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
Comprehension experiments

5.1 Measuring comprehension

The previous two chapters set out a computational model of early grammar acquisition and a procedure for generating realistic input items. The time has come to look at the behavior of the model given these two. In this chapter, we look at the ability of the model to understand the utterances it processes. Recall that, at every turn, the model is presented with an utterance in the context of a number of situations, one of which may be the situation the speaker refers to. Can SPL, given noise and uncertainty in the situation, build up an inventory of symbolic units allowing it to comprehend the utterances? This question first requires us to define what understanding means in formal terms. That is: how do we define and operationalize ‘comprehension’?

Because the input items are generated randomly, we run 10 simulations of 10,000 input items. The latter number was established on the basis of prior testing to be the amount of input items when most scores had become stable. Recall that the referential uncertainty was found to be 15 entities (events, entities) in section 4.3.2. Translating this to a number of situations, we set the number of situations co-present with the utterance to be 6 (I will henceforth call the propositional uncertainty parameter uncertainty). It is hard to establish a motivated number of situations, but given the overlap between situations (given the continuation parameters), having six situations co-present is roughly equivalent to having 15 unique entities (not counting the roles). One of these six situations is the target situation, while the other five are distractors. Furthermore, we set the value for propositional noise, $P_{\text{noise}}$, to 0.1, meaning
that in one out of ten situations, the target situation is absent.

5.1. General evaluation

A first measure of successful comprehension is the ability of the model to identify the target situation $s_{\text{target}}$ out of all candidate situation $S$. Recall that SPL always identifies a situation $s_{\text{identified}}$ as the situation the speaker was thought to refer to. The identification score of an input item, then, is 1 if $s_{\text{identified}} = s_{\text{target}}$ and 0 otherwise. Because the noise is set to 0.1 and the uncertainty to 5 situations, there are 6 situations in the situational context $S$ in 90% of the cases, and 5 in 10%. In that latter 10%, the model can, moreover, not retrieve the target situation, because it is simply absent. A chance baseline for identification is therefore $0.9 \times \frac{1}{6} = 0.15$, or one out of six for all situations in which the model can be expected to identify the target situation. Similarly, the maximum proportion of situations the model can correctly identify, or ceiling level for identification is 0.9, as in 10% of the cases, the target situation is not present.

The input items do not have a single correct mapping of the parts of the utterance to the target situation, and without such a gold standard, we cannot evaluate how well the linguistic analysis maps to parts of the situation. What we can evaluate, however, is what proportion of the utterance the model has processed, and what proportion of the identified situation (whether it is correct or incorrectly identified) is being mapped to by the best analysis. The first of these, utterance coverage is given by the proportion of the utterance $U$ that is governed by rules other than rule iii, i.e., the rule for ignoring words. In other words: the proportion of $U$ that is assigned a proper function in the analysis. Let $U_{\text{analyzed}}$ be the substring of $U$ that is governed by rules other than rule iii in the derivations underlying $a_{\text{best}}$. The utterance coverage can then be given by:

$$\text{utterance coverage} = \frac{|U_{\text{analyzed}}|}{|U|}$$ (5.1)

The second of these measures, situation coverage, works similarly, but applies to the situation. The combined mappings of all constructions used in the best analysis specify a subgraph of the (correctly or incorrectly) identified situation $s_{\text{identified}}$ that is analyzed by $a_{\text{best}}$. Let us call this subgraph $s_{\text{analyzed}}$. The situation coverage is then given as the proportion of vertices of $s$ that $s_{\text{analyzed}}$ constitutes, or:

$$\text{situation coverage} = \frac{|V(s_{\text{analyzed}})|}{|V(s_{\text{identified}})|}$$ (5.2)
5.1.2 Evaluating the used representations

Foreshadowing the study of the representations acquired by SPL in section 6, we can also inquire what the representations are that the model actually uses. For the grammatical constructions, two interesting parameters are their length (in number of constituents) and their abstraction. From Brown’s law of cumulative complexity, it follows that the inventory of linguistic representations grows more complex over time, which I take to mean that the representations become longer and the number of abstract slots increases. How this affects the choice of representations that the model actually uses in comprehension, is not evident from Brown’s law itself.

Furthermore, we cannot speak of true ‘evaluation’ of the used representations: after all, we simply do not know what representations an actual language user employs when trying to comprehend an utterance. In section 5.3 we will look at the representations and mechanisms the model employs in analyzing input items, and compare them to hypotheses within the usage-based framework.

5.2 Global evaluation

5.2.1 Identification

As can be seen in figure 5.1, the model is increasingly able to identify the correct situation, reaching an identification score between 0.7 and 0.8 after 10,000 input items, with a stabilization around 2500 items. Given a chance baseline of 0.15, the model performs well above chance, suggesting that it has learned to function relatively successfully as a communicative agent. Given a
ceiling level of 0.9, I consider the scores to be relatively close. Nonetheless, a score of 0.7 means that the model still makes a fair amount of errors (3 out of 10 cases, one of which is due to the noise).

What are those cases in which the model erroneously identifies a situation as the target situation? Looking at the errors after 10,000 input items, it seems that the only cases where the model makes errors are input items in which there are multiple, highly similar situations, and the model does not have the representational potential to tell them apart. This happens for instance when the model has misidentified attribute words like pretty or happy as markers of the role of that attribute, i.e., in an erroneous construction such as [[PERSON] [GET / get] [CHANGE-ROLE / pretty]]. When SPL has learned this construction, and next encounters two situations, one of which involves someone getting happy, and the other one involving that person getting pretty, the model is unable to choose between them, and guesses one, with a 50% chance of being correct. Other cases involve correctly learned constructions, but situations to which such constructions can equally well apply.

5.2.2 Uterance coverage

Secondly, how much of each utterance is covered by the parses at the various times? Figure 5.2a gives the results over time. The model reaches a state after approximately 1500 input items in which it is able to process almost the full utterance. It has to be kept in mind that this rapid peak may also be due to the fact that the model applies bootstrapping relatively eagerly.

When we split the values for utterance coverage over the correctly and incorrectly identified situations (figures 5.2b and 5.2c), we can observe that throughout the simulation, the analyses with incorrectly identified situations have lower utterance coverage scores. This is due to two things. First of all, there are (especially initially) several cases in which the model simply only ignores all words. Secondly, the model, in several cases, misidentifies the situation based on a partial understanding of the utterance. Given the continuity between subsequent situations, it is likely that the event and/or some participants of one situation are present in the next situation as well. When such a string of situations constitutes $S$, it is easy to see how the model, having understood one or two words, maps the analysis to the wrong situation.

Interestingly, in all simulations, the model reaches a peak in the coverage of the utterance before suffering from a slight dip in the utterance coverage, from which it recovers afterwards. When we look at the scores split over correctly and incorrectly identified target situations, we can see that the peak is found slightly earlier for the incorrectly identified ones (around 1100 – 1200) than for the correctly identified ones (around 1300 – 1400). The dip in the utterance coverage mostly occurs at the time when the model is reaching convergence in the correct identification of the situation. This means that just before the convergence, the model is applying representations that cover more of the utterance, but do so with less success. In the next stage, the model uses slightly
Comprehension experiments

Figure 5.2: Utterance coverage for 10 simulations over time.

(a) Utterance coverage for all input items.

(b) Utterance coverage for correctly identified target situations.

(c) Utterance coverage for incorrectly identified target situations.
shorter representations to analyze the utterances, covering slightly less of the utterance, but making more accurate analyses. Finally, the model starts using the longer representations again, but now in an accurate way.

What the model does here is reminiscent of a phase of syntactic creativity that is only later constrained by more ‘fitting’ representations. As we will see in section 5.3 below, and in the closer inspection of the learning mechanisms in the next chapter, the period around 2500 input items is also the moment when the model has just acquired abstract representations and has ceased to apply the syntagmatization operation frequently. This means that by then the potential for generalization, in the form of abstract constructions (constructions with few semantic constraints, obtained through paradigmatization), is present, and that afterwards the model ‘recovers’ from applying these abstractions too frequently by building up an inventory of more concrete constructions that ‘pre-empt’ the use of the abstract constructions in the analysis. The continuing accrual of relatively concrete constructions allows the model to overcome overgeneralization. As such, this robustness provides an argument for the apparent redundancy of storage, as many within the usage-based approach have argued (Langacker 1988, Beekhuizen, Bod & Verhagen 2014).

Let us have a look at an example that illustrates this. In one of the simulations, the model encounters, after some 200 input items, the utterance in example (29). The utterance illustrates a construction which is relatively rare (compared to other kinds of three-word utterances that are formed on the basis of a transitive construction). The optimal analysis the model assigns to this utterance is given in example (30). It involves an abstract intransitive construction and the bootstrapping of go. Some 300 input items later, the model encounters the same utterance, but now uses the analysis in example (31). This is a regular transitive construction, in which the action of a person on an object is expressed. With this construction, SPL erroneously takes the utterance to refer to a caused-motion event. Nonetheless, it covers the full utterance, as opposed to the analysis with the intransitive construction.

Finally, after another 300 input items, the model has an intransitive motion construction, as shown in example (32), which is combined with the known meanings of go and out. From this example, we can glean that the model eagerly applies abstract patterns to situations in which they lead to misinterpretations. These errors are overcome once a larger inventory of constructions is built up.

(29) *you go out*

(30) \[[\text{ENTITY}]\rightarrow[\text{HEARER / you}][\text{EVENT}]\rightarrow(\text{GO / go})\]

(31) \[[\text{PERSON}]\rightarrow[\text{HEARER / you}][\text{CAUSE}]\rightarrow(\text{CAUSE-MOVE / go})[\text{OBJECT}]\rightarrow(\text{ARTEFACT / out})\]

(32) \[[\text{PERSON}]\rightarrow[\text{HEARER / you}][\text{EVENT}]\rightarrow[\text{CAUSE-MOVE / go}][\text{ROLE}]\rightarrow[\text{DESTINATION-ROLE(LOCATION) / out}]\]
Comprehension experiments

5.2.3 Situation coverage

We find a highly similar pattern for the model’s understanding of the parts of the situation that are being signified by the utterance, or the **situation coverage** in figure 5.3a. The model quickly achieves high levels of understanding of the situation, with a stabilization around 1500 input items.

Again, we see a difference between the correctly and incorrectly identified target utterances (figures 5.3b and 5.3c). For input items in which the model correctly identified the target situation, the situation coverage starts out relatively high (values around 0.75), whereas for input items with incorrectly identified target situations, the situation coverage starts out low (values between 0.25 and 0.50).

An interesting future step would be to have the utterance and situation coverage affect the reinforcement of the used constructions. Currently, the
5.2. Global evaluation

(a) Identification scores for nine unique noise and uncertainty settings over time given $P_{\text{reset}} = 0.05$.

(b) Identification scores for nine unique noise and uncertainty settings over time given $P_{\text{reset}} = 1$.

Figure 5.4: Identification scores given various parameter settings.

The probability of an analysis is simply penalized for not being able to parse parts of the utterance, but if an analysis involving ignored words and ignored parts of the situation is (despite this penalty) the best analysis, the used constructions receive as much of an increase as when the analysis covers all of the utterance and the identified target situation. If we allow the model to reinforce the construction proportionally to their utterance and situation coverage, erroneous analysis, and hence (often) erroneous constructions will receive less counts, and therefore be less likely to be re-used.

The important question is: would this merely be a ‘hack’, i.e., a trick to get the model to work better, or is it in some sense a cognitively motivated operation? I do not intend to give a definitive answer to that question, but it seems to me that the firmness of the belief that something is the right analysis is a feature that can be used by a model to ‘bootstrap’ itself. Furthermore, it is not a grammar-wide optimization operation, but a local effect of the processing, and therefore still in line with desideratum D2-8 (learning-as-processing).
5.2.4 Robustness to uncertainty and noise

The parameters noise and uncertainty that I used in the experiments were set on the basis of the findings in chapter 4. It would nonetheless be interesting to see how the model behaves under different settings for these parameters. Furthermore, I set the probability of generating the next event without taking the previous one into account to 0.05. This means that subsequent frames are very likely to look alike. However, we may wonder how the model behaves if all frames are independently generated (i.e., \( P_{\text{reset}} = 1 \)).

In figure 5.4a, the identification scores for nine unique parameter settings is given if we set \( P_{\text{reset}} = 0.05 \). For each unique parameter setting, three simulations were run. The noise values were set to 0, 0.1, and 0.3, and the uncertainty values to 0, 5, and 10.

Looking at uncertainty first, we can see that the model trivially performs at ceiling level (given each noise setting) if there are no non-target situations present. Adding uncertainty causes the model to misidentify the target situation more often. However, even with 10 non-target situations present, the model still identifies the target situation correctly in six out of ten cases under the no-noise condition (where randomly guessing would yield a score of 0.09). It might furthermore be that given high levels of uncertainty, more input items would be needed to arrive at some level of communicative competence: the slopes of the developmental curves for the settings noise = 0.1, uncertainty = 10 and noise = 0.3, uncertainty = 10 do not seem to have reached a point of convergence after 10,000 input items (unlike the other curves).

With noise, we see a similar pattern. Adding more noise causes the model to learn erroneous representations and apply them, even in situations where the target situation is present. However, even with three out of ten target situations being absent, the model still identifies the target correctly, given uncertainty = 5, around 58% of the cases (where the ceiling level of the performance would be 0.70).

Setting the probability of reset \( P_{\text{reset}} \) to 1 causes a more variable performance of the model (see figure 5.4b). Whenever there is uncertainty present, this has a greater negative effect on the scores than when \( P_{\text{reset}} = 0.05 \). The reason for this is that, when the model misidentifies a situation under the condition \( P_{\text{reset}} = 0.05 \), it is very likely that it still has a correct partial identification: some referents, or the action given in the situation can be the same as the one in the target situation. Nonetheless, the model acquires some correct constructions even under the most dire settings for noise and uncertainty, given that the performance with that setting (identification = 0.3 after 10,000 input items) is still more than four times as high as the chance baseline for that setting (i.e., one out of eleven of seven of out ten situations, or ±0.07).

This means that the model’s performance decays gracefully under increasingly hard conditions. Even though we motivated the parameter settings, this result supports the idea that SPL is a robust learner.
5.3 Used representations

We have seen in section 5.2 that the model is able to comprehend sentences relatively well. In this section, we have a look at the kinds of representations the model uses in analyzing the input items. From a usage-based perspective, several topics are of interest: the use of unanalyzed chunk-like structures, the use of bootstrapping to analyze unseen words, the use of the concatenation operation, the abstractness of the used constructions, and the types of abstraction (over verbs, or over nouns). A computational model like SPL allows us to look at the representations used in the analyses.

5.3.1 The use of chunks

The usage-based approach claims that in many cases, language users operate with representations that could be further analyzed, but that are not further analyzed (Arnon 2010, McCauley & Christiansen 2014a). SPL learns lexical constructions without knowing what the word boundaries are. That is to say: it has the true word boundaries, but it may extract larger units as being a single word, both through bootstrapping and cross-situational learning. We can therefore expect the model to build up an inventory of such unanalyzed-but-analyzable lexical constructions. We furthermore expect the amount of chunks used in the analysis of input items to decay over time, as the more compositional constructions, used in a wider array of cases, will become stronger and outweigh the chunks. However, we can also expect the model to continue using some chunks, as even adult language users use unanalyzed-but-analyzable language material. It should be noted here that our definition of ‘chunk’ only covers a subset of what McCauley & Christiansen (2014a) consider chunks, namely the internally unanalyzed ones. As we will see, for the internally analyzed larger units (which I call ‘lexically specific constructions’), we do find a behavior akin to the one reported in McCauley & Christiansen (2014a), viz. that their importance in use increases over time.

All of these expectations are found in the behavior of the model. I operationalize the notion of ‘chunk’ to be any construction in which there is at least one constituent consisting of more than one word. This includes constructions with more than one constituent, but for which (at least) one of the constituents has more than one word as their phonological constraint. Figure 5.5 shows the frequency of the use of chunks over time. For the first 750 input items, the model employs chunk-like constructions relatively frequently, after which the number drops, but remains stable at around 4 used chunks per 250 input items.

What are the chunks the model uses? Table 5.1 gives the most frequently used chunks for three simulations. We can see that in simulation 0, there are mainly many chunks with play with. The chunk has been syntagmatized with entity words like matches and truck to form grammatical constructions involving the chunk. Note that all chunks in simulation 0 are ‘correct’, in that they
Figure 5.5: Frequency of the use of chunks over time (summed over simulations).

<table>
<thead>
<tr>
<th>rank</th>
<th>simulation 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[PLAY(AGENT,PATIENT) / play with] (265)</td>
</tr>
<tr>
<td>2</td>
<td>[PUT(AGENT(SPEAKER)) / I put ] (33)</td>
</tr>
<tr>
<td>3</td>
<td>[COME(AGENT,DIRECTION-ROLE(LOCATION)) / out come ] (25)</td>
</tr>
<tr>
<td>4</td>
<td>[[PLAY / play with ] [[MATCHES / matches ]]] (11)</td>
</tr>
<tr>
<td>5</td>
<td>[[PLAY / play with ] [[TRUCK / truck ]]] (8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rank</th>
<th>simulation 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[THEY / here come ] (8)</td>
</tr>
<tr>
<td>2</td>
<td>[SPEAKER / they make ] (7)</td>
</tr>
<tr>
<td>3</td>
<td>[BABY / baby take ] (3)</td>
</tr>
<tr>
<td>4</td>
<td>[SPEAKER / we go ] (1)</td>
</tr>
<tr>
<td>5</td>
<td>[[SPEAKER / we go ] [[SEE / outside ]]] (1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rank</th>
<th>simulation 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[[PUT / put them ] [[DESTINATION-ROLE / in ] [[ARTEFACT ]]] (78)</td>
</tr>
<tr>
<td>2</td>
<td>[GIVE / she give ] (46)</td>
</tr>
<tr>
<td>3</td>
<td>[[SIT(SURFACE-LOCATION) / sit on ] [[ARTEFACT / it ]]] (44)</td>
</tr>
<tr>
<td>4</td>
<td>[PUT(AGENT(HEARER)) / you put ] (29)</td>
</tr>
<tr>
<td>5</td>
<td>[[SIT / sit on ] [[SPEAKER / me ]]] (27)</td>
</tr>
</tbody>
</table>

Table 5.1: Most frequent used chunks for three simulations.
do seem to capture the meaning of the words they contain.

The extraction of play with as a chunk is interesting. SPL has acquired play with with the meaning \textit{play(AGENT,INSTRUMENT)}. This can be considered to be an error, but the word \textit{with} occurs in the input generation procedure only in one other, highly infrequent, construction, namely the \texttt{[ [ ENTITY ] [ CREATE / make ] [ ENTITY ] [ SOURCE / with ] [ ENTITY ] ]} construction (e.g., I made a cookie with dough). The meaning of \textit{with} in this construction is furthermore different from that in play with. Therefore, the model ‘decides’ to use play with essentially as a bi-syllabic word denoting the action \textit{play} and its roles.

However, play is also used without \textit{with}, in utterances like you plays game. This gives the model the opportunity to learn the meaning of play by itself, which it does: it also has a \texttt{[ PLAY / play ]} construction. However, as the play with-construction covers more of the utterance, it is given preference over the lexical play-construction in the analysis of sentences containing the substring play with. We again find that the play with-chunk, and its syntagmatic extensions are used throughout development.

Table 5.1 also shows us that there is massive variation between simulations. Both simulation 4 and 7 display less use of chunks than simulation 0. This is interesting, as it gives us the well-known difference between analytic and holistic learners (Bretherton, McNew, Snyder & Bates 1982) without parameters governing that particular behavior. That is: it is not due to a change in the model that different amounts of chunk use are found in different simulations. Rather, it is merely an effect of input order, and the chance of the subsequent co-occurrence of certain utterances. If two utterances with play with in it are found subsequently with different arguments, the model will extract a play with-construction. Perhaps this can be taken to mean that the difference in what seem like learning strategies (some learner learn many chunks, while others learn few), may (also) be an effect of input order and dispersion of the input items.

Note, finally, that there is variation in the kinds of chunks the model learns in different simulations. In simulation 4, the acquired chunks all refer to the wrong entity, and hence receive little reinforcement, whereas in simulation 7, the most frequent construction involving a chunk is a semi-open construction, with put them as its first constituent, followed by in, which has its own role, followed by any ARTEFACT. Again, this is an effect of the coincidental juxtaposition of input items and the subsequent build-up of the grammar through syntagmatization and paradigmatization, which to my mind is an exciting, albeit rather extreme hypothesis following from usage-based theory that can and should be further explored using both experiments and dense corpora.

5.3.2 The use of bootstrapping

The ability to bootstrap words into open constituents of constructions is a mechanism that allows the model to analyze utterances for which it does not know all the words. Suppose that the model encounters the utterance in ex-
ample (33). Having access to a \([ \text{HEARER} / \text{you}] [ \text{SIT} / \text{sit}] [ \text{LOCATION} / \text{on}] [ \text{OBJECT}]\) construction, all the model has to do is make the assumption that microphone refers to the OBJECT in the LOCATION-role of the sitting event, and it has learned a new word, as can be seen in example (34), which gives the best analysis of example (33).

(33) you sit on microphone

(34) \([ \text{HEARER} / \text{you}] [ \text{SIT} / \text{sit}] [ \text{LOCATION} / \text{on}] [ \text{OBJECT}] \rightarrow (\text{MICROPHONE} / \text{microphone})\]

Bootstrapping provides a strong mechanism for interpreting and acquiring novel lexical constructions. However, the risk of allowing for an operation like bootstrapping is that the model will bootstrap too freely, assigning meanings to word forms that already have well-entrenched meanings associated with them. When we look at the number of bootstrapping operations over time (figure 5.6), several things can be observed. First of all: the number of bootstrapping operations decreases over time, consistent with the idea that the learner has increasingly many (lexical) constructions in her inventory. This can be expected, as the expected number of novel, unanalyzed words decreases over time (cf. figure 5.7). However, whenever a novel word type is encountered, it is most likely to be bootstrapped into a slot of a (semi)-open grammatical construction.

More interestingly, the amount of bootstrapping operations over words that have been analyzed before (i.e., for which there is a constructional representation in SPL’s grammar) decreases less rapidly than the amount of bootstrapping operations over unanalyzed words. This means that the model bootstraps words for which it already has a representation. In some cases, this
happens in noisy input items (i.e., items in which the target situation is absent). When encountering the utterance you make picture, but the meaning $\text{MAKE}($maker($\text{HEARER}(\text{you}),\text{MADE-THING}(\text{PICTURE}))$) is absent, but another situation $\text{MAKE}($maker($\text{HEARER}),\text{MADE-THING}(\text{COOKIE}))$ is present, the model will bootstrap the word picture as meaning cookie. This is not very problematic for the model, as the bootstrapped construction $\text{[COOKIE / picture]}$ will rarely if ever be reinforced in other input items. However, this does point to a design feature of the model that might be too strict, namely that it is forced to select a situation. If the model has a strong conviction the picture means PICTURE, having no situation present that contains that conceptual element should ideally force the model to consider the input item to be noisy and not consider the analysis in which picture refers to PICTURE to be the best one.

In other cases, the grammatical construction used to bootstrap the word is erroneous. In the same simulation, the model has acquired a construction $\text{[ [PERCEIVE / you look at ] [PERCEIVER-ROLE ]]}$, where the second constituent refers to the PERCEIVER role. When encountering you look at picture, the model considers picture to refer to that agent role (as if it were a nominative case marker, essentially), and bootstrap a construction $\text{[ AGENT-ROLE / picture ]}$. Again, this construction will be used in few subsequent analyses and therefore not be reinforced, but it does lead to an erroneous analysis of the mapping between the utterance and the identified situation. Here too, the fact that the model does not take the reinforcement of a $\text{[ PICTURE / picture ]}$ construction can be considered a weakness in the design of the model.

This analysis gives us an insight in the complex interaction of mechanisms that must take place when a word is being bootstrapped. On the one hand, we have the selection preferences of a slot of a construction and the number of other elements that can fill that slot. The higher this number is, the lower the
Comprehension experiments

probability of bootstrapping something new in the slot. Interestingly, this idea runs counter to Bybee’s (2006) ideas about high type frequencies (many other constructions being able to fill a slot) making a slot more extendable. However this works, there is a top-down effect of the slot of the construction. On the other hand, there are bottom-up effects of the word. If the hearer knows with a lot of certainty that a word already refers to very different meanings, he would find it very unlikely that the speaker uses it now to refer to this particular concept. It would be as if a speaker and a hearer are looking at a painting, and the speaker says what a nice book. The hearer would not, in this situation, bootstrap book in the open slot of a what a nice-X construction, because the word form book is already used in lexical constructions referring to BOOKS. In this case, the hearer would come to the conclusion that the speaker is an uncooperative communication partner. However, if the speaker said what a nice fammer, the hearer would be prone to bootstrap the meaning of the word fammer as relating to something concerning the painting or maybe an object depicted in it. Finally, there are cases where the use of a word seems like an extension of the meaning. When the speaker says what a nice Vermeer, and the hearer does not know that one can use the name of an artist metonymically for the product of their artistry, the hearer can still make the inferential step that Vermeer refers to a product of Johannes Vermeer. A new lexical representation is then added, linking Vermeer to the concept PAINTING-BY-VERMEER. The bottom-up effects of the bootstrapping thus also concern the closeness of the bootstrapped meaning to one of the known meanings, but this is likely the way radial concepts in lexical meanings emerge (cf. Lakoff 1987).

Concluding, the bootstrapping operation as implemented in SPL is a naïve one, that does what it should do, namely learn new words, but that also applies too frequently and in an underconstrained way. A possible solution is to not only take into consideration the top-down preferences of the slots of the grammatical constructions, but also the bottom-up knowledge concerning the other constructions in which the word form is already used.

5.3.3 The use of concatenation

Concatenation is the processing mechanism that allows the model to form a more encompassing interpretation of an utterance on the basis of partial analyses. SPL explicitly frames concatenation as a back-off device for cases when no better (i.e., construction-based) analysis can be found: the probability of the rule leading to a concatenation is a small, smoothed probability depending on the number of constructional analyses that can be given. As such, we can expect its use to decrease over time. Figure 5.8 shows that this is indeed the case: the number of concatenations decreases over time.

A successful case of concatenation is given in the analysis in example (35). When processing the utterance you put animal in it, SPL uses a construction [[HEARER / you] [PUT / put] [ENTITY] [GOAL-LOCATION / in]], which is combined with the lexical [ANIMAL / animal] construction. This derivation
is finally concatenated with the [ THING / it ] construction. Note that the concatenation has meaning beyond the sum of the elements: the THING-meaning is bound to the referent filling the GOAL-LOCATION-role.

\[
( [ [ HEARER / you ] [ PUT / put ] [ ENTITY ] \rightarrow [ ANIMAL / animal ] [ GOAL-LOCATION / in ] [ THING / it ] ] ]
\]

\[
PUT(PUTTER(HEARER),PUT-THING(ANIMAL), GOAL-LOCATION(THING))
\]

The design feature of concatenation as a back-off device can be doubted, however. Perhaps using something akin to concatenation is a regular way of processing utterances (cf. Frank et al. 2012), in which case the probability model would have to be adjusted. However, there still seems to be a difference between non-conventional concatenation, as implemented in SPL, and conventional but non-hierarchical processing. I leave it to the proponents of the strong non-hierarchicality thesis to develop a working model that involves meaning.

5.3.4 The length and abstraction of the used representations

At the heart of the usage-based perspective on the acquisition of grammar is the claim that grammatical representations are built up in a gradual, bottom-up fashion. As I argued earlier, this is a cognitive take on Brown’s (1973) (observational) law of cumulative complexity, which holds that more complex representations emerge in development after, and on the basis of, simpler ones. For grammatical representations, we can take this to mean that the representations become increasingly long and increasingly abstract. However,
Comprehension experiments 181

we can wonder if this implies that the used representations become more abstract. After all, the language-learning child also encounters more concrete instances of grammatical patterns, which, under the usage-based perspective, leave traces in the mind as well. Here, the old pair ‘competence and performance’ (Chomsky 1965) comes in handy. Even within a usage-based model, the potential of a model may differ from what is doing most of the time. Whereas a model may have acquired the representational potential to make all sorts of generalizations, it may be the case that the more abstract ones are only needed in few cases, because the more concrete representations pre-empt the use of the more abstract ones in use. The competence of the model is then, of course, something derived from, or immanent in the processing involved in the performance, but conceptually, we can describe the learner’s global competence distinctly from its performance in specific cases. Again we see a case of a conceptual or analytical distinction that is ontologically non-distinct, but may methodologically or analytically be separated. Applied to a usage-based perspective, it furthermore corresponds to the distinction between a static and a dynamic take, where the competence describes the state of the language user’s potential and the performance the actual use in processing of that competence.¹

Let us start with SPL’s performance first. In chapter 6, we explore the competence side in more depth, but here, we look primarily at the nature of the constructions that the model uses to analyze the utterances. In figure 5.9, the frequency of constructions of various length and abstraction over time is given. The first thing worth noticing is that the longer representations are only used to the full extent in the 1500 – 2000 bin for length-4 constructions and the 3000 – 3500 bin for length-5 constructions. Length-2 constructions are used ‘too much’ over the first 2000 input items, which is when they are used to analyze utterances for which length-3, 4 or 5 constructions would be most suited. We see a similar pattern for length-3 constructions, being overused in the 500 – 1000 bin. All in all, this means that early on, SPL analyzes input items with longer utterances by means of shorter representations, concatenation, and ignoring words, and that the development to higher-arity constructions depends on the use of these lower-arity constructions.

For the abstraction of the used representations, the analysis is slightly more complex. The end state, after 10,000 input items, is that the longer the used representation is, the higher the chance of it being a more abstract one. About

¹Allowing myself a small digression: the idea that competence and performance are ‘implemented’ in the minds of language users as distinct ‘things’ can be seen as a case of the reification of an analytical distinction into an ontological one whereas the distinction may equally well be viewed as two perspectives on the same object. As such, it constitutes a case of Gigerenzer’s (1991) tools-to-theories heuristic, in which tools of analysis shape the conception of the objects of study. Vice versa, going from the denial of this ontological distinction to a strict what-you-see-is-what-you-get approach (more formally: the analyst’s inference of the most likely grammar on the basis of behavioral patterns) is equally fallacious as it misses the logical possibility that the learner has a more abstract representational potential, but simply not uses it because more concrete constructions pre-empt the abstract ones in all but few cases.
5.3. Used representations

Figure 5.9: Frequency and abstraction of constructions of various length used in comprehension over time (summed over simulations).
Comprehension experiments

half of the length-2 constructions have no open slots, whereas for length-5 constructions this figure is around 20%. This, of course, is an effect of the kinds of utterances they are employed for. There are simply fewer unique long utterances than there are unique short ones. Nonetheless, if this effect is realistic, it has interesting consequences for the nature of the representational system. It means that the longer a construction is, the higher the likelihood of it being more abstract, all other things being equal. Perhaps this can be taken to mean that caused-motion constructions and prepositional datives have abstract representations that are more reinforced than intransitives and transitives. With the latter two being researched less intensely than the former, this question cannot be straightforwardly answered, but it would be an interesting research avenue.

Turning to the development over time, we can see that the length-4 and length-5 constructions used early on are mostly very concrete, and that they become more abstract over time. SPL employs them, despite building up an ever-growing inventory of more concrete patterns that can be re-used. To give an example, the model encountered the utterance you put her in here after some 9700 input items. The model has relatively concrete constructions available to analyze this utterance (e.g., [you put ENTITY in ENTITY], and even [you put ENTITY in here]), but it analyzes the utterance using the abstract construction in example (36), in which only put is lexically specified.

(36) [[PERSON] [PUT / put] [OBJECT] [LOCATION-ROLE] [ENTITY]]
     | PUT(PUTTER(PERSON), MOVED-OBJECT(OBJECT),
     | LOCATION-ROLE(ENTITY))

Why does the model do so? Given the high diversity in sentences expressing a caused-motion event, we can expect the more abstract constructions to be the most-concrete used construction, and therefore get reinforced relatively frequently. Furthermore, the words used in these slots are also seen in many other contexts, and are therefore also well entrenched. With the high count, and hence high probability, of the abstract construction, and the well-entrenched lexical items, the analysis involving a more abstract representation thus becomes more likely than ones involving less abstract representations. The effect here is due to a dynamic version of Bybee’s (2006) notion of type frequency: the more the abstract representation is actually used to analyze unseen utterances (i.e., utterances with novel word types – at least in that slot), the more it gets entrenched, and can therefore be used to analyze utterances for which in principle more concrete representations can be used.

Given that the model uses a variety of concrete and abstract constructions, what are the kinds of abstraction that are useful in language comprehension? Recall that under Tomasello’s (1992) hypothesis, young learners operate with verb-island constructions, consisting of verbs and their highly-specific roles. Dodson & Tomasello (1998) added the possibility of learners using argument-frame constructions, in which the arguments but not the verb is specified.
This especially happens with pronouns, in hypothesized constructions such as [[SPEAKER/I][ACTION][OBJECT/it]].

The model gives peculiar results when we look at the amount of constructions with lexically-specific verbs slots being used (figure 5.10). The length-4 and length-5 constructions that the model uses initially all have verbs specified. Afterwards, the model discovers that there is regularity in the variation (e.g., you can PUT a ball on the table, but also TAKE it from the box), and some constructions with abstract verb slots are used. However, with the increasing accrual of more concrete patterns that (crucially) involve a specific verb, the model reverts to using verb-island-like constructions. These verb-island constructions potentially have all other slots of the construction being abstract (as in example (36) above), but the verb is fixed. Before jumping to conclusions, it should be said that the model has a low number of verbs occurring in length-4 and length-5 constructions, and the type frequency of the verbs in this slot therefore is rather low. Perhaps if the model were exposed to a wider array of verbs in these slots, it would use constructions with abstract verb slots more often.
When we look at length-3 constructions, next, we see that the abstract-verb constructions form a majority. Again, I believe this is an effect of the nature of the distribution of verbs. As many verbs occur in length-3 constructions, and as several more are at least at some point used in length-3 constructions in comprehension, the constructions with abstract verb slot receive more reinforcement, and are hence more likely to be used later (again, despite there being more concrete patterns in the models representational potential as well).

Length-2 constructions, finally, look more like length-4 and length-5 constructions than length-3 constructions. Initially, the model uses some abstract-verb constructions, but these are given up in favor of verb-island constructions later. Here too, I believe this effect is due to the nature of the distribution of the verbs in the input: there are few verbs that occur in the intransitive pattern, and therefore the model finds little use for a general intransitive.

5.4 Desiderata and explananda

In chapter 2, I set out several desiderata for a usage-based computational model. In chapter 3, I presented the Syntagmatic-Paradigmatic Learner that was intended to meet these desiderata. Using the parameter settings obtained through the study in chapter 4, the present chapter constitutes a first evaluation of the model in terms of its behavior in comprehending utterances. These parameter settings are both stricter than most models’ (there is more noise and uncertainty; cf. the comparison in section 4.2.1), and allow for more informative sets of candidate situations because of the overlap between situations. The overlap between situations constitutes a problem in identifying the correct target situation, but also makes failing to do so less problematic – when the model identifies the wrong situation, it still gets some mappings between the utterance and conceptual elements right.

Given this input procedure, we have seen in this chapter how SPL is increasingly able to understand the input items it processes. Not only does it correctly identify the target situation more frequently (around 70–80% of the cases after 10,000 input items, given a baseline of 15% and a ceiling of 90%), it also is increasingly able to analyze the full utterance and map it to many elements of the situation. Looking at the mechanisms and representations used by the models provides insight in the way SPL achieves this. By recognizing multiple words and concatenating them, the model is able to understand larger parts of the utterance. The trace these analyses leave, via syntagmatization, leads to the first grammatical constructions, which are then abstracted over if multiple similar ones have been seen. The (semi-)abstract constructions further bolster the potential for analyzing input items by allowing for novel combinations of constructions, but also by enabling the model to interpret unseen words through bootstrapping.

We are now in the position to evaluate the model against several of the desiderata and explananda. In chapter 3, I argued how the model in princi-
5.4. Desiderata and explananda

ple satisfies these, but we would like to know if that promise is made true by the behavior of the model. Concerning desideratum D2, being able to do both comprehension and production, we have seen that the model performs well in the comprehension experiment given a level of noise and uncertainty that is higher than that of most models, but with a set of candidate target situations that consists of highly similar situations, thereby aiding the model as well. This points to an often overlooked aspect in the discussion of referential uncertainty: even if the child picks out the wrong situation, or the wrong conception of a situation (as in multiply perspectivizable events, e.g., chase/flee), it will get many other things right, which eventually helps the language-learning child in verbally getting off the ground.

We have seen some remarkable effects of the quantitative grounding (D4-2) of the representational system in the usage events. Besides the obvious effects of entrenchment of more frequently processed representations, we found that the used length-3 constructions tend to be more abstract than the other constructions. I identified two reasons for this. First, the number of verb types in the ‘transitive’ construction is simply higher than that of the other constructions. Second, many length-4 and length-5 constructions develop from the length-3 constructions (using a ‘transitive’ construction, a lexical item and concatenation). The effect of this is that even more verb types are observed in length-3 constructions, thus reinforcing the abstract representation of this construction further.

This brings us to the cumulative complexity observed in the model (D6-1). We have seen that the longer constructions emerge later in development and are formed on the basis of shorter representations with concatenation. Abstract constructions show up rather early in development, but their use becomes increasingly constrained by more concrete ones, unless the abstract construction and the lexical constructions filling the slots are reinforced to such an extent that they outweigh the use of more concrete representations. As such, SPL is an avid generalizer (cf. Naigles et al. 2009), but I do not consider this property to be contrary to the usage-based perspective. It may be the case that language-learning children are not conservative in forming abstractions, but rather that their use of abstractions becomes increasingly constrained by the growing inventory of more concrete constructions. At the stage where the model has abstractions, but not many concrete constructions ‘pre-empting’ them, abstract patterns are used. This may also be the stage where overgeneralizations are found, a topic to which we will return in chapter 7. It seems that the distinction between a learner’s (usage-based) competence and her performance is a relevant conceptual distinction: a potential for abstraction does not entail its use. This view, again, is not at odds with the usage-based perspective: all representations and their degree of entrenchment is still grounded in the experienced usage events.

A further property of SPL in which it differs from the other usage-based computational models, but similar to Kwiatkowski’s (2011) model is that it acquires an inventory of both lexical and grammatical constructions at the same
Comprehension experiments

Unlike in Kwiatkowski’s model, the set of grammatical representations is unconstrained, but SPL fares well in solving this daunting task. All processing and learning mechanisms involved are needed for this task: cross-situational learning to get the model started, various forms of reinforcement to find out which representations are the most useful, concatenation to build up the grammatical constructions, abstraction to generalize, and bootstrapping to acquire lexical constructions quickly.

Despite the availability of all mechanisms at all times, some are used more in early development than others. When we look at the processing mechanisms, we can observe that the bootstrapping operation peaks in frequency early, but not at the beginning of development, and that the use of concatenation decreases over time. In this sense, the model reflects the insights of Hollich et al. (2000), who argue that various cues and various mechanisms may be at work at various times in development. Again, the competence-performance distinction is insightful: all mechanisms discussed are in principle available, but it is their use that varies over time (D6-4).

Two of the explananda are partially satisfied in the comprehension experiment. We have observed that verbs behave conservatively in the fact that most constructions used in the comprehension process are verb-island constructions (E1). The model has the potential for using other, non-verb-specific, constructions, but does not do so, suggesting that SPL finds it more useful to structure its comprehension around verb-island constructions rather than more general verb-argument constructions. However, the model does have constructions available in which the verb is not lexically specified.

Obviously, SPL does not get everything right, and several aspects of the model are worth reconsidering in future research. One that should be pointed out here is that the current implementation is highly inefficient when an utterance is analyzed without any situational context present. We should expect the model to be able to do so at some point. Simulating this artificial situation requires some changes to the model (e.g., allowing it to build up situational representations within the space of all possible situations, rather than 2 or 5 or 10). Practically, this would be highly inefficient. Perhaps the design choice present in Chang’s (2008) model, viz. to have a layer of constructional meaning first being inferred, which is only then resolved against the situational context (which, in the case of this hypothetical experiment would be absent), resolves this, and it does not seem like a major step to change the model to share this design feature. A downside of that feature, however, is that the pragmatic resolution only takes place post-hoc, that is: after the semantic analysis has been completed. A realistic analyzer would do this online. That is to say: after hearing the man, an analyzer would have to resolve it already in the situational context (“what likely definite reference to an instance of the category man can be given here?”), rather than it having to wait until the full utterance is processed.

SPL functions well as a model of acquiring communicative competence when comprehension is concerned. However, we would also like to know
how well it satisfies the second half of desideratum D2, namely production. I discuss several aspects of production in chapter 7. Before we go there, I would like to dwell on the structure of the representational knowledge of the model for a bit in chapter 6. In the present chapter, a revised take on the competence-performance distinction came to light. As one of the goals of construction grammar, or any cognitive theory of language, is to understand the representational knowledge or competence of a language user, it may be insightful to take a ‘look under the hood’.