognitive Modeling of Language
Different problems: more and less easily discernible from data

**Categorization/Clustering**
- Ex: What are the contrastive sounds of a language?
  - Vowel categories in English & Japanese
  - Hypothesis space: 3 dimensions of variation
  - English relevant dimensions: 1 and 2
  - Japanese relevant dimensions: 2 and 3

Valabha et al. 2007

**Categorization/Clustering**
- Ex: What are the contrastive sounds of a language?
  - Assumption from experimental work:
    - Relevant unit of word segmentation for infants is the syllable

Swingley 2005
Gambell & Yang 2006

**Categorization/Clustering**
- Ex: What are the contrastive sounds of a language?
  - Mapping
    - What are the word affixes that signal meaning (e.g. past tense in English)?
      - Observable data: word order
      - Generative system: syntax

**Complex systems:** What is the generative system that creates the observed (structured) data of language (ex: syntax, metrical phonology)?

Complex systems: What is the generative system that creates the observed (structured) data of language (ex: syntax, metrical phonology)?

Observable data: stress contour
Generative system: metrical phonology

**Categorization/Clustering**
- Ex: What are the contrastive sounds of a language?
  - Mapping
    - What are the word affixes that signal meaning (e.g. past tense in English)?

**Complex systems:** What is the generative system that creates the observed (structured) data of language (ex: syntax, metrical phonology)?

Complex systems: What is the generative system that creates the observed (structured) data of language (ex: syntax, metrical phonology)?

Observable data: stress contour
Generative system: metrical phonology

**Complex systems:** What is the generative system that creates the observed (structured) data of language (ex: syntax, metrical phonology)?

Observable data: stress contour
Generative system: metrical phonology
Cognitive Modeling of Language
Different problems: more and less easily discernible from data

Categorization/Clustering
Ex: What are the contrastive sounds of a language?

Extraction
Ex: Where are words in fluent speech?

Mapping
What are the word affixes that signal meaning (e.g. past tense in English)?

Complex systems: What is the generative system that creates the observed (structured) data of language (ex: syntax, metrical phonology)?

Observable data: stress contour EMphasis
Generative system: metrical phonology

( S ) S
( L ) H
( EM pha sis ) ( H L L )

Road Map
Introduction to complex linguistic systems
General problems
Parametric systems
Parametric metrical phonology

Learnability of complex linguistic systems
General learnability framework
Case study: English metrical phonology
Available data & associated waxes
Unconstrained probabilistic learning
Constrained probabilistic learning

Where next? Implications & Extensions

General Problems with Learning Complex Linguistic Systems
What children encounter: the output of the generative linguistic system EMphasis

What children must learn: the components of the system that combine to generate this observable output

EMphasis

Which syllable of a larger unit is stressed?

Are all syllables included?

Are syllables differentiated?
Chomsky (1981), Hypothesis for a language consists of a combination of generalizations about that language (grammar). But this leads to a theoretically infinite hypothesis space. Moreover, data are often ambiguous, even if parameters of variation are known.

Languages only differ in constrained ways from each other. Not all generalizations are possible.

Observation: Languages only differ in constrained ways from each other. Not all generalizations are possible.

Why this is tricky: There is often a non-transparent relationship between the observable form of the data and the underlying system that produced it. Hard to know what parameters of variation to consider.

Moreover, data are often ambiguous, even if parameters of variation are known.

Why this is tricky: There is often a non-transparent relationship between the observable form of the data and the underlying system that produced it. Hard to know what parameters of variation to consider.

EMphasis

Hypothesis for a language consists of a combination of generalizations about that language (grammar). But this leads to a theoretically infinite hypothesis space.

For example, assuming there are n binary parameters, there are 2^n core grammars to choose from.

Exponentially growing hypothesis space

Linguistic parameters gives the benefit of a finite hypothesis space. Still, the hypothesis space can be quite large.
Parametric Metrical Phonology

Metrical phonology:
What tells you to put the **Emphasis** on a particular **Syllable**

Process speakers use:
- Basic input unit: syllables
- Larger units formed: metrical feet
  - The way these are formed varies from language to language. Only syllables in metrical feet can be stressed.
  - Stress assigned within metrical feet
    - The way this is done also varies from language to language.

Observable Data: stress contour of word **Emphasis**

A Brief Tour of Parametric Metrical Phonology

Are syllables differentiated?
- No: system is quantity-insensitive (QI)
  - CVV CV CCVC

   No: system is quantity-insensitive (QI)
   - lu di crous

   No: system is quantity-insensitive (QI)
   - CVV CV CCVC

   No: system is quantity-insensitive (QI)
   - lu di crous

Yes: system is quantity-sensitive (QS)
  - Only allowed method: differ by rime weight
  - Only allowed number of divisions: 2
    - Heavy vs. Light
      - VV always Heavy
      - V always Light

Option 1: VC Heavy (QS-VC-H)
- CVV CV CCVC
- lu di crous

Option 2: VC Light (QS-VC-L)
- CVV CV CCVC
- lu di crous

Are syllables included in metrical feet?
- Yes: system has no extrametricality (Em-None)
  - VC VC VC
  - af fur noon

- No system is quantity-insensitive (QI)
  - lu di crous

  - only allowed method: differ by rime weight

  - only allowed number of divisions: 2

  - Heavy vs. Light

  - VV always Heavy

  - V always Light

  - narrowing of hypothesis space

  - Option 1: VC Heavy (QS-VC-H)

  - Option 2: VC Light (QS-VC-L)
A Brief Tour of Parametric Metrical Phonology

Are all syllables included in metrical feet?

Yes: system has no extrametricality (Em-None)

No: system has extrametricality (Em-Some)

Only allowed # of exclusions: 1

Leftmost or Rightmost syllable

narrowing of hypothesis space

A Brief Tour of Parametric Metrical Phonology

Are metrical feet unrestricted in size?

Yes: Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded)

No: Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded)

A Brief Tour of Parametric Metrical Phonology

What direction are metrical feet constructed?

Two logical options:

From the left:
Metrical feet are constructed from the left edge of the word (Ft Dir Left)

From the right:
Metrical feet are constructed from the right edge of the word (Ft Dir Right)
A Brief Tour of Parametric Metrical Phonology

**Are metrical feet unrestricted in size?**

**Yes:** Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

**No:** Metrical feet are restricted (Bounded).

- The size is restricted to 2 options: 2 or 3.
- The counting units are restricted to 2 options: syllables or moras.

**Ft Dir Left**

- 2 units per foot (Bounded-2)
- 3 units per foot (Bounded-3)

**Ft Dir Right**

- (L L L) [H L]
- (L L L) [L L]
- (B L L) [S S S]

---

A Brief Tour of Parametric Metrical Phonology

**Are metrical feet unrestricted in size?**

**Yes:** Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

**No:** Metrical feet are restricted (Bounded).

- The size is restricted to 2 options: 2 or 3.
- The counting units are restricted to 2 options: syllables or moras.

**Ft Dir Left**

- (L L L) [H L]
- (L L L) [L L]
- (B L L) [S S S]

**Ft Dir Right**

- (L L L) [H L]
- (L L L) [L L]
- (B L L) [S S S]

---

A Brief Tour of Parametric Metrical Phonology

**Are metrical feet unrestricted in size?**

**Yes:** Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

**No:** Metrical feet are restricted (Bounded).

- The size is restricted to 2 options: 2 or 3.
- The counting units are restricted to 2 options: syllables or moras.

**Ft Dir Left**

- (L L L) [H L]
- (L L L) [L L]
- (B L L) [S S S]

**Ft Dir Right**

- (L L L) [H L]
- (L L L) [L L]
- (B L L) [S S S]

---

A Brief Tour of Parametric Metrical Phonology

**Are metrical feet unrestricted in size?**

**Yes:** Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

**No:** Metrical feet are restricted (Bounded).

- The size is restricted to 2 options: 2 or 3.
- The counting units are restricted to 2 options: syllables or moras.

**Ft Dir Left**

- (L L L) [H L]
- (L L L) [L L]
- (B L L) [S S S]

**Ft Dir Right**

- (L L L) [H L]
- (L L L) [L L]
- (B L L) [S S S]

---

A Brief Tour of Parametric Metrical Phonology

**Are metrical feet unrestricted in size?**

**Yes:** Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

**No:** Metrical feet are restricted (Bounded).

- The size is restricted to 2 options: 2 or 3.
- The counting units are restricted to 2 options: syllables or moras.

**Ft Dir Left**

- (L L L) [H L]
- (L L L) [L L]
- (B L L) [S S S]

**Ft Dir Right**

- (L L L) [H L]
- (L L L) [L L]
- (B L L) [S S S]
A Brief Tour of Parametric Metrical Phonology

Are metrical feet unrestricted in size?

Yes: Metrical feet are unrestricted, delimited only by Heavy syllables if there are any (Unbounded).

No: Metrical feet are restricted (Bounded).

The size is restricted to 2 options: 2 or 3.

The counting units are restricted to 2 options: syllables or moras.

Count by syllables (Bounded-Syllabic) Count by moras (Bounded-Moraic)

Generating a Stress Contour

Process speaker uses to generate stress contour

Are syllables differentiated?

Yes.

VC syllables are Heavy.

Generating a Stress Contour

Process speaker uses to generate stress contour

Are any syllables extrametrical?

Yes.

Rightmost syllable is not included in metrical foot.

Generating a Stress Contour

Process speaker uses to generate stress contour

Which direction are feet constructed from?

From the right.

Generating a Stress Contour

Process speaker uses to generate stress contour

Are feet unrestricted?

No.

2 syllables per foot.
Generating a Stress Contour

Process speaker uses to generate stress contour.

Which syllable of the foot is stressed?
Leftmost.

(M) H
VC
em
HL
CV
EM
pha
sis

A caveat about learning parameters separately
Parameters are system components that combine together to generate output.
Choice of one parameter may influence choice of subsequent parameters.

Road Map
Introduction to complex linguistic systems
General problems
Parametric systems
Parametric metrical phonology
Learnability of complex linguistic systems
General learnability framework
Case study: English metrical phonology
Available data & associated woes
Unconstrained probabilistic learning
Constrained probabilistic learning
Where next? Implications & Extensions

Choosing among grammars
Human learning seems to be gradual and somewhat robust to noise - need some probabilistic learning component.

Probabilistic learning over parameter values

Since grammars are parameterized, child can make use of this information to constrain hypothesis space. Learn over parameters, not entire parameter value sets.

A caveat about learning parameters separately
Parameters are system components that combine together to generate output.
Choice of one parameter may influence choice of subsequent parameters.
A caveat about learning parameters separately

Parameters are system components that combine together to generate output.
Choice of one parameter may influence choice of subsequent parameters.

Point: The order in which parameters are set may determine if they are set correctly from the data.
Dresher 1999

The learning framework: 3 components

1) Hypothesis space
2) Data
3) Update procedure

Key point for cognitive modeling: psychological plausibility

Any probabilistic update procedure must, at the very least, be incremental/online.

Why? Humans (especially human children) don’t have infinite memory. Unlikely: human children can hold a whole corpus worth of data in their minds for analysis later on.
Models that do this are AI (not cognitive modeling) - they can simulate human behavior, but not necessarily the way humans produce it.
(ex: Foraker et al. 2007, Goldwater et al. 2007)

Two psychologically plausible probabilistic update procedures

Naïve Parameter Learner (NParLearner)
Probabilistic generation & testing of parameter value combinations. (incremental)
Hypothesis update: Linear reward-penalty
(Bush & Mosteller 1951)

Bayesian Learner (BayesLearner)
Probabilistic generation & testing of parameter value combinations. (incremental)
Hypothesis update: Bayesian updating
(Chew 1971, binomial distribution)
Probabilistic learning for English

Proportional generation and testing of parameter values (Yang 2002)

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

*QI = 0.5  QSVCL = 0.5  QSVCH = 0.5  Em-Some = 0.5  Em-None = 0.5  Em-Left = 0.5  Ft Dir Left = 0.5  Bounded = 0.5  Bounded-Moraic = 0.5  Bounded-2 = 0.5  Bounded-Moraic-2 = 0.5  Bounded-Syl = 0.5  Ft Dir Right = 0.5  Ft Dir Left = 0.5  Ft Dir Right = 0.5  Ft Dir Left = 0.5  Bounded = 0.5  Bounded-Moraic = 0.5  Bounded-2 = 0.5  Bounded-Moraic-2 = 0.5  Bounded-Syl = 0.5  Ft Dir Right = 0.5  Ft Dir Left = 0.5  Bounded = 0.5  Bounded-Moraic = 0.5  Bounded-2 = 0.5  Bounded-Moraic-2 = 0.5  Bounded-Syl = 0.5  Ft Dir Right = 0.5  Ft Dir Left = 0.5

Initially all are equiprobable

---

Case study: English metrical phonology

Adult English system values:

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

Estimate of child input: caretaker speech to children between the ages of 6 months and 2 years (CHILDES) (Brent & Bernsman-Right corpora; MacWhinney 2000)

Total Words: 540505  Mean Length of Utterance: 3.5

Words parsed into syllables using the MRC Psycholinguistic database (Wilson, 1998) and assigned likely stress contours using the American English CALL HOME database of telephone conversation (Canavan et al., 1997)

---

English Data

Non-trivial language: English (full of exceptions)

Noisy data: 27% incompatible with correct English grammar on at least one parameter value

Hard - therefore interesting!

Exceptions:

- QI, QSVCL, Em-None, Ft Dir Left, Unbounded, Bounded-3, Bounded-Moraic, Ft Hl Right
Probabilistic learning for English

Probabilistic generation and testing of parameter values (Yang 2002)

The learner then uses this grammar to generate a stress contour for the observed data point.

If the generated stress contour matches the observed stress contour, the grammar successfully "parses" the data point. All participating parameter values are rewarded.

If the generated stress contour does not match the observed stress contour, the grammar does not successfully "parse" the data point. All participating parameter values are punished.

Parameter values $v_1$ vs. $v_2$

- $p_0 = p_0 - (1-p_0)$
- $p_0 = (1-p_0)$
- reward $v_1$
- punish $v_1$

After learning: expect probabilities of parameter values to converge near endpoints (above/below some threshold).

QI = 0.3
OSVCL = 0.8
Em-Some = 0.1
... = Em-None = 0.9

BayesLearner: Bayesian update of binomial distribution (Chao 1991)

Parameter value $v$

- $P = 1 + \text{success}$
- $P = 0 + \text{error}$
- reward success $+ 1$
- punish success $+ 0$
Probabilistic learning for English

Probabilistic generation and testing of parameter values (Yang 2002)

Update parameter value probabilities

After learning: expect probabilities of parameter values to converge near endpoints (above/below some threshold).

\[ \text{QI} = 0.3 \quad \text{QSVCL} = 0.6 \quad \text{Em-None} = 0.1 \]

Once set, a parameter value is always used during generation, since its probability is 1.0.

\[ \text{QI = 0.3 \quad QSVCL = 0.6 \quad Em-None = 0.1} \]

\[ \text{QP: if QS, QSVCL or QSVCH?} \]

Note: total data seen + 1 counter for p

Invoke when

Examples of incorrect target grammars

BayesLearner: Bounded-Syllabic

NParLearner: Bounded

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

Goal: Converge on English values after learning period is over

QI, QSVCH, Em-None, Em-Right, Ft Dir Right, Bounded, Bounded-2, Bounded-Syllabic, Ft Hd Left

Probabilistic learning for English: Modifications

Probabilistic generation and testing of parameter values (Yang 2002)

Update parameter value probabilities

Batch-learning (for very small batch sizes): smooth out some of the irregularities in the data

Implementation (Yang 2002):

Success = increase parameter value’s batch counter by 1

Failure = decrease parameter value’s batch counter by 1

Invoke update procedure (Linear Reward-Penalty or Bayesian Updating) when batch limit \( b \) is reached. Then, reset parameter’s batch counters.

Parameter values v1 vs. v2

\[ p_1 = p_1 \cdot (1 - p_0) \quad p_0 = (1 - p_1) \]

\[ p_1 = 1 - p_1 \quad p_0 = 1 - p_0 \]

\[ \text{reward v1} \quad \text{punish v1} \]

BayesLearner: Bayesian update of binomial distribution (Chow 1961)

Invoke when the batch counter for \( p_{v1} \) or \( p_{v2} \) equals \( b \)

Note: total data seen + 1

\[ \text{Parameter value v} \]

\[ \text{Ft Dir Right, Bounded, Bounded-2, Bounded-Syllabic, Ft Hd Left} \]
Probabilistic learning for English

Goal: Converge on English values after learning period is over

<table>
<thead>
<tr>
<th>Model</th>
<th>Success rate (1000 runs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NParLearner, 0.01 ≤ b ≤ 0.05</td>
<td>1.2%</td>
</tr>
<tr>
<td>BayesLearner</td>
<td>0.0%</td>
</tr>
<tr>
<td>NParLearner + Batch, 0.01 ≤ b ≤ 0.05, 2 ≤ b ≤ 10</td>
<td>0.8%</td>
</tr>
<tr>
<td>BayesLearner + Batch, 2 ≤ b ≤ 10</td>
<td>1.0%</td>
</tr>
</tbody>
</table>

Probabilistic learning for English: Modifications

Probabilistic generation and testing of parameter values (Yang 2002)

Human infants may already have knowledge of FT-HD Left (Jusczyk, Cutler, & Redarz 1993) and QS (Turk, Jusczyk, & Gerken 1995).

Build this bias into a model: set probability of QS = FT-HD Left = 1.0. These will always be chosen during generation.

QS, QSVCL, or QSVCH?

Update parameter value probabilities + Batch Learning

Probabilistic learning for English

Goal: Converge on English values after learning period is over

<table>
<thead>
<tr>
<th>Model</th>
<th>Success rate (1000 runs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NParLearner, 0.01 ≤ b ≤ 0.05</td>
<td>1.2%</td>
</tr>
<tr>
<td>BayesLearner</td>
<td>0.0%</td>
</tr>
<tr>
<td>NParLearner + Batch, 0.01 ≤ b ≤ 0.05, 2 ≤ b ≤ 10</td>
<td>0.8%</td>
</tr>
<tr>
<td>BayesLearner + Batch, 2 ≤ b ≤ 10</td>
<td>1.0%</td>
</tr>
</tbody>
</table>

Probabilistic learning for English

Goal: Converge on English values after learning period is over

<table>
<thead>
<tr>
<th>Model</th>
<th>Success rate (1000 runs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NParLearner, 0.01 ≤ b ≤ 0.05</td>
<td>1.2%</td>
</tr>
<tr>
<td>BayesLearner</td>
<td>0.0%</td>
</tr>
<tr>
<td>NParLearner + Batch, 0.01 ≤ b ≤ 0.05, 2 ≤ b ≤ 10</td>
<td>0.8%</td>
</tr>
<tr>
<td>BayesLearner + Batch, 2 ≤ b ≤ 10</td>
<td>1.0%</td>
</tr>
</tbody>
</table>

Probabilistic learning for English

Goal: Converge on English values after learning period is over

<table>
<thead>
<tr>
<th>Model</th>
<th>Success rate (1000 runs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NParLearner, 0.01 ≤ b ≤ 0.05</td>
<td>1.2%</td>
</tr>
<tr>
<td>BayesLearner</td>
<td>0.0%</td>
</tr>
<tr>
<td>NParLearner + Batch, 0.01 ≤ b ≤ 0.05, 2 ≤ b ≤ 10</td>
<td>0.8%</td>
</tr>
<tr>
<td>BayesLearner + Batch, 2 ≤ b ≤ 10</td>
<td>1.0%</td>
</tr>
</tbody>
</table>

Where else can we modify?

1. Hypothesis space
2. Data
3. Update procedure

The best isn't so great
Where else can we modify?

1. Hypothesis space
   - Prior knowledge, biases: QS, F1, HD Left known...

2. Data
   - What about the data the learner uses?

3. Update procedure
   - Linear Reward-Penalty, Bayesian, Batch...

Data Intake Filtering

“Selective Learning”

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

“Selective” Intuition: Use the really good data only.

One instantiation of “really good” = highly informative.

One instantiation of “highly informative” = data viewed by the learner as unambiguous (Fodor, 1998; Dreher, 1999; Lightfoot, 1999; Pearl & Weinberg, 2007)
Practical matters: Feasibility of unambiguous data

Existence?

"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."

Identification?

Even if unambiguous data existed, how could a child identify them?

Practical matters: Feasibility of unambiguous data

Existence?

Depends on data set (empirically determined).

Identification?

Identifying unambiguous data:

Dresher (1999); Lightfoot (1999): heuristic pattern-matching to observable form of the data.

Cues are available for each parameter value, known already by the learner.

Parsing (Fodor 1998; Sakas & Fodor 2001): extract necessary parameter values from all successful parses of data point.

Both operate over a single data point at a time: compatible with incremental learning.

Probabilistic learning from unambiguous data

Each parameter has 2 values.

(Pearl 2008)
Probabilistic learning from unambiguous data

Each parameter has 2 values.

Advantage in data: How much more unambiguous data there is for one value over the other in the data distribution.

Assumption (Yang 2002): The value with the greater advantage will be the one a probabilistic learner will converge on over time.

Allows us to be fairly agnostic about the exact nature of the probabilistic learning, provided it has this behavior.

Probabilistic learning from unambiguous data

The order in which parameters are set may determine if they are set correctly from the data.

Dresher 1999

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.

Probabilistic learning from unambiguous data

Dresher 1999

Success guaranteed as long as parameter-setting order constraints are followed.

Parsing
Group 1: QS, FHLd, Left, Bounded
Group 2: Rr, Dir Right, QS/VC-Heavy
Group 3: Em-Some, Em-Right, Bounded-2, Bounded-Syl

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

Road Map

Introduction to complex linguistic systems
- General problems
- Parametric systems
- Parametric metrical phonology

Learnability of complex linguistic systems
- General learnability framework
- Case study: English metrical phonology
- Available data & associated woes
- Unconstrained probabilistic learning
- Constrained probabilistic learning

Where next? Implications & Extensions

Where we are now

Cognitive modeling: aimed at understanding how humans solve problems, generating human behavior by using psychologically plausible methods.

Language: learning complex systems is difficult. Success comes from integrating biases into probabilistic learning models.

Bias on hypothesis space: linguistic parameters already known, some values already known

Bias on data: interpretive bias to use highly informative data

0.7
0.3
0.5
0.5
0.8
0.2

Where we can go

(1) Interpretive bias:
- How successful on other difficult learning cases (noisy data sets, other complex systems)?
- Are there other methods of implementing interpretative biases that lead to successful learning?
- How necessary is an interpretive bias? Are there cleverer probabilistic learning methods than can succeed?

+ biases?
Where we can go
(1) Interpretive bias:
   How successful on other difficult learning cases (noisy data sets, other complex systems)?
   Are there other methods of implementing interpretive biases that lead to successful learning?
   How necessary is an interpretive bias? Are there cleverer probabilistic learning methods than can succeed?

(2) Hypothesis space bias:
   Is it possible to infer the correct parameters of variation given less structured information a priori (e.g. larger units than syllables are required)? [Model Selection]

(3) Informing AI/ML:
   Can we import the necessary biases for learning complex systems into language applications (e.g. speech generation)?

The big idea
Complex linguistic systems may well require something beyond probabilistic methods in order to be learned, and learned as well as humans learn them.

What this likely is: learner biases in hypothesis space and data intake (how to deploy probabilistic learning)

What we can do: take insights from cognitive modeling and apply them to problems in artificial intelligence and machine learning, & vice versa

Thank You
Amy Weinberg
Bill Idsardi
Bill Sakas
Janet Fodor

The audiences at
University of California, Los Angeles Linguistics Department
University of Southern California Linguistics Department
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