Computational Modeling for Language Acquisition Research

John Nerbonne
j.nerbonne@rug.nl

Center for Language and Cognition
University of Groningen

November 2011
The puzzle to solve

ljuuzuibutsjhiuljuuz
ljuuztbzjubhbjompwfljuuz
xibutuibu
ljuuz
epzpvxbounpsfnjmlipofz
ljuuzljuuzephhjf
opnjaxibuepftbljuuztbz
xibuepftbljuuuztbz
ephhjfe
eph
opnjaxibuepftuifephhjftbz
xibuepftuifephhjftbz
mjuumfcbczczcjejf
cbczcjsejf
zpvepoumjlfuibupof
plbznppnzublfuijtpvu
dpx
uifdpxtbztnppnp
xibuepftuifdpxtbzpnpn
The puzzle to solve

Children need to:

- segment the input to linguistic units (words, endings, etc.).
- assign meanings to these units.
- figure out which combinations of these units are acceptable in the language.
- ...
The Problem of Language Acquisition

Some observations:

▶ Human languages are complex.
▶ Language acquisition by children is (arguably) fast and robust.
▶ Children do not receive explicit instruction during language acquisition.
▶ Children do not receive negative input (at least not with respect to form), e.g., by means of corrections. Even if they do, they seem to ignore it.
The Problem of Language Acquisition

Some observations:

▶ Human languages are complex.
▶ Language acquisition by children is (arguably) fast and robust.
▶ Children do not receive explicit instruction during language acquisition.
▶ Children do not receive negative input (at least not with respect to form), e.g., by means of corrections. Even if they do, they seem to ignore it.

An argument:

▶ The input children receive is not enough for learning (Argument from Poverty of Stimulus, APS).
  ▶ Human languages are not learnable from positive input (i.e., without corrections).
  ▶ Children do not receive certain type of input critical for learning certain phenomena.
Computational Models of Language Acquisition

Answers to the questions on language acquisition should eventually come from neuroscience. But we seem to be far from this yet. Meanwhile...

- Experimental work (developmental psycholinguistics), investigates the behavioral aspects of language acquisition.
- Formal learning theory provides mathematical/formal methods to investigate the limits of learnability.
- Computational simulations offer methods to tests for these theories under different conditions, some of which are not possible/practical to test experimentally.
Answers to the questions on language acquisition should eventually come from neuroscience. But we seem to be far from this yet. Meanwhile...

- Experimental work (developmental psycholinguistics), investigates the behavioral aspects of language acquisition.
- Formal learning theory provides mathematical/formal methods to investigate the limits of learnability.
- Computational simulations offer methods to test these theories under different conditions, some of which are not possible/practical to test experimentally.

In this lecture, we are interested in last two (complementary and interacting) methods of studying language acquisition. These methods are particularly useful for testing claims of APS.
Computational Models on Language Acquisition Debate

▶ **Nativism**: Language acquisition is guided by a Universal Grammar, consisting of principles and parameters. Learning is achieved by setting a small number of (binary) parameters.

\[
\text{pro-drop} \quad \text{wh} \quad \text{v2}
\]
Computational Models on Language Acquisition Debate

- **Nativism**: Language acquisition is guided by a Universal Grammar, consisting of principles and parameters. Learning is achieved by setting a small number of (binary) parameters.

\[
\begin{align*}
\text{pro-drop} & \quad \text{wh} & \quad \text{v2} \\
- & \quad + & \quad + & \rightarrow & \text{Dutch}
\end{align*}
\]
Computational Models on Language Acquisition Debate

▶ **Nativism**: Language acquisition is guided by a **Universal Grammar**, consisting of principles and parameters. Learning is achieved by setting a small number of (binary) parameters.

<table>
<thead>
<tr>
<th>pro-drop</th>
<th>wh</th>
<th>v2</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>
Computational Models on Language Acquisition Debate

- **Nativism**: Language acquisition is guided by a *Universal Grammar*, consisting of principles and parameters. Learning is achieved by setting a small number of (binary) parameters.

<table>
<thead>
<tr>
<th>pro-drop</th>
<th>wh</th>
<th>v2</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Computational Models on Language Acquisition Debate

- **Nativism**: Language acquisition is guided by a **Universal Grammar**, consisting of principles and parameters. Learning is achieved by setting a small number of (binary) parameters.

\[
\text{pro-drop} \quad \text{wh} \quad \text{v2} \\
- \quad + \quad + \quad \rightarrow \quad \text{Dutch} \\
- \quad + \quad - \quad \rightarrow \quad \text{English} \\
+ \quad - \quad - \quad \rightarrow \quad \text{Turkish}
\]

- **Empiricism**: e.g., **Connectionist models**
  Learning is achieved by general purpose learning algorithms, e.g., backpropagation.
Language acquisition debate: summary

**Ground rules:**

- *There must be some innate component:*
  - The child born in the same household learns the language, but the kitten does not.
  - No free lunch theorem: we know from the machine learning theory that there is no universal learning algorithm.

- *Learning is a part of the language acquisition:* children learn the language(s) spoken in their environment, not a universal language.

The main dispute is on the nature of the innate component and the learning mechanisms, either they are language specific, or general cognitive mechanisms.
Overview

- The problem of language acquisition.
- **Computational Learning Theory & Computational Simulations**
- Example 1: learning segmentation.
- Example 2: learning syntax.
Complexity of (formal) Languages: Chomsky hierarchy

Mathematical studies show that we can define language classes in terms of (i) types of rules (grammars); (ii) type of sequence sets (sentences); and even types of computations (vis-à-vis memory requirements).
Complexity of (formal) Languages: Chomsky hierarchy

Mathematical studies show that we can define language classes in terms of (i) types of rules (grammars); (ii) type of sequence sets (sentences); and even types of computations (vis-à-vis memory requirements).

Human language syntax seems to require slightly more expressive power than context-free languages.
How to formalize language acquisition?

- Define the language class.
How to formalize language acquisition?

- Define the language class.
- Define the input.
How to formalize language acquisition?

- Define the **language class**.
- Define the **input**.
How to formalize language acquisition?

- Define the language class.
- Define the input.
- Define when it’s learned.
How to formalize language acquisition?

- Define the language class.
- Define the input.
- Define when it’s learned.
How to formalize language acquisition?

- Define the **language class**.
- Define the **input**.
- Define when it’s learned.
How to formalize language acquisition?

- Define the language class.
- Define the input.
- Define when it’s learned.
Identification in the Limit (Gold 1967)

- **The learning task**: Finding a grammar that identifies the target language in the relevant class of languages.

- **Input**: A sequence of sentences of the target language, in any possible order. Every sentence of the target language appears at least once in the input sequence.

- **The learner** is assumed to be an effective learner: After every input, the learner picks a language that is consistent with the input so far. The details of the algorithm are irrelevant.

- **Learning criterion**: The learner converges to the target grammar, and sticks to it indefinitely.

A class of languages is **identifiable in the limit** if there is a learner that meets the learning criterion with finite amount of input.
Learnability Results

▶ None of the language classes in the Chomsky hierarchy is identifiable in the limit.
▶ With positive and negative input, all classes in the Chomsky hierarchy is learnable.
▶ Putting bounds on the grammar makes most of these classes learnable.
▶ A large number of positive and negative results are found after Gold (1967), but linguistic (and language acquisition) literature seems to ignore it.
# IIL and Language Acquisition

<table>
<thead>
<tr>
<th></th>
<th>Identification in the limit</th>
<th>Language acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Language Class</strong></td>
<td>Formal Languages</td>
<td>Human Languages</td>
</tr>
<tr>
<td><strong>Learning criterion</strong></td>
<td>Identify the target language (TL)</td>
<td>?</td>
</tr>
<tr>
<td><strong>Amount of Data &amp; resources</strong></td>
<td>As much as needed</td>
<td>Limited by human learning conditions</td>
</tr>
<tr>
<td><strong>Input Data</strong></td>
<td>Any sequence of sentences from the TL</td>
<td>Child-directed speech</td>
</tr>
</tbody>
</table>
## IIL and Language Acquisition

<table>
<thead>
<tr>
<th></th>
<th>Identification in the limit</th>
<th>Language acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning criterion</td>
<td>Identify the target language (TL)</td>
<td>?</td>
</tr>
<tr>
<td>Amount of Data &amp; resources</td>
<td>As much as needed</td>
<td>Limited by human learning conditions</td>
</tr>
<tr>
<td>Language Class</td>
<td>Formal Languages</td>
<td>Human Languages</td>
</tr>
<tr>
<td>Input Data</td>
<td>Any sequence of sentences from the TL</td>
<td>Child-directed speech</td>
</tr>
</tbody>
</table>

Particularly for modeling input, computational simulations seem to provide a better fit than purely mathematical studies.
Computational Models for Language Acquisition

Computational models of human language acquisition has to meet some criteria that is not always applicable for engineering oriented CL applications.

- Modes should use realistic input, such as naturally occurring child directed speech.
- Any additional source of information, or heuristics should be justifiable.
- Learning should proceed on-line: models should not require all the input data available at once.
- Models should not pose unrealistic bounds on memory and computation resources.
- The assumptions and predictions of the model should match (at least should not conflict with) psycholinguistic evidence.
Overview

- The problem of language acquisition.
- Computational Learning Theory & Computational Simulations
- Example: learning segmentation.
Segmentation: back to the puzzle

ljuuzuibutsjhiuljuuz
ljuuztbzjubhbjompwfljuuz
xibutuibu
ljuuz
epzpvxbounpsfnnjmlipofz
ljuuzljuuzephhjf
opnjxibuepftbljuuztbz
xibuepftbljuuztbz
ephhjfe
ephhjf
opnjxibuepftuifephhjftbz
xibuepftuifephhjftbz
mjuumfcbczczcjej
cbcncjczej
zpvepoumjlfuibupof
plbznppnnzublfuijtpvu
dpx
ufdpxtbztznppnp
xibuepftuifdpxtbzoij
Segmentation: back to the puzzle

Children need to segment the input (divide it into words).

Despite:

- no reliable indicators to boundaries,
- variation in pronunciation,
- noise (mistakes, non-words).
Lexical knowledge (knowledge about words) helps

But...

▶ Even a complete lexicon does not solve the segmentation problem completely
Lexical knowledge (knowledge about words) helps

But...

- Even a complete lexicon does not solve the segmentation problem completely

*Example re-produced from: (Shillcock, 1995)*
Lexical knowledge (knowledge about words) helps

But...

- Even a complete lexicon does not solve the segmentation problem completely
- To acquire a lexicon, you need to extract word from continuous speech (need segmentation).

*Example re-produced from: (Shillcock, 1995)*
Cues for segmentation

- Lexical knowledge.
- Stress pattern of the words.
- Slight differences in pronunciation
- Distributional regularities / phonotactic constraints.
  —We’ll look at what this is next!
Cues for segmentation

- Lexical knowledge.
- Stress pattern of the words.
- Slight differences in pronunciation
- **Distributional regularities** / phonotactic constraints.
  — We’ll look at what this is next!
Distributional regularities

*Predictability within units is high, predictability between units is low.*
Distributional regularities

*Predictability within units is high, predictability between units is low.*

An old method using distributional regularities is suggested by (Haris, 1955): *The boundaries (between words or words and endings) occur where a large variety of possible sounds (letters) can follow.*

read- readi-
Distributional regularities

*Predictability within units is high, predictability between units is low.*

An old method using distributional regularities is suggested by (Haris, 1955): *The boundaries (between words or words and endings) occur where a large variety of possible sounds (letters) can follow.*

\[
\text{read}-\{\text{a,e,i,j,o,s,y,-,',$}\} \quad (10) \quad \text{readi}-\{\text{e, l, n}\} \quad (3)
\]

read  
reads  
reading  
reader  
readjusted  
...

There are a number of ways to formalize the concept: including conditional probability, entropy, mutual information...
**Distributional regularities**

*Predictability within units is high, predictability between units is low.*

An old method using distributional regularities is suggested by (Haris, 1955): *The boundaries (between words or words and endings) occur where a large variety of possible sounds (letters) can follow.*

\[
\text{read-}\{a,e,i,j,o,s,y,-,\}$ (10) \quad \text{readi-}\{e, l, n\} (3)
\]

read, reads, reading, reader, readjusted, 
readied, readily

...  

There are a number of ways to formalize the concept: including conditional probability, entropy, mutual information...
Segmentation: back to the puzzle

\[(u | j) = 11.4 \quad P(u | z) = 2.08 \]

John Nerbonne, CLCG
Segmentation: back to the puzzle

\[
P(u | j) = 11^{27} = 0.44
\]
\[
P(u | z) = 2^{23} = 0.08
\]
Segmentation: back to the puzzle

ljuuzuibutsjhiuljuuz
ljuutztbzuhibbjompwfljuuz
xibutuibu
ljuuz
epzpvxbounpsfnjmlipozf
ljuuzljuuzepphhjf
opnjxibuepftbljuuztbz
xibuepftbljuuztbz
ephhjfeiph

ephjjf
opnjxibuepftuifephhjjftbz
xibuepftuifephhjjftbz
mjuumfcbczczcjsejf
cbczczcjsejf
zpvepoumlfluibupofof
plbznppnzu blfuijtpvu
dpx
uifdpxtbzttnppnpp
xibuepftuifdpx tbzopnj
Segmenntation: back to the puzzle

\[ P(u|j) = \frac{11}{27} = 0.4 \quad P(u|z) = \frac{2}{23} = 0.08 \]
Segmentation puzzle: a solution

The boundaries are inserted using the method suggested by Harris (where successor or predecessor values peak).
Predictability based models: psychological relevance

Children very early in life (8-months) seem to be sensitive to this type of information in the speech (Saffran, Aslin, Newport 1996)

- Infants are habituated to artificial speech segments built from a simple vocabulary.
- They are tested with non-familiar patterns and familiar patterns.
- On the basis of very short training 8-month-old infants attended familiar examples significantly longer than the unfamiliar ones.
Summary

- Computational models are useful, especially for investigating characteristics of input. But, they can tell us more.
- Formal learning theory provides useful formal tools to assess learnability of well-defined models.
- Computational simulations forces us to formalize our intuitions. Only then, we can actually assess validity of our intuitions.
More reading

References

More on learnability debate

Example P&P model

Example Connectionist model