



Universiteit
Leiden
The Netherlands

Constructions emerging : a usage-based model of the acquisition of grammar

Beekhuizen, B.F.

Citation

Beekhuizen, B. F. (2015, September 22). *Constructions emerging : a usage-based model of the acquisition of grammar*. LOT dissertation series. LOT, Utrecht. Retrieved from <https://hdl.handle.net/1887/35460>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/35460>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/35460> holds various files of this Leiden University dissertation

Author: Beekhuizen, Barend

Title: Constructions emerging : a usage-based model of the acquisition of grammar

Issue Date: 2015-09-22

Summary

Understanding how children acquire the language of their community within a limited amount of time is a central question in linguistics. The usage-based constructivist approach to language acquisition holds that children do so by using domain-general learning mechanisms such as social cognition and pattern recognizing mechanisms. Computational cognitive modeling (simulating a child's behavior by formalizing and implementing important aspects of these hypotheses as software) is becoming an increasingly important method in the field of language acquisition. This dissertation addresses four central issues in the field of language acquisition and computational cognitive modeling:

- Achieving greater comprehensiveness of computational cognitive models: the model should be able to produce, as well as interpret utterances, and not just a part of the process.
- Achieving greater naturalism in the computational modeling of the acquisition of meaning: the interpretability of utterances should be as realistic as possible.
- A reappraisal of the starting-small hypothesis within the usage-based framework: children do not only break down larger wholes into their component parts, they also learn to arrive at larger linguistic structures by combining smaller ones.
- A reassessment of proposed learning mechanisms (cognitive) and algorithms (computational): many learning mechanisms are still framed in deductivist or rationalist terms, two perspectives on cognition which do not connect naturally to the usage-based approach.

Besides these particular theoretical issues, I set out a list of general theoretical desiderata and empirical explananda the model has to satisfy in chapter 2. Previous models have made important contributions by focussing on parts of

this list and my main aim in developing yet another model was to bring these insights together.

If we want to build a comprehensive model, that is: one that can interpret as well as produce utterances, we need to have a hypothesis of how children arrive at an understanding of the communicative intention without the help of language. Computational models that deal with meaning typically have a set of situations to which the utterance potentially refers. In chapter 4, I studied the realism of this assumption. I found that the levels for noise (the absence of the meaning of an element of the utterance from experience) and uncertainty (the overwhelming presence of possible meanings that are not referred to in experience) typically used in computational modeling studies are low compared to the ones we find in actual caregiver-child interaction. I studied the latter by looking at a corpus of videotaped caregiver-child interaction and annotated the corpus for all conceptual elements reasonably thought to be present in the situation around the speech situation. Another insight following from this study was that chains of events are highly dependent on each other: if the mother engages in an action with a ball, it is very likely that she will engage in another action with the ball afterwards, or perhaps in the same action with another object. Given the tediousness of hand-coding the data, this method did not prove scalable to the demands of a computational model. The study of these properties of interaction 'in the wild', however, did lead to an adaptation of Alishahi & Stevenson's (2010) input generation procedure. In this adapted procedure, we generate pairs of an utterance and the situational context in which the utterance occurs, with the latter consisting of a set of situations, one of which is the target situation, unless the target situation is absent. Notably, the similarity of the situations within the situational context, and between subsequent situational contexts to each other is given by the similarity we found in the caregiver-child interaction. Furthermore, the setting of the parameters for noise and uncertainty was derived from the video data as well.

In chapter 3 I formalize the model: the Syntagmatic-Paradigmatic Learner (SPL). The model starts off with no linguistic-representational content, and learns to comprehend as well as produce utterances. SPL processes utterances in a context of situations, and in doing so, gradually builds up a construction, an inventory of both lexical and grammatical constructions. The 'learning mechanisms' involved in the learning process are best thought of as mere traces of processing operations, rather than actual hypothesis-testing operations (which is the metaphor, grounded in deductivist thought, that is often used to describe the acquisition of linguistic representations). SPL uses the representational format of the construction, a pairing of signifying elements (both phonological and conceptual) and a signified conceptualization.

For every processed input item, the model arrives at an optimal analysis, and does so without engaging in utterance-wide optimization. That is: SPL processes the utterance linearly and while keeping track of only the most likely analysis up to that point. The best analysis constitutes the input for SPL's

learning mechanisms. Through a set of learning mechanisms, SPL gradually builds up an inventory of constructions allowing it to comprehend and produce utterances. The learning mechanisms constitute the central innovation of the model in the aim to stay close to the usage-based approach as set out by Langacker (1988). I believe this aim has been fulfilled in the design of the model in several ways. Crucially, all of the learning mechanisms, with perhaps the exception of cross-situational learning, are online mechanisms. That is: they do not constitute post-hoc operations on the construction (the inventory of constructions), but rather reflect the traces left by the processing of the input item. These traces are found at several levels.

First, a trace of the most concrete representations of the utterances the processes is left in the representational system of SPL through the use of most-concrete constructions. This operation has the effect that highly concrete representations, if they are reinforced often enough, can become stronger over time. We can interpret this as the formation of category prototypes: the well-reinforced, highly-concrete representations are readily available to the model in analyzing and generating utterances.

Second, the mechanism of reinforcing the most-concrete used constructions, i.e. the most-concrete constructions, allows the model to accrue reinforcement mass for those constructions that are used frequently. The effect of this operation is that abstract constructions may obtain reinforcement if they are used to analyze utterances. Because the model only reinforces the most-concrete used construction, the reinforcement operation rewards patterns that are actually used. The usefulness of a construction is therefore determined by its frequency of use. Notably, this design feature implements Bybee's (2006) notion of type frequency. An abstract construction will typically only be reinforced once for each unique usage event for which it is used in an analysis. If the same usage event is encountered again, it is very likely that the more concrete construction blocks the use of the more abstract one. Routinization through high token frequency follows from the same learning operation: if a construction is used frequently, it is more readily available for subsequent analyses. If this construction happens to be a highly concrete one (i.e., one with many constituents lexically specified) the model will acquire such a construction as a routine.

Third, the model builds up increasingly long constructions through the use of the syntagmatization operation. Syntagmatization is the trace left by the processing of multiple, smaller, constructions for which the model has found no analysis in which they are connected to each other with a grammatical construction. These smaller constructions then form the constituents of a novel, wider, construction. Syntagmatization is the primary means through which SPL builds up grammatical constructions.

Finally, the paradigmaticization operation allows the model its potential to generalize to unseen usage events. By taking the joint structure of any two constructions that have been reinforced, the paradigmaticization 'extracts' abstractions from more concrete constructions. These abstractions, however, are

only extracted in the implementational sense: as no selection over them takes place, they can be considered immanent in the more concrete constructions from which they are abstracted, by simply restating their overlap. However, through the reinforcement of the most-concrete used construction, they can be reinforced themselves, in a way akin to Langacker's (2009) description of how abstractions may obtain unit status without the more concrete patterns doing so. This way, selection of 'good' or 'useful' abstractions takes place, but without any selection mechanism performing a global evaluation of the usefulness of a novel abstraction.

The model gets off the ground by the cross-situational learning mechanism, which compares recent usage events and extracts any reliable overlap as initial lexical constructions. Another way of obtaining lexical constructions is through the bootstrap operation. Bootstrapping is a property of the utterance analysis mechanism that fills a non-phonologically-specified slot of a construction with a substring of the utterance, by assuming that substring is an actual word filling that slot.

Both cross-situational learning and bootstrapping allow for the extraction of chunks: lexical constructions that are larger than a single word in the 'adult' language. These chunks, unlike what many within the usage-based framework assume, are not broken down by the paradigmaticization operation. This would require the model to engage in a post-hoc re-analysis of the chunks, which was an operation I wanted to avoid, as it makes learning more than a mere by-product of processing.

I argued in chapter 3 that the developed model reasonably succeeds in satisfying the desiderata set out in chapter 2. To the best of my knowledge, it constitutes the first usage-based computational model that is able to analyze and produce utterances while starting its development with no representational content. Furthermore, I believe it most closely instantiates the full set of ideas put forward within the usage based perspective: the representations are both qualitatively and quantitatively grounded in the linguistic usage events: their reinforcement depends on their frequency of use in analyzing linguistic usage events. Any learned abstractions are furthermore immanent: they merely restate commonalities across more concrete constructions rather than extracting novel cognitive representations from the more concrete constructions. In analyzing utterances, SPL reasonably satisfies the constraints on the realism of processing. Although this was not the focus of this dissertation, it satisfies the baseline conditions that processing is incremental over the utterance and does not involve the search for an optimal analysis over the full utterance.

I evaluated SPL's behavior both in a comprehension (chapter 5) and a production (chapter 7) experiment. In the comprehension experiment, I looked at the performance of the model in identifying the correct situation out of all possible situations the utterance could refer to, as well as the coverage of the utterance and the situation with the best analysis. On all three measures, SPL gradually becomes a more competent language user over time. Similarly, for production, SPL was tested by having it generate utterances on the basis of

a situation and its construction at that point in time. The generated utterances become longer over time, and increasingly capture the linguistic material found in the utterance that would have been produced by the input generation procedure. Interestingly, the model displayed high scores of precision, or correctness, from the outset: whatever it produced was mostly correct. This is in line with the finding that children mainly make errors of omission (leaving out elements present in adults' speech), but few errors of commission (producing linguistic elements an adult would not produce).

Next, I looked at the robustness of the model. Recall that we set the parameters for the similarity of the situations in the situational context, as well as the noise and uncertainty of the situational context on the basis of the empirical study of caregiver-child interaction. We may, however, ask how the model performs given different values for these parameters. I found that if the situations are similar to each other, the model is relatively robust to higher levels of noise and uncertainty (on the measures discussed above). Generating each situation independently of the previous one creates a situational context in which the situations are more dissimilar from each other, and in that condition, noise and uncertainty do affect the model's performance negatively. This suggests that the coherence of the situational contexts in which children have their early linguistic experiences plays an important role in bootstrapping a linguistic system: even if the child misidentifies the precise situation, the erroneously identified situation likely contains many elements that are correct.

It is, however, at a more detailed level that the interesting behavioral patterns can be seen, and especially from the failure of the model to behave as we expect, we learn important things about how the mechanisms work. In the two experimental chapters, I studied several behavioral patterns of the model in qualitative detail, to try to understand why the model behaves in certain ways.

In the production experiments, we observed that the number of expressed arguments grew over time as an effect of an increasing number of syntagmatized and subsequently paradigmatic constructions being acquired. I was not able to simulate the prevalence of subject omissions, but argued that this is likely due to a lack of pragmatics and of a right-edge processing bias. What I did find was that the omission of early arguments was not only a matter of a small vocabulary: for many aspects of the situation the model had to express, it had a lexical construction available, but it simply did not have a grammatical construction ready to fit the lexical construction in.

A central question in language acquisition is why children sometimes overgeneralize argument-structure (and other) constructions and how they retreat from this overgeneralization. The overgeneralization of argument structure constructions and the subsequent retreat were modeled in chapter 7. The answer of SPL to these two questions is that it quickly builds up an inventory of abstract, generalizable, grammatical constructions (which it, however, hardly uses in comprehension) that it combines with verbs that cannot occur in these constructions (e.g., *you fall ball*). The presence of an alternative con-

struction pre-empts this kind of combinations after a phase of overgeneralization. I argued that pre-emption works in two ways. First, the more entrenched this alternative construction is, the quicker the model retreats from overgeneralization. Second, we find an entrenchment effect of the 'correct' construction: when the model experiences more cases of *ball fall* with a causative meaning (someone dropping a ball), the constructions underlying such utterances are reinforced more, and because of this, highly general constructions allowing for the overgeneralization become less entrenched. I argued that, rather than describing this as entrenchment per se, we could better regard this effect as 'latent pre-emption', that is: as a pre-emption effect that is not seen in the behavior (the model does not produce *ball fall*, as it is less expressive than *you drop ball*), but that does block the use of a novel, erroneous, combination of an abstract construction and a verb.

One interesting property of computational models is that we can study their representations independently of the model's behavior. I did so in chapter 6. A first finding reported there is that, even though all learning mechanisms are available over time, their use varies over time. For the acquisition of lexical constructions we found that cross-situational learning, the naïve method by means of which the model extracts similarities across linguistic usage events, is only used for the first few hundreds of input items. Afterwards, the model has built up an inventory of semi-open and open grammatical constructions that it can use to bootstrap the meaning of words it has not seen. The paradigmaticization operation, secondly, displays interesting 'bursts' of activity over time, meaning that the model does not arrive at abstractions gradually, but encounters exemplars that 'unlock' new subspaces of the design space of linguistic representations.

The abstractions learned by SPL display the interesting property that they are not directly obvious from the behavior of the model in comprehension and production. If we would not have looked under the hood of the model, we might have arrived at the erroneous conclusion that its representational system is very concrete. This is a false line of reasoning: given the usage-based tenet that language users prefer the use of more concrete constructions over more abstract ones (as implemented in the probability model of SPL), we expect the highly concrete constructions to show up most of the time. However, representationally, the model has great potential for making generalizations. In fact, generalizations are found rather early, and the model spends the later iterations mainly by adding more relatively concrete constructions to the abstract ones that pre-empt the latter. This is not strange, given the overgeneralization behavior we observe in both children and SPL: once abstraction is available, the model will use it for expressivity's sake, unless it has something more concrete that is equally expressive.

An interesting feature of the abstractions found in the model is that they clearly reflect the type frequencies of the items occurring in them: the transitive construction is strongly reinforced as a non-verb-specific construction, because many verbs occur in it, whereas the caused-motion construction is

only seen with two verbs, and hence reinforced in verb-island-like constructions rather than as constructions that abstract over verbs.

Reversing the perspective, we furthermore saw how certain words are more readily learned as independent lexical constructions whereas others are primarily learned as the constituents of grammatical constructions. Notably, words referring to entities ('nouns'), are typically learned as independent entities. For the other kinds of words, there was more variation, both between the words and between simulations. Pronouns are used in a lot of different contexts, hence boosting the likelihood of their independent acquisition, but they are also used frequently *within* particular constructions. What we find for pronouns, as well as for prepositions and verbs displaying similar distributions, is that they are acquired independently in some simulations, but as 'bound' elements of constructions in others. I identified three possible factors that determined a word's independence. First, the more different elements occur in a slot, the more likely it is that the abstraction over them will be used in comprehension and production, and the more likely it is that the filler word will be acquired independently. Second, the frequency of the word in the slot: the higher this value is, the more likely it is that it will not be acquired independently, as it will be reinforced as part of a grammatical construction often. Finally, the word's 'promiscuity' matters: if a word occurs across the slots of many grammatical constructions, it is more likely that it will be acquired independently.

On several aspects of the representations, we found high degrees of 'individual' variation between the simulations: the abstraction of the representations as well as the relative independence of various words varied between simulations. This is interesting, as the various simulations display grossly the same behavior – they perform equally well on the global tasks in comprehension and production.

