



UvA-DARE (Digital Academic Repository)

A Usage-Based Model of Early Grammatical Development

Beekhuizen, B.F.; Bod, L.W.M.; Fazly, A.; Stevenson, S.; Verhagen, A.

Published in:

ACL 2014: The Fifth Workshop on Cognitive Modeling and Computational Linguistics (CMCL)

[Link to publication](#)

Citation for published version (APA):

Beekhuizen, B., Bod, R., Fazly, A., Stevenson, S., & Verhagen, A. (2014). A Usage-Based Model of Early Grammatical Development. In V. Demberg, & T. O'Donnell (Eds.), *ACL 2014: The Fifth Workshop on Cognitive Modeling and Computational Linguistics (CMCL): proceedings of the workshop: June 26, 2014, Baltimore, Maryland, USA* (pp. 46-54). Stroudsburg, PA: Association for Computational Linguistics.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

A Usage-Based Model of Early Grammatical Development

Barend Beekhuizen

LUCL

Leiden University

b.f.beekhuizen@hum.leidenuniv.nl

Rens Bod

ILLC

University of Amsterdam

l.w.m.bod@uva.nl

Afsaneh Fazly and Suzanne Stevenson

Department of Computer Science

University of Toronto

afsaneh,suzanne@cs.toronto.edu

Arie Verhagen

LUCL

Leiden University

a.verhagen@hum.leidenuniv.nl

Abstract

The representations and processes yielding the limited length and telegraphic style of language production early on in acquisition have received little attention in acquisitional modeling. In this paper, we present a model, starting with minimal linguistic representations, that incrementally builds up an inventory of increasingly long and abstract grammatical representations (form+meaning pairings), in line with the usage-based conception of language acquisition. We explore its performance on a comprehension and a generation task, showing that, over time, the model better understands the processed utterances, generates longer utterances, and better expresses the situation these utterances intend to refer to.

1 Introduction

A striking aspect of language acquisition is the difference between children’s and adult’s utterances. Simulating early grammatical production requires a specification of the nature of the linguistic representations underlying the short, telegraphic utterances of children. In the usage-based view, young children’s grammatical representations are thought to be less abstract than adults’, e.g. by having stricter constraints on what can be combined with them (cf. Akhtar and Tomasello 1997; Bannard et al. 2009; Ambridge et al. 2012). The representations and processes yielding the restricted length of these early utterances, however, have received little attention. Following Braine (1976), we adopt the working hypothesis that the early learner’s grammatical representations are more limited in length (or: arity) than those of adults.

Similarly, in computational modeling of grammar acquisition, comprehension has received more

attention than language generation. In this paper we attempt to make the mechanisms underlying early production explicit within a model that can parse and generate utterances, and that incrementally learns constructions (Goldberg, 1995) on the basis of its previous parses. The model’s search through the hypothesis space of possible grammatical patterns is highly restricted. Starting from initially small and concrete representations, it learns incrementally long representations (**syntagmatic growth**) as well as more abstract ones (**paradigmatic growth**). Several models address either paradigmatic (Alishahi and Stevenson, 2008; Chang, 2008; Bannard et al., 2009) or syntagmatic (Freudenthal et al., 2010) growth. This model aims to explain both, thereby contributing to the understanding of how different learning mechanisms interact. As opposed to other models involving grammars with semantic representations (Alishahi and Stevenson, 2008; Chang, 2008), but similar to Kwiatkowski et al. (2012), the model starts without an inventory of mappings of single words to meanings.

Based on motivation from usage-based and construction grammar approaches, we define several learning principles that allow the model to build up an inventory of linguistic representations. The model incrementally processes pairs of an utterance U , consisting of a string of words $w_1 \dots w_n$, and a set of situations S , one of which is the situation the speaker intends to refer to. The other situations contribute to propositional uncertainty (the uncertainty over which proposition the speaker is trying to express; Siskind 1996). The model tries to identify the intended situation and to understand how parts of the utterance refer to certain parts of that situation. To do so, the model uses its growing inventory of linguistic representations (Section 2) to analyze U , producing a set of structured semantic analyses or parses (Fig. 1, arrow 1; Section 3).

The resulting best parse, U and the selected situation are then stored in a memory buffer (arrow 2), which is used to learn new constructions (arrow 3) using several learning mechanisms (Section 4). The learned constructions can then be used to generate utterances as well. We describe two experiments: in the comprehension experiment (Section 5), we evaluate the model’s ability to parse the stream of input items. In the generation experiment (Section 6), the model generates utterances on the basis of a given situation and its linguistic knowledge. We evaluate the generated utterances given different amounts of training items to consider the development of the model over time.

2 Representations

We represent linguistic knowledge as **constructions**: pairings of a signifying form and a signified (possibly incomplete) semantic representation (Goldberg, 1995). The **meaning** is represented as a graph with the nodes denoting entities, events, and their relations, connected by directed unlabeled edges. The conceptual content of each node is given by a set of semantic features. We assume that meaning representations are rooted trees. The signifying **form** consists of a positive number of **constituents**. Every constituent has two elements: a phonological form, and a pointer to a node in the signified meaning (in line with Verhagen 2009). Both can be specified, or one can be left empty. Constituents with unspecified phonological forms are called **open**, denoted with ϵ in the figures. The **head constituent** of a construction is defined as the constituent that has a pointer to the root node of the signified meaning. We furthermore require that no two constituents point to the same node of the signified meaning.

This definition generalizes over lexical elements (one phonologically specified constituent) as well as larger linguistic patterns. Fig. 2, for instance, shows two larger constructions being combined with each other. We call the set of constructions the learner has at some moment in time the **constructicon** C (cf. Goldberg 2003).

3 Parsing

3.1 Parsing operations

We first define a **derivation** d as an assembly of constructions in C , using four parsing operations defined below. In parsing, derivations are constrained by the utterance U and the situations

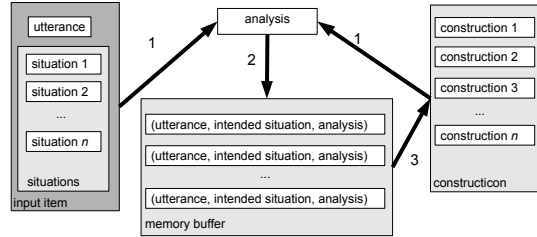


Figure 1: The global flow of the model

S , whereas in production, only a situation s constrains the derivation. The leaf nodes of a derivation must consist of phonological constraints of constructions that (in parsing) are satisfied by U . All constructions used in a derivation must map to the same situation $s \in S$. A construction c maps to s iff the meaning of c constitutes a subgraph of s , with the features on each of the nodes in the meaning of c being a subset of the features on the corresponding node of s . Moreover, each construction must map to a different part of s . This constitutes a mutual exclusivity effect in analyzing U : every part of the analysis must contribute to the composite meaning. A derivation d thus gives us a mapping between the composed meaning of all constructions used in d and one situation $s \in S$. The aggregate mapping specifies a subgraph of s that constitutes the interpretation of that derivation.

The central parsing operation is the COMBINATION operator \circ . In $c_i \circ c_j$, the leftmost open constituent of c_i is combined with c_j . Fig. 2 illustrates COMBINATION. COMBINATION succeeds if both the semantic pointer of the leftmost open constituent of c_i and the semantic pointer of the head constituent of c_j map to the same semantic node of a situation s

Initially, the model has few constructions to analyze the utterance with. Therefore, we define three other operations that allow the model to create a derivation over the full utterance without combining constructions. First, a known or unknown word that cannot be fit into a derivation, can be IGNORED. Second, an unknown word can be used to fill an open constituent slot of a construction with the BOOTSTRAP operator. Bootstrapping entails that the unknown word will be associated with the semantics of the node. Finally, the learner can CONCATENATE multiple derivations, by linearly sequencing them, thus creating a more complex derivation without combining con-

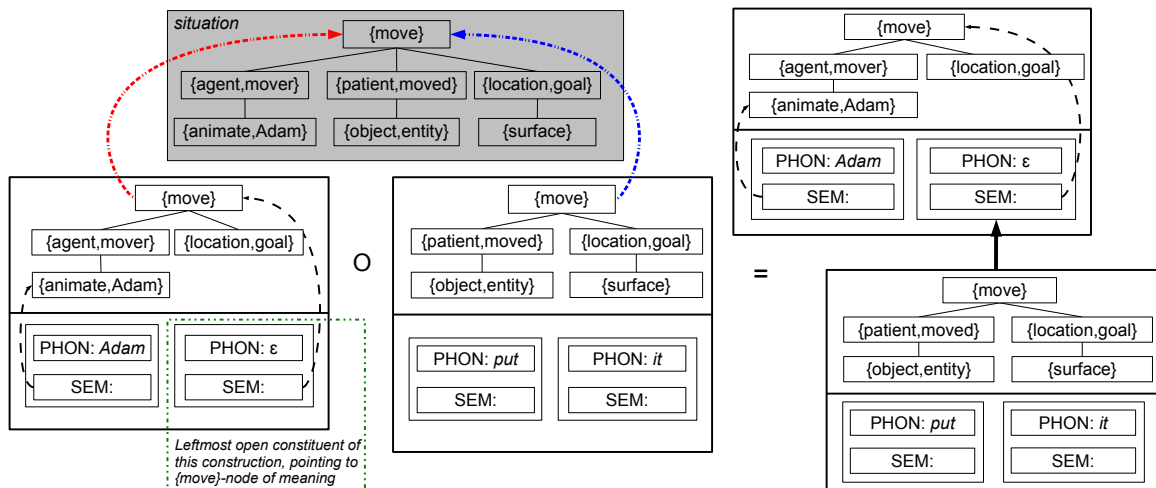


Figure 2: Combining constructions. The dashed lines represent semantic pointers, either from constituents to the constructional meaning (black) or from the constructions to the situation (red and blue).

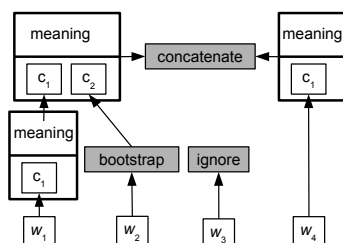


Figure 3: The CONCATENATE, IGNORE and BOOTSTRAP operators (internal details of the constructions left out).

structions. This allows the learner to interpret a larger part of the situation than with COMBINATION only. The resulting sequences may be analyzed in the learning process as constituting one larger construction, consisting of the parts of the concatenated derivations. Fig. 3 illustrates these three operations.

3.2 Selecting the best analysis

Multiple derivations can be highly similar in the way they map parts of U to parts of an $s \in S$. We define a **parse** to be a set of derivations that have the same internal structure and the same mappings to a situation, but that use different constructions in doing so (cf. multiple licensing; Kay 2002). We take the most probable parse of U to be the best analysis of U . The most probable parse points to a situation, which the model then assumes to be the **identified situation** or $s_{\text{identified}}$. If no parse can be

made, $s_{\text{identified}}$ is selected at random from S .

The probability of a parse p is given by the sum of the probabilities of the derivations d subsumed under that parse, which in turn are defined as the product of the probabilities of the constructions c used in d .

$$P(p) = \sum_{d \in p} P(d) \quad (1)$$

$$P(d) = \prod_{c \in d} P(c) \quad (2)$$

The probability of a construction $P(c)$ is given by its relative frequency (count) in the construction C , smoothed with Laplace smoothing. We assume that the simple parsing operations of IGNORE, BOOTSTRAP, and CONCATENATION reflect usages of an unseen construction with a count of 0.

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

The most probable parse, U and $s_{\text{identified}}$ are added to the memory buffer. The memory buffer has a pre-set maximal length, discarding the oldest exemplars upon reaching this length. In the future, we plan to consider more realistic mechanisms for the memory buffer, such as graceful degradation, and attention effects.

4 Learning mechanisms

The model uses the best parse of the utterance to develop its knowledge of the constructions in the construction C . Two simple operations, UPDATE and ASSOCIATION, are used to create initial constructions and reinforce existing ones respectively. Two additional operations, PARADIGMATIZATION and SYNTAGMATIZATION, are key to the model's ability to extend these initial representations by inducing novel constructions that are richer and more abstract than existing ones.

4.1 Direct learning from the best parse

The best parse is used to UPDATE C . For this mechanism, the model uses the concrete meaning of $s_{\text{identified}}$ rather than the (potentially more abstract) meaning of the constructions in the best parse.¹ Every construction in the parse is assigned the subgraph of $s_{\text{identified}}$ it maps to as its new meaning, and the count of the adjusted construction is incremented with 1, or added to C with a count of 1, if it does not yet exist. This includes applications of the BOOTSTRAP operation, creating a mapping of the previously unknown word to a situational meaning.

ASSOCIATE constitutes a form of simple cross-situational learning over the memory buffer. The intuition is that co-occurring word sequences and meaning components that remain unanalyzed across multiple parses might themselves comprise the form-meaning pairing of a construction. If the unanalyzed parts of two situations contain an overlapping subgraph, and the unanalyzed parts of two utterances an overlapping subsequence of words, the two are mapped to each other and added to C with a count of 0.

4.2 Qualitative extension of the best parse

Syntagmatization Some of the processes described thus far yield analyses of the input in which constructions are linearly associated but lack appropriate relational structure among them. The model requires a process, which we call SYNTAGMATIZATION, that enables it to induce further hierarchical structure.

In order for the learner to acquire constructions in which the different constituents point to different parts of the construction's meaning, the ASSO-

CIATE operation does not suffice. We assume that the learner is able to learn such constructions by using concatenated derivations. The process we propose is SYNTAGMATIZATION. In this process, the various concatenated derivations are taken as constituents of a novel construction. This instantiates the idea that joint processing of two (or more) events gradually leads to a joint representation of these, previously independent, events.

More precisely, the process starts by taking the top nodes T of the derivations in the best parse, where T consists of the single top node if no CONCATENATION has been applied, or the set of concatenated nodes of the parse tree if CONCATENATION has been applied (e.g. for the derivation in Fig. 3, $|T| = 2$). For each top node $t \in T$, we take the root node of the construction's meaning, and define its **semantic frame** to consist of all children (roles) and grandchildren (role-fillers) of the node in the situation it maps to. The model then forms a novel construction c_{syn} by taking all the constructions in the parse whose semantic root nodes point to a node in this semantic frame, referring to those as the set R of semantically related constructions. As the novel meaning of c_{syn} , the model takes the subgraph of the situation mapped to by the joint mapping of all constructional meanings of constructions in R .

R , as well as all phonologically specified constituents of t itself, are then linearized as the constituents of c_{syn} . The novel construction thus constitutes a construction with a higher arity, 'joining' several previously independent constructions. Fig. 4 illustrates the syntagmatization mechanism.

Paradigmatization Due to our usage-driven approach, all learning mechanisms so far give us maximally concrete constructions. In order for the model to generalize beyond the observed input, some degree of abstraction is needed. The model does so with the PARADIGMATIZATION mechanism. This mechanism recursively looks for minimal abstractions (cf. Tomasello 2003, 123) over the constructions in C and adds those to C , thus creating a full-inheritance network (cf. Langacker 1989, 63-76).

An abstraction over a set of constructions is made if there is an overlapping subgraph between the meanings of the constructions, where every node of the subgraph is the non-empty feature set intersection between two mapped nodes of the constructional meanings. Furthermore, the con-

¹This follows Langacker's (2009) claim that the processed concrete usage events should leave traces in the learner's mind.

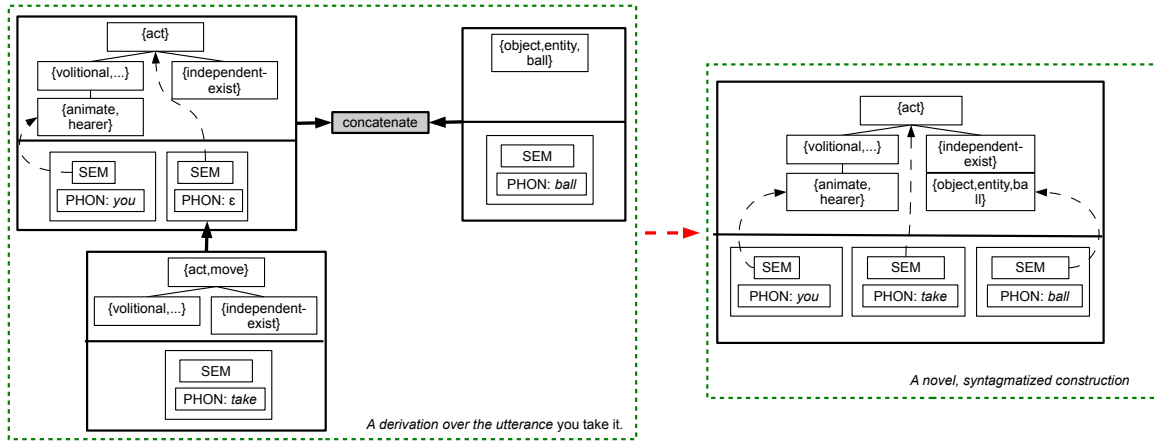


Figure 4: The SYNTAGMATIZATION mechanism. The mechanism takes a derivation as its input and reinterprets it as a novel construction of higher arity).

stituents must be mappable: both constructions have the same number of constituents and the paired constituents point to a mapped node of the meaning. The meaning of the abstracted construction is then set to this overlapping subgraph, which is the lowest possible semantic abstraction over the constructions. The constituents of this new abstraction have a specified phonological form if the more concrete constructions share the same word, and an unspecified one otherwise. The count of an abstracted construction is given by the cardinality of the set of its direct descendants in the network. This generalizes Bybee’s (1995) idea about type frequency as a proxy for productivity to a network structure. Fig. 5 illustrates the paradigmmatization mechanism.

5 Experimental set-up

The model is incrementally presented with U, S pairings based on Alishahi & Stevenson’s (2010) generation procedure. In this procedure, an utterance and a semantic frame expressing its meaning (a situation) are generated. The generation procedure follows distributions occurring in a corpus of child-directed speech. As we are interested in the performance of the model under propositional uncertainty, we add a parametrized number of randomly sampled situations, so that S consists of the situation the speaker intends to refer to (s_{correct}) and a number of situations the speaker does not intend to refer to.² Here, we set the number of ad-

²We are currently researching the effects of sampling non-correct situations that have a greater likelihood of overlap

ditional situations to be 1 or 5; the other parameter of the model, the size of the memory buffer, is set to 5 exemplars.

For the **comprehension** experiment, we evaluate the model’s performance parsing the input items, averaging over every 50 U, S pairs. We track the ability to **identify** the intended situation from S . Identification succeeds if the best parse maps to s_{correct} , i.e. if $s_{\text{identified}} = s_{\text{correct}}$. Next, **situation coverage** expresses what proportion of $s_{\text{identified}}$ has been interpreted and thus how rich the meanings of the used constructions are. It is defined as the number of nodes of the interpretation of the best parse, divided by the number of nodes of $s_{\text{identified}}$. Finally, **utterance coverage** tells us what proportion of U has been parsed with constructions (excluding IGNORED; including BOOTSTRAPPED words). The measure expresses the proportion of the signal that the learner (correctly or incorrectly) is able to interpret.

For exploring language **production**, the model receives a situation, and (given the construction) finds the most probable, maximally expressive, fully lexicalized derivation expressing it. That is: among all derivations terminating in phonologically specified constituents, it selects the derivations that cover the most semantic nodes of the given situation. In the case of multiple such derivations, it selects the most probable one, following the probability model in Section 3. We only allow for the COMBINATION operator in the derivations, as BOOTSTRAPPING and IGNORE re-

with the intended situation, to reflect more realistic input (cf. Siskind 1996).

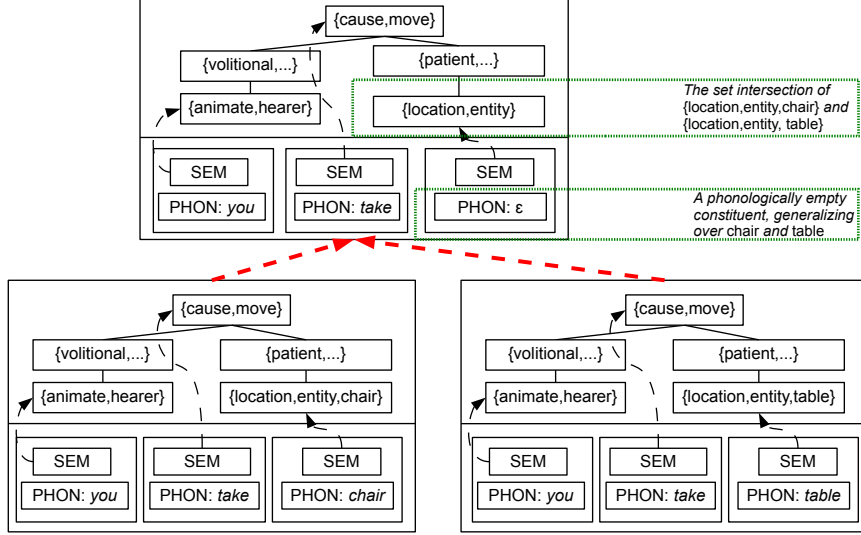


Figure 5: The PARADIGMATIZATION mechanism. The construction on top is an abstraction obtained over the two constructions at the bottom.

fer to words in a given U , and CONCATENATE is a back-off method for analyzing more of U than the construction allows for. The situations used in the generation experiment do not occur in the training items, so that we truly measure the model’s ability to generate utterances for novel situations.

The phonologically specified leaf nodes of the best derivation constitute the generated utterance U_{gen} . U_{gen} is evaluated on the basis of its **mean length**, in number of words, its **situation coverage**, as defined in the comprehension experiment, and its **utterance precision** and **utterance recall**. To calculate these, we take the maximally overlapping subsequence U_{overlap} between the actual utterance U_{act} associated with the situation and U_{gen} . Utterance precision (how many words are generated correctly) and utterance recall (how many of the correct words are generated) are defined as:

$$\text{Utterance precision} = \frac{|U_{\text{overlap}}|}{|U_{\text{gen}}|} \quad (4)$$

$$\text{Utterance recall} = \frac{|U_{\text{overlap}}|}{|U_{\text{act}}|} \quad (5)$$

Because the U, S -pairs on which the model was trained, are generated randomly, we show results for comprehension and production averaged over 5 simulations.

6 Experiments

A central motivation for the development of this model is to account for early grammatical production: can we simulate the developmental pattern of the growth of utterance length and a growing potential for generalization? The same constructions underlying these productions should, at the same time, also account for the learner’s increasing grasp of the meaning of U . To explore the model’s performance in both domains, we present a comprehension and a generation experiment.

6.1 Comprehension results

Fig. 6a gives us the results over time of the comprehension measures given a propositional uncertainty of 1, i.e. one situation besides s_{correct} in S . Overall, the model understands the utterances increasingly well. After 2000 input items, the model identifies s_{correct} in 95% of the cases. With higher levels of propositional uncertainty (not shown here), performance is still relatively robust: given 5 incorrect situations in S , s_{correct} is identified in 62% of all cases (random guessing gives a score of 17%, or $\frac{1}{6}$). Similarly, the proportion of the situation interpreted and the proportion of the utterance analyzed go up over time. This means that the model builds up an increasing repertoire of constructions that allow it to analyze larger parts of the utterance and the situations it identifies. It is important to realize that these mea-

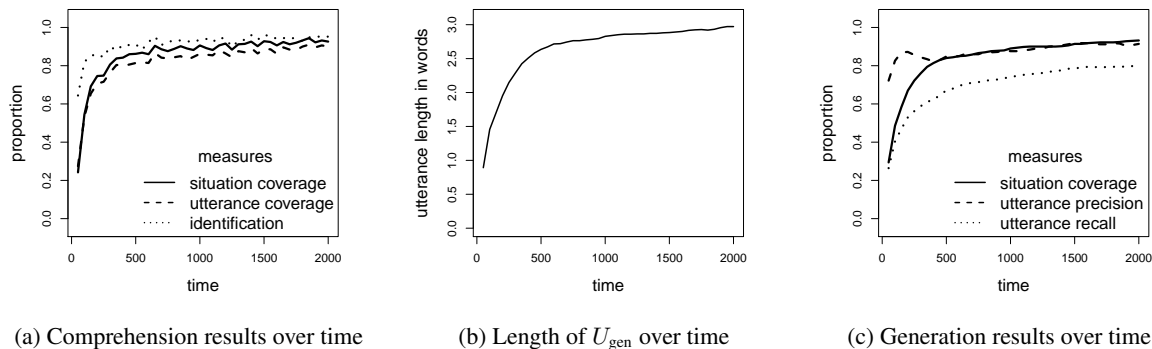


Figure 6: Quantitative results for the comprehension and generation experiments

asures do not display what proportion of the utterance or situation is analyzed correctly.

6.2 Generation results

Quantitative results Fig. 6b shows that the average utterance length increases over time. This indicates that the number of constituents of the used constructions grows. Next, Fig. 6c shows the performance of the model on the generation task. After 2000 input items, the model generates productions expressing 93% of the situation, with an utterance precision of 0.91, and an utterance recall of 0.81. Given a propositional uncertainty of 5, these go down to 79%, 0.76 and 0.59 respectively.

Comparing the utterance precision and recall over time, we can see that the utterance precision is high from the start, whereas the recall gradually increases. This is in line with the observation that children predominantly produce errors of omission (leaving linguistic material out an adult speaker would produce), and few errors of comission (producing linguistic material an adult speaker would not produce).

Qualitative results Tracking individual productions given specific situations over time allows us to study in detail what the model is doing. Here, we look at one case qualitatively. Given the situation for which the U_{act} is *she put them away*, the model generates, over time, the utterances in Table 1. The brackets show the internal hierarchical structure of the derivation. This development illustrates several interesting aspects of the model. First, as discussed earlier, the model mostly makes errors of omission: earlier productions leave out more words found in the adult utterances. Only at $t = 550$, the model makes an error of commission, using the word *in* erroneously.

t	[[she] put]	[she [put]]	[[she] [put] [in]]	[[she] put them [away]]	[[she] put [them]]	[[she] put them [away]]	[[she] put [them] away]	[[she] put them away]
50								
500								
550								
600								
950								
1000								
1050								
1400								

Table 1: Generations over time t for one situation.

Starting from $t = 600$ (except at $t = 950$), the model generates the correct utterance, but the derivations leading to this production differ. At $t = 550$, for instance, the learner combines a completely non-phonologically specific construction for which the constituents refer to the agent, action and goal location, with three ‘lexical’ constructions that fill in the words for those items.. The constructions used after $t = 550$ are all more specific, combining 3, or even only 2 constructions ($t \geq 1400$) where the entire sequence of words “put them away” arises from a single construction.

Using less abstract constructions over time seems contrary to the usage-based idea that constructions become more abstract over the course of acquisition. However, this result follows from the way the probability model is defined. More specific constructions that are able to account for the input will entail fewer combinations, and a derivation with fewer combination operations will often be more likely than one with more such operations. Given equal expressivity of the situation, the former derivation will be selected over the latter in generation.

The effect is indeed in line with another concept hypothesized to play a role in language acquisition on a usage-based account, viz. pre-emption (Gold-

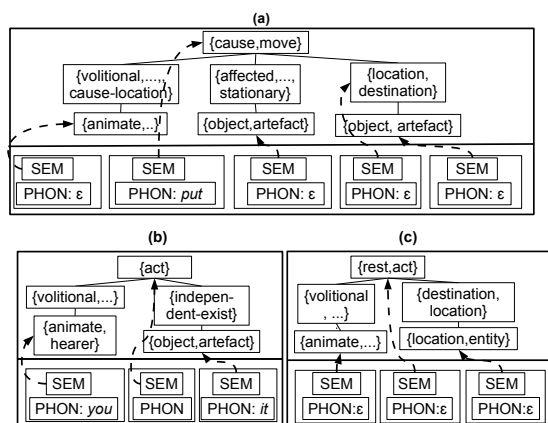


Figure 7: Some representations at $t = 2000$

berg, 2006, 94-95). Pre-emption is the effect that a language user will select a more concrete representation over the combination of more abstract ones. The effect can be reconceptualized in this model as an epiphenomenon of the way the probability model works: simply because combining fewer constructions in a derivation is often more probable than combining more constructions, the former derivation will be selected over the latter. Pre-emption is typically invoked to explain the blocking of overgeneralization patterns, and an interesting future step will be to see if the model can simulate developmental patterns for well-known cases of overgeneralization errors.

The potential for abstraction The paradigmatic operation allows the model to go beyond observed concrete instances of form-meaning pairings: without it, unseen situations could never be fully expressed. Despite this potential, we have seen that the model relies on highly concrete constructions. The concreteness of the used patterns, however, does not imply the absence of more abstract representations. Fig. 7 gives three examples of constructions in C in one simulation. Construction (a) could be seen as a verb-island construction (Tomasello, 1992, 23-24). The second constituent is phonologically specified with *put*, and the other arguments are open, but mapped to specific semantic functions. This pattern allows for the expression of many caused-motion events. Construction (b) is the inverse of (a): the arguments are phonologically specified, but the verb-slot is open. This would be a case of a pronominal argument frame [*you V it*], which have been found to be helpful in the bootstrapping of verbal mean-

ings (Tomasello, 2001). Finally, (c) presents a case of full abstraction. This construction licenses utterances such as *I sit here, you stay there* and erroneous ones like *he sits on* (which, again, will be pre-empted in the generation of utterances if more concrete constructions licence *he sits on it*).

Summarizing, abstract constructions are acquired, but only used for those cases in which no concrete construction is available. This is in line with the usage-based hypotheses that abstract constructions do emerge, but that for much of language production, a language user can rely on highly concrete patterns. A next step will be to measure the development of abstractness and length over the constructions themselves, rather than the parses and generations they allow.

7 Conclusion

This, admittedly complex, model forms an attempt to model different learning mechanisms in interaction from a usage-based constructionist perspective. Starting with an empty set of linguistic representations, the model acquires words and grammatical constructions simultaneously. The learning mechanisms allow the model to build up increasingly **abstract**, as well as increasingly **long** constructions. With these developing representations, we showed how the model gets better over time at understanding the input item, performing relatively robustly under propositional uncertainty.

Moreover, in the generation experiment, the model shows patterns of production (increasingly long utterances) similar to those of children. An important future step will be to look at these productions more closely and investigate if they also converge on more detailed patterns of development in the production of children (e.g. item-specificity, as hypothesized on the usage-based view). Despite highly concrete constructions sufficing for most of production, inspection of the acquired representations tells us that more abstract constructions are learned as well. Here, an interesting next step would be to simulate patterns of overgeneralization in children's production.

Acknowledgements

We would like to thank three anonymous reviewers for their valuable and thoughtful comments. We gratefully acknowledge the funding of BB through NWO of the Netherlands (322.70.001) and AF and SS through NSERC of Canada.

References

- Nameera Akhtar and Michael Tomasello. 1997. Young Children's Productivity With Word Order and Verb Morphology. *Developmental Psychology*, 33(6):952–965.
- Afra Alishahi and Suzanne Stevenson. 2008. A Computational Model of Early Argument Structure Acquisition. *Cognitive Science*, 32(5):789–834.
- Afra Alishahi and Suzanne Stevenson. 2010. A computational model of learning semantic roles from child-directed language. *Language and Cognitive Processes*, 25(1):50–93.
- Ben Ambridge, Julian M Pine, and Caroline F Rowland. 2012. Semantics versus statistics in the retreat from locative overgeneralization errors. *Cognition*, 123(2):260–79.
- Colin Bannard, Elena Lieven, and Michael Tomasello. 2009. Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences of the United States of America*, 106(41):17284–9.
- Martin D.S. Braine. 1976. *Children's first word combinations*. University of Chicago Press, Chicago, IL.
- Joan Bybee. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes*, 10(5):425–455.
- Nancy C.-L. Chang. 2008. *Constructing Grammar: A computational model of the emergence of early constructions*. Dissertation, University of California, Berkeley.
- Daniel Freudenthal, Julian Pine, and Fernand Gobet. 2010. Explaining quantitative variation in the rate of Optional Infinitive errors across languages: a comparison of MOSAIC and the Variational Learning Model. *Journal of Child Language*, 37(3):643–69.
- Adele E. Goldberg. 1995. *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago University Press, Chicago, IL.
- Adele E Goldberg. 2003. Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5):219–224.
- Adele E. Goldberg. 2006. *Constructions at Work. The Nature of Generalization in Language*. Oxford University Press, Oxford.
- Paul Kay. 2002. An Informal Sketch of a Formal Architecture for Construction Grammar. *Grammars*, 5:1–19.
- Tom Kwiatkowski, Sharon Goldwater, Luke Zettlemoyer, and Mark Steedman. 2012. A Probabilistic Model of Syntactic and Semantic Acquisition from Child-Directed Utterances and their Meanings. In *Proceedings EACL*.
- Ronald W. Langacker. 1989. *Foundations of Cognitive Grammar, Volume I*. Stanford University Press.
- Ronald W. Langacker. 2009. A dynamic view of usage and language acquisition. *Cognitive Linguistics*, 20(3):627–640.
- Jeffrey M Siskind. 1996. A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61(1-2):39–91.
- Michael Tomasello. 1992. *First Verbs: A study of early grammatical development*. Cambridge University Press, Cambridge, UK.
- Michael Tomasello. 2001. Perceiving intentions and learning words in the second year of life. In Melissa Bowerman and Stephen C. Levinson, editors, *Language Acquisition and Conceptual Development*, chapter 5, pages 132–158. Cambridge University Press, Cambridge, UK.
- Michael Tomasello. 2003. *Constructing a language: A Usage-Based Theory of Language Acquisition*. Harvard University Press, Cambridge, MA.
- Arie Verhagen. 2009. The conception of constructions as complex signs. Emergence of structure and reduction to usage. *Constructions and Frames*, 1:119–152.