Bootstrapping into Filler-Gap: An Acquisition Story

Marten van Schijndel
Department of Linguistics
The Ohio State University

November 22, 2013
**Background**

**Filler-Gap**
A non-local dependency that potentially spans an unbounded # of lexemes.

e.g. That’s \{the ball\} John kicked \__\__

e.g. That’s \{the ball\} Mary said John kicked \__\__

This is hard because:

- Filler must be remembered
- Where is the gap?
Motivation

How could children learn this?

Goal

- Simplest model of filler-gap?
# Background

## Psychology

Children can’t use filler-gap until 5 years
[de Villiers and Roeper, 1995]

## Computational Linguistics

An uncommon phenomenon that doesn’t boost performance much
[Rimell et al., 2009, Nivre et al., 2010, Nguyen et al., 2012]
Experimental Results

[Seidl et al., 2003]
Preferential looking paradigm

**Wh-**

Wh-S: What hit the apple?
Wh-O: What did the flower hit?

**Control**

Where is the flower?
Acquisition Pattern?

Developmental timeline of wh- question comprehension (13, 15, 20)

[Seidl et al., 2003]
Acquisition Pattern

Developmental timeline of wh- question comprehension (15, 20)
Parentheses = marginal comprehension
[Gagliardi et al., 2011]
Model Motivation

What are children learning?

**Complex Grammatical Constraints**

Under certain conditions:
- Arguments may occur in non-canonical syntactic positions.
  - e.g., questions introduce an expected future gap (SLASH, A-bar).

**Different Possible Orderings**

- The flower **hit** the **apple**.
- What **hit** the **apple**.
- What did the flower **hit**?
Model Motivation

Different Word Orderings

- **SOV**: Japanese, Hindi, German
- **SVO**: English, Mandarin, Spanish
- **VSO**: Zapotec, Irish
- **VOS**: Malagasy, Baure
Model Motivation

OT: Different Constraint Orderings

Yield different phonological realizations [Boersma, 1997]
e.g. nasal place assimilation

<table>
<thead>
<tr>
<th></th>
<th>*GESTURE(tip)</th>
<th>*REPLACE(cor)</th>
</tr>
</thead>
<tbody>
<tr>
<td>an+pa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[anpa]</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>[ampa]</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>*REPLACE(cor)</th>
<th>*GESTURE(tip)</th>
</tr>
</thead>
<tbody>
<tr>
<td>an+pa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[anpa]</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>[ampa]</td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>
Model

- Gradual Learning Algorithm [Boersma, 1997]
- Structure mapping: nouns used to learn verbs [Yuan et al., 2012]

Assumptions

- Children can identify nouns [Shi et al., 1998]
- Ns and roles are 1-to-1 [Gertner and Fisher, 2012]
- Abstract factors (#N) are used by learners [Xu, 2002]
- Children are bad at recursion [Diessel and Tomasello, 2001]

Implementation Assumptions

- Distributions are Gaussian
Model
The cat bumped the dog.
Wh-S: Which cat bumped the dog?
Wh-O: Which *cat* did *the dog* bump?*
Initialization 2.0

- Split distributions into mixtures of distributions
  - 1) strong due to canonical evidence
  - 2) weak, but finds arguments from anywhere
Wh-S: Which cat bumped the dog?
Wh-O: Which cat did the dog bump?
With priors, our initial model looks like this.
Evaluation

1. Extract CDS from Eve corpus
   ('you', 'S') ('get', 'V') ('one', 'O') .
   ('what', 'O') are ('you', 'S') ('doing', 'V') ?
   ('you', 'S') ('have', 'V') another cookie right on the table .

2. Chunk nouns (NLTK)
   (N;you)(V;get)(N;one) .
   (N;what)(X;are)(N;you)(V;doing) ?
   (N;you)(V;have)(N;cookie)(X;right)(X;on)(N;table) .

3. Run inference
Inference

Expectation-Maximization

• Estimate labels using distributions over previous observations
• Estimate new distributions using labelled data
• Iterate until converged (≈4 iterations)
Results
Results
Relative Development

[Gagliardi and Lidz, 2010, Gagliardi et al., 2011]

T-Rel
T-S: Show me the dog that bumped the cat.
T-O: Show me the cat that the dog bumped.

W-Rel
Wh-S: Show me the dog who bumped the cat.
Wh-O: Show me the cat who the dog bumped.

Results
• ‘Wh-’ and ‘that’ relative comprehension ∼15 months
• ‘Wh-’ easier than ‘that’
Relative Differences

**That: Confusion with dem/det?**

- That is a book.
- Gimme that!
- Gimme that book!
- Find the cookie that the mouse ate.

**Wh-: Helped by questions?**

- Who kicked the bucket?
- Who did the burglar assault?
- Find the mouse who the cat ate.
## Results: Quantitative

### Overall Accuracy

Arguments correctly labelled

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>.56</td>
<td>.66</td>
<td>.60</td>
</tr>
<tr>
<td>Trained</td>
<td>.54</td>
<td>.71</td>
<td>.61*</td>
</tr>
</tbody>
</table>

Eve (n = 3944)

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>.55</td>
<td>.62</td>
<td>.58</td>
</tr>
<tr>
<td>Trained</td>
<td>.53</td>
<td>.67</td>
<td>.59*</td>
</tr>
</tbody>
</table>

Adam (n = 3622)

* (p < .01)
### Results: Quantitative

#### Agent Prediction

<table>
<thead>
<tr>
<th></th>
<th>Recall</th>
<th></th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>.67</td>
<td>Trained</td>
<td>.65</td>
</tr>
<tr>
<td>Transitive (n = 1000)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### [Connor et al., 2010] (pseudo-comparable)

<table>
<thead>
<tr>
<th></th>
<th>Recall</th>
<th></th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak (10) lexical</td>
<td>.71</td>
<td>Strong (365) lexical</td>
<td>.74</td>
</tr>
<tr>
<td>Gold Args</td>
<td>.77</td>
<td>Gold Args</td>
<td>.58</td>
</tr>
<tr>
<td>Transitive</td>
<td></td>
<td>Intransitive</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Recall</th>
<th></th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak (10) lexical</td>
<td>.59</td>
<td>Strong (365) lexical</td>
<td>.41</td>
</tr>
<tr>
<td>Gold Args</td>
<td>.58</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Marten van Schijndel

Filler-Gap

November 22, 2013 29 / 42
**Results: Quantitative**

But those numbers reflect overall performance...

We can try a coarse filler-gap filter.

**Extract sentences where:**
- O precedes V
- S not immediately followed by V

**Filler-gap Corpora**

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>.53</td>
<td>.57</td>
<td>.55</td>
</tr>
<tr>
<td>Trained</td>
<td>.55</td>
<td>.67</td>
<td>.61*</td>
</tr>
</tbody>
</table>

Eve FG (n = 1345)

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>.53</td>
<td>.52</td>
<td>.52</td>
</tr>
<tr>
<td>Trained</td>
<td>.54</td>
<td>.63</td>
<td>.58*</td>
</tr>
</tbody>
</table>

Adam FG (n = 1287)

* (p < .01)
## Results: Quantitative

Eve FG Corpus

### Subject/Object

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>691</td>
<td>.66</td>
<td>.83</td>
<td>.74</td>
</tr>
<tr>
<td>Object</td>
<td>654</td>
<td>.35</td>
<td>.31</td>
<td>.33</td>
</tr>
</tbody>
</table>

Initial Model

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>.64</td>
<td>.84</td>
<td>.72†</td>
</tr>
<tr>
<td>Object</td>
<td>.45</td>
<td>.52</td>
<td>.48*</td>
</tr>
</tbody>
</table>

Trained Model

### That/Wh-

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wh-</td>
<td>363</td>
<td>.63</td>
<td>.45</td>
<td>.52</td>
</tr>
<tr>
<td>That</td>
<td>68</td>
<td>.43</td>
<td>.48</td>
<td>.45</td>
</tr>
</tbody>
</table>

Initial Model

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wh-</td>
<td>.73</td>
<td>.75</td>
<td>.74*</td>
</tr>
<tr>
<td>That</td>
<td>.44</td>
<td>.57</td>
<td>.50†</td>
</tr>
</tbody>
</table>

Trained Model

* (p < .01) † (p < .05)
Conclusion

It is possible to acquire filler-gap without (complex) syntax. The current model offers additional benefits:

- Reflects developmental S-O asymmetry
- Reflects developmental That-Wh asymmetry
- Robust to varied initializations
  - positions: -3,3 ; -1,1 ; -0.1,0.1
  - sd: filler preverbal prob must outweigh skip-penalty
Questions?

Thanks to everyone who gave feedback on this project: Lacqueys, Clippers, Dave Howcroft, Evan Jaffe, William Schuler, and Peter Culicover, but especially Micha Elsner.
How does this model compare to Connor et al ‘10?

Connor et al are interested in modeling SRL acquisition and in replicating 1-1 role bias error (21 months).

**Plausibility**

- Connor et al ‘10 productively learn 5 roles
  - This increases their specificity
  - Children do not generalize above 2 roles until after 31 months (earliest) [Goldberg et al., 2004, Bello, 2012]

- Connor et al’s results raise questions about structure mapping
  Single N is patient 40% of the time?

**1-1 Role Bias**

- Connor et al (gold training): 63-82% 1-1 bias error
- Our initial model: 77% 1-1 bias error
Model: Relativizers

Initial model with function Gaussians
Model: Relativizers

Initial relative model with priors
Trained model with function Gaussians


The acquisition of finite complement clauses in english: A corpus-based analysis.
*Cognitive Linguistics*, 12:1–45.

Morphosyntactic cues impact filler-gap dependency resolution in 20- and 30-month-olds.
In *Poster session of BUCLD35*.

*Language Learning and Development*.

Predicted errors in childrens early sentence comprehension.
*Cognition*, 124:85–94.


