Bootstrapping into Filler-Gap: An Acquisition Story

Marten van Schijndel and Micha Elsner
Department of Linguistics
The Ohio State University

June 24, 2014
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

• Simple model of filler-gap
**Questions**

Wh-S: \{What\} ___ ate the apple?
Wh-O: \{What\} did the monkey eat ___?

**Relatives**

Wh-rS: Find \{the boy\} who ___ bumped the girl.
Wh-rO: Find \{the boy\} who the girl bumped ___.
That-rS: Find \{the boy\} that ___ bumped the girl.
That-rO: Find \{the boy\} that the girl bumped ___.
**Acquisition Pattern**

Developmental timeline of wh- question comprehension
Parentheses = marginal comprehension
That-relatives acquired slower than wh-relatives
[Seidl et al., 2003, Gagliardi et al., 2014]
Acquisition Pattern

1-1 Role Bias

Subject  Object

- John gorped
- Mary gorped John
- John and Mary gorped

Interpreted by Gertner and Fisher (2012) as ‘Agent-first bias’
But we will show: can be modeled as 1-1 role bias
Developmental timeline of 1-1 role bias errors (21, 25)
Children stop this error by 25 months
**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).

Problem:
Syntax isn’t great yet

- Role conjunction not comprehended
  - [Gertner and Fisher, 2012]
- Ditransitives not generalized until later
  - [Goldberg et al., 2004, Bello, 2012]
Model Motivation

What are children learning?

Different Possible Orderings

The flower hit the apple.
What hit the apple.
What did the flower hit?

Plausible:
Word ordering patterns are fairly widespread (e.g. SOV, SVO, etc)

Previously used in BabySRL [Connor et al., 2008, 2009, 2010]
Model

- Inspired by Gradual Learning Algorithm [Boersma, 1997]
- Structure mapping: nouns used to learn verbs [Yuan et al., 2012]
- Roles assigned via ordered, latent distributions

Assumptions

- (14m) Children can chunk nouns [Waxman and Booth, 2001]
- (pre-25m) Ns and roles are 1-to-1 [Gertner and Fisher, 2012]
- (9m) Abstract factors (#N) are used by learners [Xu, 2002]
- (4-5y) Children are bad at recursion [Diessel and Tomasello, 2001]

Implementation Assumptions

- Generate position of arguments relative to verb
- Sampled from Gaussian distributions
- Samples assumed to be independent
Model
The cat bumped the dog.
Model

Possible parses...

\[ P(SVO) = P(-1 \mid S) \cdot P(1 \mid O) \]
\text{The cat bumped the dog.}

\[ P( OVS ) = P(-1 \mid O) \cdot P(1 \mid S) \]
\text{The cat bumped the dog.}

\[ P(VO) = P(-1 \mid \text{skip}) \cdot P(1 \mid O) \]
\text{The cat bumped the dog.}

\[ P(SV) = P(-1 \mid S) \cdot P(1 \mid \text{skip}) \]
\text{The cat bumped the dog.}

\ldots
The cat bumped the dog.
Model

\[ P(-1 \mid S) \cdot P(1 \mid O) \]

Wh-S: Which cat bumped the dog?
Wh-O: Which cat did the dog bump?*
Model

Initialization 2.0

- Split distributions into mixtures of distributions
  - 1) strong due to canonical evidence
  - 2) weak, but finds arguments from anywhere
\[ P(-1 \mid S_C) \cdot P(1 \mid O_C) \]

**Wh-S:** Which cat bumped the dog?
Model

\[ P(-3 \mid O_N) \cdot P(-1 \mid S_C) \]

Wh-O: Which cat did the dog bump?
With priors, our initial model looks like this.
Evaluation

1. Extract CDS from Eve corpus
   
   (‘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 (EM)
   
   • Estimate labels using distributions over previous observations
   • Estimate new distributions using labelled data
Results
## 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 = 4820)

<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 = 4461)

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

But those numbers reflect overall performance. . .

We can try a coarse filler-gap filter.

**Extract sentences where either:**

- 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>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>.66</td>
<td>.83</td>
<td>.74</td>
</tr>
<tr>
<td>Trained</td>
<td>.64</td>
<td>.84</td>
<td>.72†</td>
</tr>
</tbody>
</table>

Subject (n = 691)

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>.35</td>
<td>.31</td>
<td>.33</td>
</tr>
<tr>
<td>Trained</td>
<td>.45</td>
<td>.52</td>
<td>.48*</td>
</tr>
</tbody>
</table>

Object (n = 654)

### That/Wh-

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>.63</td>
<td>.45</td>
<td>.53</td>
</tr>
<tr>
<td>Trained</td>
<td>.73</td>
<td>.75</td>
<td>.74*</td>
</tr>
</tbody>
</table>

Wh- (n = 689)

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>.43</td>
<td>.48</td>
<td>.45</td>
</tr>
<tr>
<td>Trained</td>
<td>.44</td>
<td>.57</td>
<td>.50†</td>
</tr>
</tbody>
</table>

That (n = 125)

* (p < .01) † (p < .05)
1-1 Role Bias

How often is NNV labelled as SOV? (1-1 role bias error)

- Our initial model: 66% error (1-1 bias)

Current model is comparable to Baby SRL
**Initialization Analysis**

**Very Robust**
- positions: -3,3 ; -1,1 ; -0.1,0.1
- variance: 0.5 – 4
- caveat: filler preverbal prob must outweigh skip-penalty
Do we really want this setup?
Is the non-canonical subject useful? (According to BIC)
“Helps” capture imperatives…
But kids know imperatives…

‘Put the cookie on the table!’
‘[You] put the cookie on the table!’
Then non-canonical subject isn’t useful (according to BIC)

Suggests dynamic Gaussian generation is possible
Future Work

- Add lexicalization
- Dynamically generate Gaussians
- Model non-English (verb-medial) languages
- Bootstrap linear grammar into hierarchic grammar
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
Questions?

Thanks to:

- Peter Culicover
- William Schuler
- Laura Wagner
- Attendees of the OSU 2013 Fall Ling. Colloquium Fest

This work was partially funded by an OSU Dept. of Linguistics Targeted Investment for Excellence (TIE) grant for collaborative interdisciplinary projects conducted during the academic year 2012-13.
Results: 1-1 bias

How often NNV is labelled SOV

Current Model

<table>
<thead>
<tr>
<th></th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>.66</td>
</tr>
<tr>
<td>Trained</td>
<td>.13</td>
</tr>
</tbody>
</table>

(n = 1000)

Trained Baby SRL

<table>
<thead>
<tr>
<th></th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg-Arg</td>
<td>.65</td>
</tr>
<tr>
<td>Arg-Verb</td>
<td>0</td>
</tr>
</tbody>
</table>

[Connor et al., 2008]

<table>
<thead>
<tr>
<th></th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg-Arg</td>
<td>.82</td>
</tr>
<tr>
<td>Arg-Verb</td>
<td>.63</td>
</tr>
</tbody>
</table>

[Connor et al., 2009]
### Results: 1-1 bias

#### 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>Transitive (n = 1000)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Initial</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Trained</td>
<td>.96</td>
</tr>
</tbody>
</table>

#### [Connor et al., 2010]

<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>Weak (10) lexical</td>
<td>.59</td>
</tr>
<tr>
<td>Strong (365) lexical</td>
<td>.74</td>
<td>Strong (365) lexical</td>
<td>.41</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>
### Role Bias Summary

How often is the agent correctly labelled?

<table>
<thead>
<tr>
<th>Transitives (1173 sents)</th>
<th>Intransitives (1513 sents)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Connor et al. (2010): 71-77%</td>
<td>• Connor et al. (2010): 41-59%</td>
</tr>
<tr>
<td>• Lexicalization helps</td>
<td></td>
</tr>
<tr>
<td>• Initial current model: 67%</td>
<td>• Initial current model: 100%</td>
</tr>
<tr>
<td>Trained current model: 65%</td>
<td>Trained current model: 96%</td>
</tr>
<tr>
<td>• Completely unlexicalized</td>
<td></td>
</tr>
</tbody>
</table>

Current model is comparable to Baby SRL for transitives
Current model does much better on intransitives
The boy/girl is gorping.
The girl is gorping the boy.
The girl and the boy are gorping.


