A Usage-Based Model of Early Grammatical Development

Barend Beekhuizen\textsuperscript{1,2}

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

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

12 December 2014

Coglingdays 6, Ghent
Motivation

- How does a constructicon emerge in language acquisition?

Theoretical issues

- Metaphors of abstraction: hypotheses vs. processing
- Starting big and starting small: the horizontal dimension
- The linking problem: ‘all together now’

Model and experiments

- The Syntagmatic-Paradigmatic Learner
- Comprehension experiment
- Generation experiment
Motivation

- How does a constructicon emerge in language acquisition?

Theoretical issues

- Metaphors of abstraction: hypotheses vs. processing
- Starting big and starting small: the horizontal dimension
- The linking problem: ‘all together now’

Model and experiments

- The Syntagmatic-Paradigmatic Learner
- Comprehension experiment
- Generation experiment
Motivation

- How does a constructicon emerge in language acquisition?

Theoretical issues

- Metaphors of abstraction: hypotheses vs. processing
- Starting big and starting small: the horizontal dimension
- The linking problem: ‘all together now’

Model and experiments

- The Syntagmatic-Paradigmatic Learner
- Comprehension experiment
- Generation experiment
Understanding the acquisition of grammar

- Adult state (cf. Goldberg 1995): Constructions of various degrees of abstraction
- How to get there? (Tomasello 2003, Goldberg 2006)
- Clear sketch of processes: Langacker (2009):
  - Learning is a by-product of processing
  - ‘Units’ (constructions) emerge through selective reinforcement
  - Units are of the same make as the input items (conceptual and phonological structure)
  - Abstraction is not ‘creating something new’, but rather the potential that is immanent in a number of more concrete experiences
Three theoretical issues
A Usage-Based Model of Early Grammatical Development

Three theoretical issues

Abstraction

What *is* abstraction

- Usage-based work blends two frames:
  - abstraction as *hypotheses* about generalizability
  - abstraction as a *by-product* of processing
What *is* abstraction

- Usage-based work blends two frames:
  - abstraction as hypotheses about generalizability
  - abstraction as a by-product of processing

Relevance of this?

- Hypothesis frame brings along:
  - Search for new hypotheses
  - Evaluating them against some data
  - Corroborating or rejecting them
- Legacy: too post-hoc/offline for usage-based view
- Chang (2006) is based on the hypothesis view.

Proposal #1: Abstraction

- Emphasize the ‘by-product’ view in explanation of abstraction
Chunks and syntagms: The horizontal dimension

- What is the width/arity/‘length’ of early units
- Impression: infant has built up (e.g.) full caused-motion utterances, then abstracts i.o to form paradigms.
- **Starting big**: infants learn chunks, later break them down and do ‘blame assignment’ (Tomasello 2003)
- **Problems with starting big**
  - Storage of unstructured phonological wholes is problematic
  - Profile early SVO/SV/VO-productions (Theakston et al. 2012)
  - Argument omission in early production
Chunks and syntagms: The horizontal dimension

- What is the width/arity/‘length’ of early units
- Impression: infant has built up (e.g.) full caused-motion utterances, then abstracts i.o to form paradigms.
- Starting big: infants learn chunks, later break them down and do ‘blame assignment’ (Tomasello 2003)
- Problems with starting big
  - Storage of unstructured phonological wholes is problematic
  - Profile early SVO/SV/VO-productions (Theakston et al. 2012)
  - Argument omission in early production

Proposal #2: Syntagmatization

- ‘Longer’ constructions emerge through a gradual build-up of the horizontal dimension of constructions as a by-product of processing
The linking problem

▶ (Beekhuizen, Bod & Verhagen 2014): All processes have to work at the same time: developmental continuity
  ▶ Learning lexical and grammatical constructions
  ▶ Syntagmatization and paradigmmatization (abstraction)

▶ (Instrumental) assumption/idealization:
  Lexical cx > syntagmatize > paradigmmatize (Chang 2006, Alishahi & Stevenson 2010)
The linking problem

- (Beekhuizen, Bod & Verhagen 2014): All processes have to work at the same time: developmental continuity
  - Learning lexical and grammatical constructions
  - Syntagmatization and paradigmmatization (abstraction)

- (Instrumental) assumption/idealization:
  Lexical cx > syntagmatize > paradigmatize (Chang 2006, Alishahi & Stevenson 2010)

- In a sense true:
  - Syntagmatization requires analyses using lexical constructions
  - Paradigmatization requires syntagmatized constructions

- But not as consecutive stages.
A Usage-Based Model of Early Grammatical Development

Three theoretical issues

Linking problem

The linking problem

- (Beekhuizen, Bod & Verhagen 2014): All processes have to work at the same time: developmental continuity
  - Learning lexical and grammatical constructions
  - Syntagmatization and paradigmatization (abstraction)
- (Instrumental) assumption/idealization:
  Lexical cx > syntagmatize > paradigmatize (Chang 2006, Alishahi & Stevenson 2010)
- In a sense true:
  - syntagmatization requires analyses using lexical constructions
  - paradigmatization requires syntagmatized constructions
- But not as consecutive stages.

Proposal #3: Processing continuity

- Mechanisms of abstraction, syntagmatization, and different varieties of form-meaning association operate simultaneously.
The model
The cycle of the model

- Model receives utterance paired with set of situations
- Model comes up with the best analysis for this utterance
- Used constructions are reinforced

Key innovations

- All mechanisms of analysis and learning are available throughout time
- In analyzing, the model can concatenate multiple partial analyses
- These are starting point for novel syntagms
- Abstraction is a blind process
- (Stat.) pre-emption, semantic fit, prototype effects follow from same analysis mechanism.
Representations learned from input:

- Constructions, cf. construction grammar (Goldberg 1995)

Definition

- Pairings of
  - a meaning
  - a string of constituents, each containing
    - a phonological form (possibly empty)
    - a semantic pointer
Analyzing

- Model tries to find which parts of utterance map to parts of a situation
- By creating derivations of constructions, using four interpretation mechanisms:
  - **Combine**: fill a 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
- Often many possibilities: select most probable one (see Beekhuizen et al. 2014)
Analyzing

- Model tries to find which parts of utterance map to parts of a situation
- By creating **derivations** of constructions, using four interpretation mechanisms:
  - **COMBINE**: fill a 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
- Often many possibilities: select most probable one (see Beekhuizen et al. 2014)
Analyzing

- Model tries to find which parts of utterance map to parts of a situation
- By creating derivations of constructions, using four interpretation mechanisms:
  - **Combine**: fill a 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
- Often many possibilities: select most probable one (see Beekhuizen et al. 2014)
A Usage-Based Model of Early Grammatical Development

- Model
- Analyzing

**Figure:** The **COMBINATION** mechanisms
Figure: The IGNORE, BOOTSTRAP, and CONCATENATE mechanisms
Learning on the basis of best analysis

Idea of learning-as-processing (Langacker 2009)

Idea of learned units being ‘of the same matter’ as input items (ibid.)

Four learning mechanisms

- **ASSOCIATE** parts of utterance and parts of a situation matching over recent experiences (cross-situational learning)
- **REINFORCE** used rules
- **SYNTAGMATIZATION**: store concatenation as a new construction
- **PARADIGMATIZATION**: add (more abstract) overlap between similar constructions to constructicon
Learning on the basis of best analysis

Idea of learning-as-processing (Langacker 2009)

Idea of learned units being ‘of the same matter’ as input items (ibid.)

Four learning mechanisms

- **ASSOCIATE** parts of utterance and parts of a situation matching over recent experiences (cross-situational learning)
- **REINFORCE** used rules
- **SYNTAGMATIZATION**: store concatenation as a new construction
- **PARADIGMATIZATION**: add (more abstract) overlap between similar constructions to constructicon
Learning on the basis of best analysis

Idea of learning-as-processing (Langacker 2009)

Idea of learned units being ‘of the same matter’ as input items (ibid.)

Four learning mechanisms

- **ASSOCIATE** parts of utterance and parts of a situation matching over recent experiences (cross-situational learning)
- **REINFORCE** used rules
- **SYNTAGMATIZATION**: store concatenation as a new construction
- **PARADIGMATIZATION**: add (more abstract) overlap between similar constructions to constructicon
A Usage-Based Model of Early Grammatical Development

A parse over the utterance you take ball.

A novel, syntagmatized construction
A Usage-Based Model of Early Grammatical Development

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

A phonologically empty constituent, generalizing over chair and table

Figure: Paradigmatization
Experiments
Experiments

Experimental set-up

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.
A Usage-Based Model of Early Grammatical Development

Experiments

Comprehension experiment

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

Experiments

Generation experiment

Figure: Generation scores over time.
### Table: Generations

<table>
<thead>
<tr>
<th>time</th>
<th>generated production</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-450</td>
<td>[ [ she ] put ]</td>
</tr>
<tr>
<td>500</td>
<td>[ she [ put ] ]</td>
</tr>
<tr>
<td>550</td>
<td>[ [ she ] [ put ] [ in ] ]</td>
</tr>
<tr>
<td>600-900</td>
<td>[ [ she ] put them [ away ] ]</td>
</tr>
<tr>
<td>950</td>
<td>[ [ she ] put [ them ] ]</td>
</tr>
<tr>
<td>1000</td>
<td>[ [ she ] put them [ away ] ]</td>
</tr>
<tr>
<td>1050</td>
<td>[ [ she ] put [ them ] away ]</td>
</tr>
<tr>
<td>1400</td>
<td>[ [ she ] put them away ]</td>
</tr>
</tbody>
</table>
The growth of grammar

- A look under the hood
- How abstract are the constructions at a given point in time?
- How long are they?
- Case: constructions with *take* and their network
A Usage-Based Model of Early Grammatical Development

Experiments

The growth of grammar
The growth of grammar

AGT ACTION

AGT take

You take  I take  She take  We take out

T = 750
Experiments

The growth of grammar

T = 1000

X EVENT

AGT ACTION

AGT take

You take

I take

She take

We take out

AGT ACTION LOC

AGT take LOC

She take out

She take Sarah

AGT put in

...
A Usage-Based Model of Early Grammatical Development

Experiments

The growth of grammar

T = 1250

X EVENT

AGT ACTION

AGT take

You take
I take
She take
We take
out
She take
out
She take
Sarah
You take
back
I take
sharpener
She take
it

AGT ACTION LOC

AGT take LOC

She

AGT ACTION `X

AGT give it to REC

AGT put in

AGT put PAT on LOC

... ...

AGT ACT PAT REL X

AGT ACT PAT REL LOC

AGT take PAT

She
take
you
to
hospital
Experiments

The growth of grammar

A Usage-Based Model of Early Grammatical Development

\[ T = 1500 \]
Theoretical points

- Abstraction as a by-product of processing
- Starting-small: syntagmatization
- Linking everything up
Conclusion

Theoretical points

- Abstraction as a by-product of processing
- Starting-small: syntagmatization
- Linking everything up

Empirical validation

- Gradual convergence in perception and production
- Constructions become increasingly long and abstract
Theoretical points

- Abstraction as a by-product of processing
- Starting-small: syntagmatization
- Linking everything up

Empirical validation

- Gradual convergence in perception and production
- Constructions become increasingly long and abstract

Raising new questions

- How is abstraction constrained under the 'by-product' view?
- Relation to Bybee’s type frequency or Baayen’s hapaxes
- Analyzing is now a rational decision making process: can we get rid of this legacy as well?
Thanks to:

- Rens Bod, Afsaneh Fazly, Suzanne Stevenson, and Arie Verhagen for many fruitful discussions
- The anonymous reviewers for their thoughtful comments,
- Afra Alishahi for providing me with the code of the generation procedure,
- NWO (Netherlands) for funding
- The organization and participants of the Coglingdays,