A Usage-Based Model of Early Grammatical Development

Barend Beekhuizen\textsuperscript{1,2} Rens Bod\textsuperscript{2} Afsaneh Fazly\textsuperscript{3} Suzanne Stevenson\textsuperscript{3} Arie Verhagen\textsuperscript{1}

\textsuperscript{1}Leiden University Centre for Linguistics, Leiden University

\textsuperscript{2}Institute for Logic, Language, and Computation, University of Amsterdam

\textsuperscript{3}Department of Computer Science, University of Toronto

26 June 2014
Workshop on Cognitive Modeling and Computational Linguistics
Motivation

- Explaining early production in first language acquisition
- Interaction of learning mechanisms

Model

- Semantic parser and generator
- Incrementally acquiring constructions from parses

Experiments

- Parsing experiment
- Generation experiment
Motivation

- Explaining early production in first language acquisition
- Interaction of learning mechanisms

Model

- Semantic parser and generator
- Incrementally acquiring constructions from parses

Experiments

- Parsing experiment
- Generation experiment
Motivation

- Explaining early production in first language acquisition
- Interaction of learning mechanisms

Model

- Semantic parser and generator
- Incrementally acquiring constructions from parses

Experiments

- Parsing experiment
- Generation experiment
The two-word phase

Daddy get!

Questions

- What are the representations behind these truncated utterances?
- How do these representations develop?

Earlier proposals

- Braine (1976), Schlesinger (1971)
- Usage-based proposals: Theakston et al. (2012)
  - Main focus is on abstraction (paradigmatic) rather than increasing syntagmatic knowledge
Introduction

Early production

The two-word phase

*Daddy* get!

Questions

- What are the representations behind these truncated utterances?
- How do these representations develop?

Earlier proposals

- Braine (1976), Schlesinger (1971)
- Usage-based proposals: Theakston et al. (2012)
  - Main focus is on abstraction (paradigmatic) rather than increasing syntagmatic knowledge
The two-word phase

*Daddy get!*

Questions

▶ What are the representations behind these truncated utterances?
▶ How do these representations develop?

Earlier proposals

▶ Braine (1976), Schlesinger (1971)
▶ Usage-based proposals: Theakston et al. (2012)
  ▶ Main focus is on abstraction (paradigmatic) rather than increasing syntagmatic knowledge
Goal 1

- Work out a model that explains syntagmatic development in early production from a usage-based vantage point
- Assuming that the length or arity of representations increases

Current models

<table>
<thead>
<tr>
<th>model</th>
<th>syntagms</th>
<th>grammatical</th>
<th>lexical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chang (2008)</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freudenthal et al. (2009)</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alishahi &amp; Stevenson (2010)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2012)</td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Goal 1

- Work out a model that explains syntagmatic development in early production from a usage-based vantage point
- Assuming that the length or arity of representations increases

Current models

<table>
<thead>
<tr>
<th>model</th>
<th>acquires:</th>
<th>syntagms</th>
<th>grammatical</th>
<th>lexical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chang (2008)</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freudenthal et al. (2009)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alishahi &amp; Stevenson (2010)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2012)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
A Usage-Based Model of Early Grammatical Development

Introduction

Interacting learning mechanisms

▶ All constructions acquired with **same mechanisms**
  (construction grammar: Tomasello 2003, Goldberg 2006)
▶ ‘Linking’ lexical acquisition, schematization, growth of rules
  (Beekhuizen, Bod & Verhagen 2014)

Interacting learning mechanisms

▶ So: responsible mechanisms interact

Current models

<table>
<thead>
<tr>
<th>model</th>
<th>acquires:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>syntagms</td>
</tr>
<tr>
<td>Chang (2008)</td>
<td>✓</td>
</tr>
<tr>
<td>Freudenthal et al. (2009)</td>
<td></td>
</tr>
<tr>
<td>Alishahi &amp; Stevenson (2010)</td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2012)</td>
<td></td>
</tr>
</tbody>
</table>
**Introduction**

- All constructions acquired with same mechanisms (construction grammar: Tomasello 2003, Goldberg 2006)
- ‘Linking’ lexical acquisition, schematization, growth of rules (Beekhuizen, Bod & Verhagen 2014)

**Interacting learning mechanisms**

- So: responsible mechanisms **interact**

**Current models**

<table>
<thead>
<tr>
<th>model</th>
<th>syntagms</th>
<th>grammatical</th>
<th>lexical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chang (2008)</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freudenthal et al. (2009)</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alishahi &amp; Stevenson (2010)</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2012)</td>
<td>✓ ✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A Usage-Based Model of Early Grammatical Development

Introduction

Interacting learning mechanisms

- All constructions acquired with same mechanisms (construction grammar: Tomasello 2003, Goldberg 2006)
- ‘Linking’ lexical acquisition, schematization, growth of rules (Beekhuizen, Bod & Verhagen 2014)

Interacting learning mechanisms

- So: responsible mechanisms interact

Current models

<table>
<thead>
<tr>
<th>model</th>
<th>acquires:</th>
<th>syntagms</th>
<th>grammatical</th>
<th>lexical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chang (2008)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freudenthal et al. (2009)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alishahi &amp; Stevenson (2010)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2012)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
Goal 2

Account of language acquisition in which

- both lexical and grammatical constructions are acquired with the same set of mechanisms

**Current models**

<table>
<thead>
<tr>
<th>model</th>
<th>syntagms</th>
<th>grammatical</th>
<th>lexical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chang (2008)</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Freudenthal et al. (2009)</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alishahi &amp; Stevenson (2010)</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2012)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>this model</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Goal 2
Account of language acquisition in which
- both lexical and grammatical constructions are acquired with the same set of mechanisms

Current models

<table>
<thead>
<tr>
<th>model</th>
<th>acquires:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>syntagms</td>
</tr>
<tr>
<td>Chang (2008)</td>
<td>✓</td>
</tr>
<tr>
<td>Freudenthal et al. (2009)</td>
<td>✓</td>
</tr>
<tr>
<td>Alishahi &amp; Stevenson (2010)</td>
<td>✓</td>
</tr>
<tr>
<td>Kwiatkowski et al. (2012)</td>
<td>✓</td>
</tr>
<tr>
<td>this model</td>
<td>✓</td>
</tr>
</tbody>
</table>
Input data

- Learner receives utterances paired with sets of situations
- Interpretability requirement (O'Grady 1997)

Definitions

- Utterance $U$ is a string of words $w_1, w_2, \ldots, w_n$
- $S$ is a set of situations $s$
- Propositional uncertainty: $|S| > 1$ (Siskind 1996)
- Propositional noise: $s_{correct} \not\in S$ (Siskind 1996)
Input data

- Learner receives utterances paired with sets of situations
- Interpretability requirement (O'Grady 1997)

Definitions

- Utterance $U$ is a string of words $w_1, w_2, \ldots, w_n$
- $S$ is a set of situations $s$
  - Propositional uncertainty: $|S| > 1$ (Siskind 1996)
  - Propositional noise: $s_{correct} \notin S$ (Siskind 1996)
A Usage-Based Model of Early Grammatical Development

Model
Input data

- Input data
  - Learner receives utterances paired with sets of situations
  - Interpretability requirement (O'Grady 1997)

Definitions

- Utterance $U$ is a string of words $w_1, w_2, \ldots, w_n$
- $S$ is a set of situations $s$
- Propositional uncertainty: $|S| > 1$ (Siskind 1996)
- Propositional noise: $s_{correct} \notin S$ (Siskind 1996)
Representations learned from input:

- Constructions, cf. construction grammar (Goldberg 1995)

**Definition**

- Pairings of
  - a meaning (tree)
  - a string of constituents, each containing
    - a phonological form
    - a semantic pointer
Parsing

- Model tries to find which parts of $U$ map to which parts of an $s \in S$.
- By creating derivations of constructions given $U, S$, using four interpretation mechanisms:
  - **Combine**: fill a phonologically open constituent of one construction with another construction
  - **Concatenate**: create a list of derivations
  - **Bootstrap**: fill a phonologically open constituent with an unknown word
  - **Ignore**: don’t integrate the word in the derivation
- Constraints on derivations:
  - All constructions in a derivation should map to the same $s \in S$.
  - Each construction in a derivation maps to a different node of the meaning (isomorphy)
Model

Parsing/Generating

Parsing

- Model tries to find which parts of $U$ map to which parts of an $s \in S$.
- By creating derivations of constructions given $U, S$, using four interpretation mechanisms:
  - **Combine**: fill a phonologically open constituent of one construction with another construction
  - **Concatenate**: create a list of derivations
  - **Bootstrap**: fill a phonologically open constituent with an unknown word
  - **Ignore**: don’t integrate the word in the derivation

- Constraints on derivations:
  - All constructions in a derivation should map to the same $s \in S$.
  - Each construction in a derivation maps to a different node of the meaning (isomorphy)
Parsing

- Model tries to find which parts of $U$ map to which parts of an $s \in S$.
- By creating derivations of constructions given $U, S$, using four interpretation mechanisms:
  - **Combine**: fill a phonologically open constituent of one construction with another construction
  - **Concatenate**: create a list of derivations
  - **Bootstrap**: fill a phonologically open constituent with an unknown word
  - **Ignore**: don’t integrate the word in the derivation
- Constraints on derivations:
  - All constructions in a derivation should map to the same $s \in S$.
  - Each construction in a derivation maps to a different node of the meaning (isomorphy)
A Usage-Based Model of Early Grammatical Development

Figure: The COMBINATION mechanisms
Figure: The IGNORE, BOOTSTRAP, and CONCATENATE mechanisms
Best analysis of $U$

- Multiple derivations may map to the same parts of a situation
- Then: take them together as a parse $t$
- Best analysis of $U$ is the most probable parse:

\[ P(t) = \sum_{d \in P} P(d) \]  \hspace{1cm} (1)

- Let a derivation $d = \langle c_1, c_2, \ldots, c_n \rangle$

\[ P(d) = \prod_{i=1}^{n} P(c_i) \]  \hspace{1cm} (2)

\[ P(c) = \frac{c.\text{count} + 1}{\sum_{c' \in C} c'.\text{count} + |C| + 1} \]  \hspace{1cm} (3)

- Where \texttt{BOOTSTRAP} and \texttt{CONCATENATE} count as one unseen $c$, and \texttt{IGNORE} as two
Best analysis of $U$

- Multiple derivations may map to the same parts of a situation
- Then: take them together as a parse $t$
- Best analysis of $U$ is the most probable parse:

$$P(t) = \sum_{d \in p} P(d)$$  \hspace{1cm} (1)

- Let a derivation $d = \langle c_1, c_2, \ldots, c_n \rangle$

$$P(d) = \prod_{i=1}^{n} P(c_i)$$  \hspace{1cm} (2)

$$P(c) = \frac{c.count + 1}{\sum_{c' \in C} c'.count + |C| + 1}$$  \hspace{1cm} (3)

- Where `BOOTSTRAP` and `CONCATENATE` count as one unseen $c$, and `IGNORE` as two
Best analysis of $U$

- Multiple derivations may map to the same parts of a situation.
- Then: take them together as a parse $t$.

Best analysis of $U$ is the most probable parse:

1. \[ P(t) = \sum_{d \in p} P(d) \]  
2. \[ P(d) = \prod_{i=1}^{n} P(c_i) \]  
3. \[ P(c) = \frac{c\text{.count} + 1}{\sum_{c' \in C} c'\text{.count} + |C| + 1} \]

Where BOOTSTRAP and CONCATENATE count as one unseen $c$, and IGNORE as two.
Best analysis of $U$

- Multiple derivations may map to the same parts of a situation
- Then: take them together as a parse $t$

Best analysis of $U$ is the most probable parse:

\[ P(t) = \sum_{d \in p} P(d) \]  \hspace{1cm} (1)

- Let a derivation $d = \langle c_1, c_2, \ldots, c_n \rangle$

\[ P(d) = \prod_{i=1}^{n} P(c_i) \]  \hspace{1cm} (2)

\[ P(c) = \frac{c\cdot\text{count} + 1}{\sum_{c' \in C} c'\cdot\text{count} + |C| + 1} \]  \hspace{1cm} (3)

- Where $\text{BOOTSTRAP}$ and $\text{CONCATENATE}$ count as one unseen $c$, and $\text{IGNORE}$ as two
Learning on the basis of best analysis

Idea of learning-as-processing (Langacker 2009)

Four learning mechanisms

- **ASSOCIATE** parts of \( U \) and parts of an \( s \) matching over recent \( U, s \) pairs (cross-situational learning)
- **UPDATE** count of used rules
- **SYNTAGMATIZATION**: store concatenation as a new construction
- **PARADIGMATIZATION**: store (more abstract) overlap between similar constructions as a new construction
Learning on the basis of best analysis

Idea of learning-as-processing (Langacker 2009)

Four learning mechanisms

- **ASSOCIATE** parts of $U$ and parts of an $s$ matching over recent $U, s$ pairs (cross-situational learning)
- **UPDATE** count of used rules
- **SYNTAGMATIZATION**: store concatenation as a new construction
- **PARADIGMATIZATION**: store (more abstract) overlap between similar constructions as a new construction
Learning on the basis of best analysis

Idea of learning-as-processing (Langacker 2009)

Four learning mechanisms

- **ASSOCIATE** parts of $U$ and parts of an $s$ matching over recent $U,s$ pairs (cross-situational learning)
- **UPDATE** count of used rules
- **SYNTAGMATIZATION**: store concatenation as a new construction
- **PARADIGMATIZATION**: store (more abstract) overlap between similar constructions as a new construction
A parse over the utterance you take ball.

A novel, syntagmatized construction

**Figure:** Syntagmatization
A phonologically empty constituent, generalizing over chair and table.

The set intersection of \{location, entity, chair\} and \{location, entity, table\}.

**Figure:** Paradigmatization
Training

- Model **incrementally** presented with $U, S$ pairs
- On the basis of Alishahi & Stevenson’s (2010) generation procedure
- $|S| = 2$ (propositional uncertainty is 1)
- Non-correct $s \in S$ randomly generated
- 5 simulations of 2000 input items.
Comprehension experiment

- After every input item measure comprehension.
- Averaging over 50 input items in each of 5 simulations:

<table>
<thead>
<tr>
<th>Identification</th>
<th>What proportion of $s_{\text{correct}}$ is identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Situation coverage</td>
<td>What proportion of the identified $s$ the best parse maps to</td>
</tr>
<tr>
<td>Utterance coverage</td>
<td>What proportion of $U$ is not IGNORED in the best parse</td>
</tr>
</tbody>
</table>
Generation experiment

- After every 50 input items
- Present the model with 50 situations it has never seen
- Generation is parsing but only using COMBINATION
- Best parse is most probable, most expressive one
- The generated $U$ is the yield of best parse given $s$.
- The actual $U \rightarrow$ the generation model.
- Average over 50 situations in each of 5 simulations:

<table>
<thead>
<tr>
<th>length</th>
<th>Length of $U$ given best parse</th>
</tr>
</thead>
<tbody>
<tr>
<td>situation coverage</td>
<td>What proportion of $s$ is expressed by the best parse</td>
</tr>
<tr>
<td>utterance precision</td>
<td>What proportion of the generated $U$ corresponds to actual $U$ for $s$</td>
</tr>
<tr>
<td>utterance recall</td>
<td>What proportion of the actual $U$ corresponds to the generated $U$</td>
</tr>
</tbody>
</table>
A Usage-Based Model of Early Grammatical Development

Experiment

Comprehension experiment

Figure: Comprehension scores over time.
Figure: Mean length of $U$ generated over time.
A Usage-Based Model of Early Grammatical Development

Generation experiment

Figure: Generation scores over time.
Comprehension

- Over time, more of $U$ and $s$ is understood
- Over time, $s_{\text{correct}}$ is identified more frequently
- Model can deal with some uncertainty

Generation

- Length increases
- Utterance recall gradually goes up (omission)
- Utterance precision is high from the start (commission)
- Qualitative analysis: in paper
Comprehension

- Over time, more of $U$ and $s$ is understood
- Over time, $s_{\text{correct}}$ is identified more frequently
- Model can deal with some uncertainty

Generation

- Length increases
- Utterance recall gradually goes up (omission)
- Utterance precision is high from the start (comission)
- Qualitative analysis: in paper
Concluding:

- **goal 1** Increasing length of utterances in production: ✓
- **goal 2** Interacting mechanisms (lexical & grammatical): ✓

Future work

- **Conceptually:**
  - Model is (admittedly) complex
  - Can we simplify the model?

- **Empirically:**
  - Test battery of studies on early transitive constructions, both comprehension and production
  - Can we simulate diverse experiments and observations?
Concluding:

- **goal 1** Increasing length of utterances in production: ✓
- **goal 2** Interacting mechanisms (lexical & grammatical): ✓

Future work

- **Conceptually:**
  - Model is (admittedly) complex
  - Can we simplify the model?

- **Empirically:**
  - Test battery of studies on early transitive constructions, both comprehension and production
  - Can we simulate diverse experiments and observations?
Concluding:

- **goal 1** Increasing length of utterances in production: ✓
- **goal 2** Interacting mechanisms (lexical & grammatical): ✓

Future work

- **Conceptually:**
  - Model is (admittedly) complex
  - Can we simplify the model?

- **Empirically:**
  - Test battery of studies on early transitive constructions, both comprehension and production
  - Can we simulate diverse experiments and observations?
Thanks to:

- The anonymous reviewers for valuable and thoughtful comments,
- The organization and participants of CMCL,
- Afra Alishahi for providing us with the code of the generation procedure,
- NWO (Netherlands) for funding Barend Beekhuizen,
- NSERC (Canada) for funding Afsaneh Fazly and Suzanne Stevenson.