Exploring the assumptions of language acquisition models

Christos Christodouloupolous, Cynthia Fisher and Dan Roth

Midwest Speech & Language Days 2015
Models of language acquisition

“The girl chases the boy”
Models of language acquisition

“The girl chases the boy”
Models of language acquisition

“The girl chases the boy”
Models of language acquisition

“The girl chases the boy”
“The boy runs”
PropBank corpus
[Palmer et al. 2005]

Core arguments:
- **A0** - Agent
- **A1** - Patient
- **A2** - Recipient

... 

Modifiers:
- Locative
- Temporal
- Manner

... 

“The girl chases the boy”

**A0** pred **A1**
BabySRL [Connor et al. 2008; 2010]

BabySRL corpus
Adam, Eve, Sarah [Brown, 1973]

“The girl chases the boy”
A0  pred  A1
BabySRL [Connor et al. 2008; 2010]

BabySRL corpus
Adam, Eve, Sarah [Brown, 1973]
Adult utterances (cleaned up)

“The girl chases the boy”
A0  pred  A1
BabySRL [Connor et al. 2008; 2010]

BabySRL corpus
Adam, Eve, Sarah [Brown, 1973]
Adult utterances (cleaned up)
Focus on verb predicates

“The girl chases the boy”
A0 pred A1
BabySRL [Connor et al. 2008; 2010]

BabySRL corpus

Adam, Eve, Sarah [Brown, 1973]
Adult utterances (cleaned up)
Focus on verb predicates
1 verb 2 args (24% of sent.)

"The girl chases the boy"

A0  pred  A1
Experiment 1: Supervised learning

Given **perfect feedback**, do simple, **bottom-level** features capture anything useful about semantic roles/verb preferences?
Experiment 1: Supervised learning

• Supervised classifier (average perceptron)
  • LBJava [Rizzolo and Roth, 2010]
• Train on BabySRL corpus
• Test on novel verb sentences
Experiment 1: Supervised learning

- Supervised classifier (average perceptron)
  - LBJava [Rizzolo and Roth, 2010]
- Train on BabySRL corpus
- Test on novel verb sentences
  - Intransitive: “The bunny krads”
  - Transitive: “The boy krads the girl”
  - Ditransitive: “The girl krads the boy a bunny”
Experiment 1: Features

- **Most frequent label**

  A0
  The girl chases the boy

  A1
Experiment 1: Features

- **Most frequent label**

- **Lexical features**
  
  **A0**
  The girl chases the boy
  
  **A1**
  chase-girl  chase-boy
Experiment 1: Features

- **Most frequent label**
- **Lexical features**
- **Noun Pattern**

<table>
<thead>
<tr>
<th>A0</th>
<th>A1</th>
</tr>
</thead>
<tbody>
<tr>
<td>The girl chases the boy</td>
<td>chase-boy</td>
</tr>
<tr>
<td>chase-girl</td>
<td>1st of 2</td>
</tr>
</tbody>
</table>
| 2nd of 2 | }
Experiment 1: Features

- **Most frequent** label
- **Lexical features**
- **Noun Pattern**
- **Verb Position**

```
A0
The girl chases the boy
chase-girl
1st of 2
Before

A1
chase-boy
2nd of 2
After
```
Experiment 1: Results

The bar chart represents the results of Experiment 1, showing the frequency of different categories. The categories are labeled as follows:

- Most Freq.
- Lex
- +NPat
- +VPos
- +NPat & VPos

The chart uses different colors to represent these categories:

- Red: A1A1
- Orange: A0A0
- Green: A0A2
- Blue: A0A1
Experiment 1: Results

The bar chart shows the distribution of results across different categories: Most Freq., Lex, +NPat, +VPos, and +NPat & VPos. The categories are color-coded as follows:

- A1A1 (red)
- A0A0 (orange)
- A0A2 (green)
- A0A1 (blue)

The chart indicates a significant concentration of results in the Lex category, with the A1A1 category dominating in terms of count.
Experiment 1: Results

![Bar chart showing the results of Experiment 1. The chart compares categories like Most Freq., Lex, +NPat, +VPos, and +NPat & VPos. The categories are distinguished by color: A1A1 (red), A0A0 (orange), A0A2 (green), and A0A1 (blue).]
Experiment 1: Results

- Most Freq.
- Lex
- +NPat
- +VPos
- +NPat & VPos

Legend:
- A1A1
- A0A0
- A0A2
- A0A1
Experiment 1: Results
Experiment 1: Results

![Bar chart showing results for different conditions with category labels and data points: Most Freq., Lex, +NPat, +VPos, and +NPat & VPos. The chart indicates variations in knowledge with colored bars for different conditions.]

Predicate knowledge
Multiple predicates

“Remember how we play the surprise game?”

“Remember how we play the surprise game?”
Multiple predicates

“Remember how we play the surprise game?”

“Remember how we play the surprise game?”

<table>
<thead>
<tr>
<th></th>
<th># sent</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 verb</td>
<td>10,356</td>
<td>69.86</td>
</tr>
<tr>
<td>2 verbs</td>
<td>3,614</td>
<td>24.38</td>
</tr>
</tbody>
</table>
Effect of multiple predicates (Noun Pattern)
Effect of multiple predicates (Noun Pattern)
Effect of multiple predicates (Noun Pattern)
Effect of multiple predicates (Noun Pattern)

- All
- First
- Last

Syntactic surface the same e.g. NPat: 4th out of 5
Experiment 1: Supervised learning

Given **perfect feedback**, do simple, **bottom-level** features capture anything useful about semantic roles/verb preferences?
Experiment 1: Supervised learning

Given **perfect feedback**, do simple, **bottom-level**
features capture anything useful about semantic
roles/verb preferences?

Yes, but predicate knowledge is crucial
Experiment 2: Unsupervised learning

Can we predict arguments/predicates using distributional \textit{clusters} and a few \textit{seed nouns}?
Experiment 2: Unsupervised learning

Can we predict arguments/predicates using distributional clusters and a few seed nouns?

Syntactic Bootstrapping via Structure-Mapping

[Gleitman, 1990; Fisher et al. 2010]
Experiment 2: Unsupervised learning

- HMM over 2.2M tokens (CHILDES)
  - 80 induced clusters, list of function words
- List of seed nouns [Dale and Fenson, 1996]
- Noun identification

  “Cluster contains more than $k$ seed nouns”
## Experiment 2: Verb Identification

<table>
<thead>
<tr>
<th></th>
<th>She</th>
<th>krads</th>
<th>a</th>
<th>red</th>
<th>truck</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMM</td>
<td>45</td>
<td>51</td>
<td>19</td>
<td>60</td>
<td>73</td>
</tr>
<tr>
<td>N Ident.</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Funct.</td>
<td></td>
<td></td>
<td>F</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Experiment 2: Verb Identification

<table>
<thead>
<tr>
<th></th>
<th>She</th>
<th>krads</th>
<th>a</th>
<th>red</th>
<th>truck</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMM</td>
<td>45</td>
<td>51</td>
<td>19</td>
<td>60</td>
<td>73</td>
</tr>
<tr>
<td>N Ident.</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Funct.</td>
<td></td>
<td>F</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Experiment 2: Verb Identification

<table>
<thead>
<tr>
<th>N Ident.</th>
<th>FUNCT.</th>
<th>She</th>
<th>klands</th>
<th>a</th>
<th>red</th>
<th>truck</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMM</td>
<td></td>
<td>45</td>
<td>51</td>
<td>19</td>
<td>60</td>
<td>73</td>
</tr>
<tr>
<td>N Ident.</td>
<td></td>
<td>N</td>
<td></td>
<td></td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>FUNCT.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bar chart showing frequency distribution for different argument counts:
- 0 args
- 1 arg
- 2 args
- 3 args

Values: 51 and 60
Experiment 2: Verb Identification

<table>
<thead>
<tr>
<th></th>
<th>She</th>
<th>krads</th>
<th>a</th>
<th>red</th>
<th>truck</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMM</td>
<td>45</td>
<td>51</td>
<td>19</td>
<td>60</td>
<td>73</td>
</tr>
<tr>
<td>N Ident.</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Funct.</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Graph showing distribution of arguments:
- 0 args
- 1 arg
- 2 args
- 3 args

Number of arguments for words:
- 51: 3 args
- 60: 2 args
Experiment 2: Results
Experiment 2: Results
Experiment 2: Results

![Graph showing results of Experiment 2 with categories: arg-F, verb-F, and verbRand-F. The graph plots values from 0 to 1 at intervals of 0.25, with x-axis values from 1 to 73.]
Experiment 2: Parameters

- Random/frequent seed noun selection
- Variants + plurals of seed nouns
- Verb/predicate evaluation
- Multiple predicates
- Seed noun threshold $k$
- Null predictions
- Function words
Experiment 2: Parameters

- Random/frequent seed noun selection
- Variants + plurals of seed nouns
- Verb/predicate evaluation
- Multiple predicates
- Seed noun threshold $k$
- Null predictions
- Function words

Experiment 2:

- Parameters

- $0$ $0.2$ $0.4$ $0.6$ $0.8$ $1$
- $1$ $25$ $49$ $73$
Experiment 2: Parameters

- Random/frequent seed noun selection
- Variants + plurals of seed nouns
- Verb/predicate evaluation
- Multiple predicates
- Seed noun threshold \( k \)
- Null predictions
- Function words

Graph showing the progression of verb, verb FREQ, verbRand, and verbRand FREQ over different values of a parameter.
Experiment 2: Parameters

- Random/frequent seed noun selection
- Variants + plurals of seed nouns
- Verb/predicate evaluation
- Multiple predicates
- Seed noun threshold $k$
- Null predictions
- Function words
Experiment 2: Parameters

- Random/frequent seed noun selection
- Variants + plurals of seed nouns
- Verb/predicate evaluation
- Multiple predicates
- Seed noun threshold $k$
- Null predictions
- Function words

\[
\begin{align*}
\text{Freq} & : \text{verb} \quad \text{verbRand} \\
\text{Freq + Var} & : @24 \text{ seed nouns}
\end{align*}
\]
Experiment 2: Parameters

- Random/frequent seed noun selection
- Variants + plurals of seed nouns
- Verb/predicate evaluation
- Multiple predicates
- Seed noun threshold $k$
- Null predictions
- Function words

![Graph showing comparison between verb and verbRand]

@24 seed nouns
Experiment 2: Parameters

- Random/frequent seed noun selection
- Variants + plurals of seed nouns
- Verb/predicate evaluation
- Multiple predicates
- Seed noun threshold \( k \)
- Null predictions
- Function words

![Graph showing comparison between 'verb' and 'verbRand' under different thresholds.](image-url)
Experiment 2: Parameters

- Random/frequent seed noun selection
- Variants + plurals of seed nouns
- Verb/predicate evaluation
- Multiple predicates
- Seed noun threshold $k$
- Null predictions
- Function words

![Graph showing performance for relaxed and strict conditions with 24 seed nouns.](image-url)
Experiment 2: Unsupervised learning

Can we predict arguments/predicates using distributional **clusters** and a few **seed nouns**?
Experiment 2: Unsupervised learning

Can we predict arguments/predicates using distributional **clusters** and a few **seed nouns**?

Yes, with as few as 24 seed nouns
Experiment 2: Unsupervised learning

Can we predict arguments/predicates using distributional clusters and a few seed nouns?

Yes, with as few as 24 seed nouns need to consider multiple predicates
Conclusions

• BabySRL model of language acquisition
  • Evidence for syntactic bootstrapping
• Exploration of assumptions
  • Data representation
  • Evaluation
  • Psycholinguistic validity
Future Directions

• BabySRL from scratch [Connor et al. 2012]

• Beyond single predicates
  • Multiple verbs
  • Prepositions

• Relaxing perfect feedback (scene ambiguity)
  • Superset
  • Bootstrapped Animacy
Future Directions

• BabySRL from scratch [Connor et al. 2012]

• Beyond single predicates
  • Multiple verbs
  • Prepositions

• Relaxing perfect feedback (scene ambiguity)
  • Superset
  • Bootstrapped Animacy

Thanks