A Usage-Based Model of Early Grammatical Development

Barend Beekhuizen[1,2] Rens Bod[2] Afsaneh Fazly[3]
Suzanne Stevenson[3] Arie Verhagen[1]

[1]Leiden University Centre for Linguistics, Leiden University

[2]Institute for Logic, Language, and Computation, University of Amsterdam

[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)
- Generative proposals: Pinker (1984), Boster (1997), Lebeaux (2000), Hyams & Wexler (1993)
- 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)
- Generative proposals: Pinker (1984), Boster (1997), Lebeaux (2000), Hyams & Wexler (1993)
- 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)
- Generative proposals: Pinker (1984), Boster (1997), Lebeaux (2000), Hyams & Wexler (1993)
- 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

		acquires:	
model	syntagms	grammatical	lexical
Chang (2008)		✓	
Freudenthal et al. (2009)	✓		
Alishahi & Stevenson (2010)		✓	
Kwiatkowski et al. (2012)		✓	✓

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

		acquires:	
model	syntagms	grammatical	lexical
Chang (2008)		✓	
Freudenthal et al. (2009)	✓		
Alishahi & Stevenson (2010)		✓	
Kwiatkowski et al. (2012)		✓	✓

- 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

		acquires:	
model	syntagms	grammatical	lexical
Chang (2008)		✓	
Freudenthal et al. (2009)	✓		
Alishahi & Stevenson (2010)		✓	
Kwiatkowski et al. (2012)		✓	✓

- 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

		acquires:	
model	syntagms	grammatical	lexical
Chang (2008)		✓	
Freudenthal et al. (2009)	✓		
Alishahi & Stevenson (2010)		✓	
Kwiatkowski et al. (2012)		✓	✓

- 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

		acquires:	
model	syntagms	grammatical	lexical
Chang (2008)		✓	
Freudenthal et al. (2009)	✓		
Alishahi & Stevenson (2010)		✓	
Kwiatkowski et al. (2012)		✓	✓

Goal 2

Account of language acquisition in which

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

Current models

		acquires:	
model	syntagms	grammatical	lexical
Chang (2008)		✓	
Freudenthal et al. (2009)	✓		
Alishahi & Stevenson (2010)		✓	
Kwiatkowski et al. (2012)		✓	✓
this model	✓	✓	✓

Goal 2

Account of language acquisition in which

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

Current models

		acquires:	
model	syntagms	grammatical	lexical
Chang (2008)		✓	
Freudenthal et al. (2009)	✓		
Alishahi & Stevenson (2010)		✓	
Kwiatkowski et al. (2012)		✓	✓
this model	✓	✓	✓

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

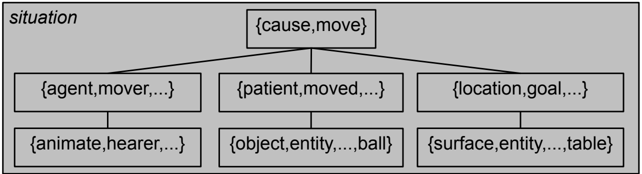
Definitions

- 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)

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

Definitions

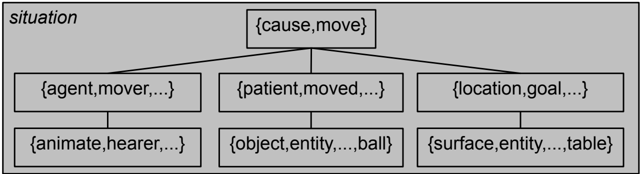
- Utterance U is a string of words $w_1, w_2, \ldots, w_n$
- S is a set of situations s
- Propositional uncertainty: $|S| > 1$ (Siskind 1996)
- Propositional noise: $s_{correct} \notin S$ (Siskind 1996)



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

Definitions

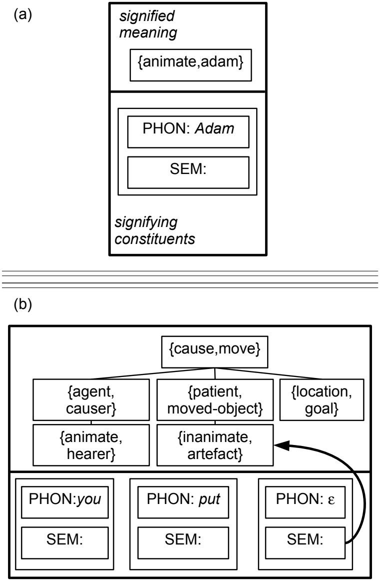
- Utterance U is a string of words $w_1, w_2, \ldots, w_n$
- S is a set of situations s
- Propositional uncertainty: $|S| > 1$ (Siskind 1996)
- Propositional noise: $s_{correct} \notin S$ (Siskind 1996)



- 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)

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)

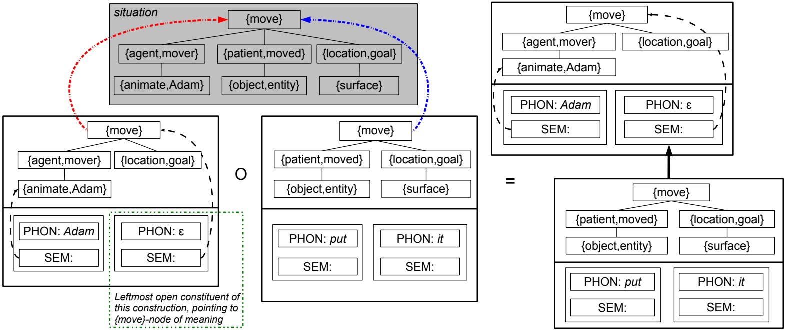


Figure: The COMBINATION mechanisms

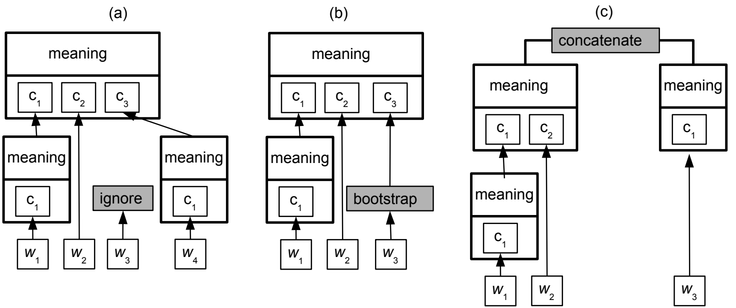


Figure: The IGNORE, BOOTSTRAP, and CONCATENATE mechanisms

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:
  - $$P(t) = \sum_{d \in p} P(d) \quad (1)$$
  - Let a derivation $d =< c_1, c_2, \ldots, c_n >$
  - $$P(d) = \prod_{i=1}^{n} P(c_i) \quad (2)$$
  - $$P(c) = \frac{c.count + 1}{\sum_{c' \in C} c'.count + |C| + 1} \quad (3)$$
  - 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:
  - $$P(t) = \sum_{d \in p} P(d) \quad (1)$$
  - Let a derivation $d = < c_1, c_2, \ldots, c_n >$
  - $$P(d) = \prod_{i=1}^{n} P(c_i) \quad (2)$$
  - $$P(c) = \frac{c.count + 1}{\sum_{c' \in C} c'.count + |C| + 1} \quad (3)$$
  - 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:
  - $$P(t) = \sum_{d \in p} P(d) \quad (1)$$
  - Let a derivation $d =< c_1, c_2, \ldots, c_n >$
  - $$P(d) = \prod_{i=1}^{n} P(c_i) \quad (2)$$
  - $$P(c) = \frac{c.count + 1}{\sum_{c' \in C} c'.count + |C| + 1} \quad (3)$$
  - 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:
  - $$P(t) = \sum_{d \in p} P(d) \qquad (1)$$
  - Let a derivation $d =< c_1, c_2, \ldots, c_n >$
  - $$P(d) = \prod_{i=1}^{n} P(c_i) \qquad (2)$$
  - $$P(c) = \frac{c.count + 1}{\sum_{c' \in C} c'.count + |C| + 1} \qquad (3)$$
  - Where BOOTSTRAP and CONCATENATE count as one unseen c, and 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

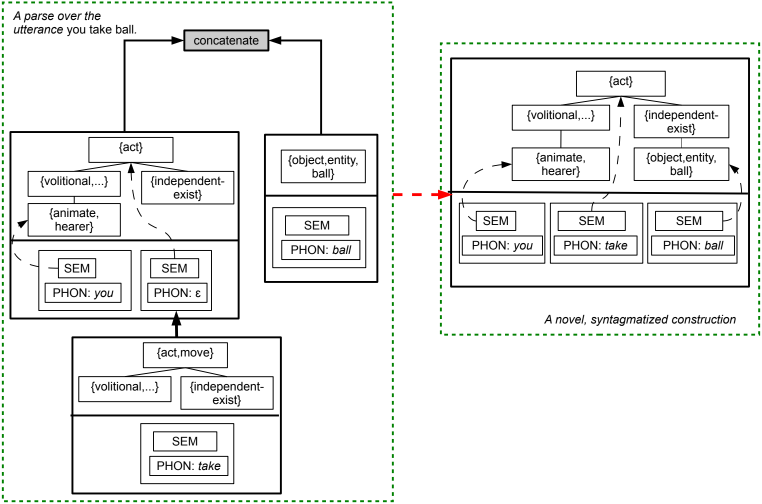


Figure: Syntagmatization

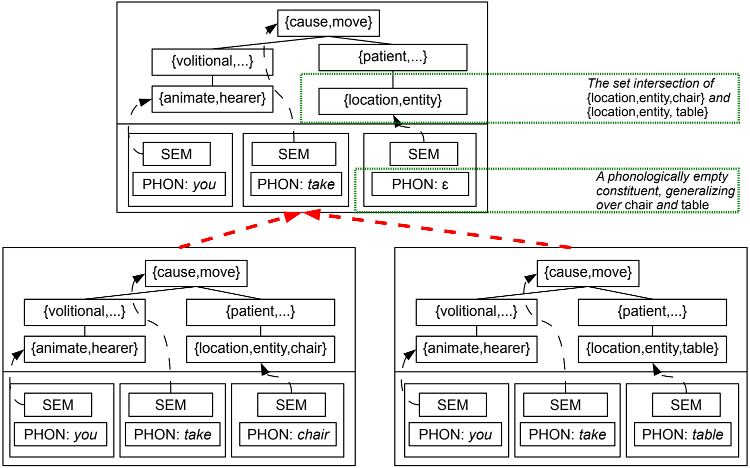


Figure: Paradigmatization

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:

identification	What proportion of s_{correct} is identified
situation coverage	What proportion of the identified s the best parse maps to
utterance coverage	What proportion of U is not IGNORED in the best parse

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:

length	Length of U given best parse
situation coverage	What proportion of s is expressed by the best parse
utterance precision	What proportion of the generated U corresponds to actual U for s
utterance recall	What proportion of the actual U corresponds to the generated U

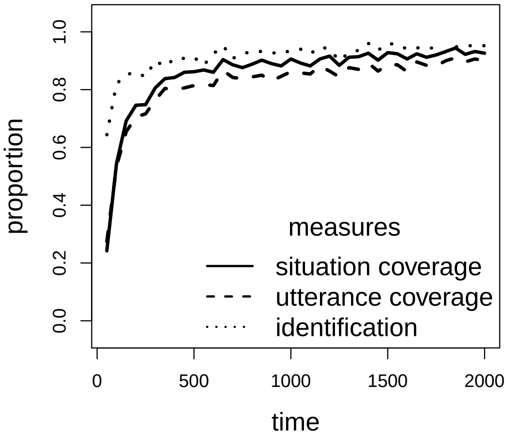


Figure: Comprehension scores over time.

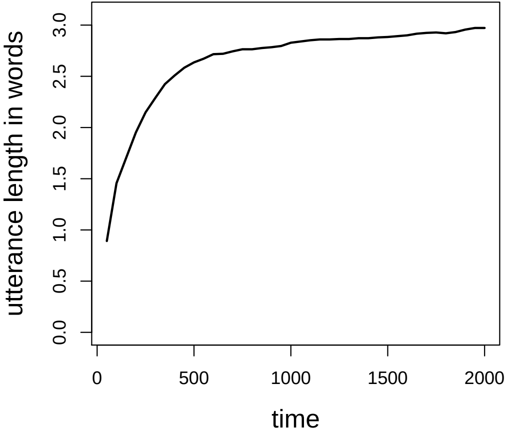


Figure: Mean length of *U* generated over time.

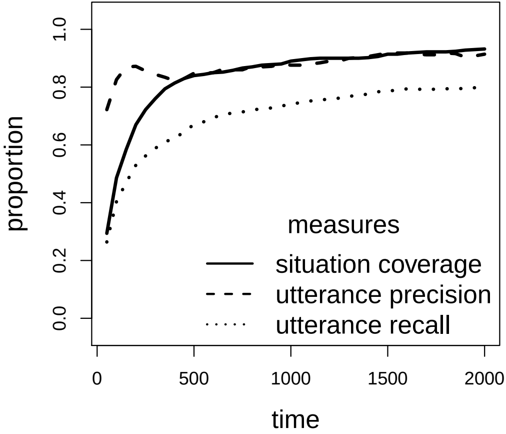


Figure: Generation scores over time.

Comprehension

- Over time, more of U and s is understood
- Over time, s_{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

Comprehension

- Over time, more of U and s is understood
- Over time, s_{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.