Constructions emerging : a usage-based model of the acquisition of grammar
Beekhuizen, B.F.

Author: Beekhuizen, Barend
Title: Constructions emerging : a usage-based model of the acquisition of grammar
Issue Date: 2015-09-22

CHAPTER 6

Entering the black box

In chapter 5, we looked at the behavior of the model in understanding the input items that it processes. At several points, I referred to the idea that SPL's potential for analyzing utterances may go beyond the behavior that it shows in comprehending input items. Unlike with human subjects, a computational model such as SPL allows us to 'take a look under the hood', and find out what the inner workings of the model are. In this chapter, I explore two of these. these. First, it is interesting to inspect the frequency with which the learning operations are applied. Despite their availability thoughout ontogenetic development (cf. desideratum D6-4), their actual use may vary. What does this tell us about the actual use of the model's processing competence? Second, we look at the representations learned by the model. Recall from chapter 5 that it may be that the model uses only a limited subset of all representations it has acquired. In that chapter, I suggested that this be taken as the usage-based instantiation of the (representational) competence-performance distinction. In this chapter, we look at the representational competence of the model.

6.1 Learning mechanisms

We can inspect how frequently the various learning mechanisms are applied by the model. A first reason to do so, is that it provides us with further insight in the way the model works. Can the application of learning mechanisms for instance be linked to the law of cumulative complexity? Furthermore, any patterns we detect in the application of the learning mechanisms can inspire novel hypotheses about the course of language acquisition in the child.

6.1.1 Lexical learning

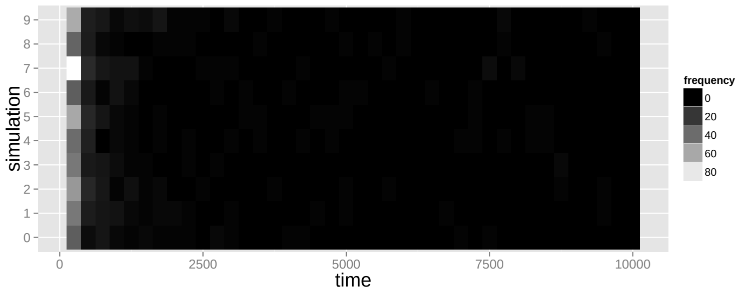
The hypothesis that the available mechanisms vary in their importance has been framed most clearly by Lila Gleitman in various publications (Gleitman 1990, Gleitman et al. 2005). Although cast within a nativist framework, the idea can be easily transferred to a usage-based one. In Gleitman's account, simple associative learning is a capacity available at any time in ontogeny, but its use may be restricted to early development. Afterwards, after all, the learner has acquired several grammatical representations that it may use in a top-down way to analyze a substring of the utterance for which it does not have a lexical representation yet. Gleitman calls this 'syntactic bootstrapping', and the process is instantiated in SPL as the bootstrapping operator of rule **vi**, whereby any phonological string can be fit into an non-phonologically specified slot of a construction. If the analysis involving the application of bootstrapping turns out to be the best one, a lexical construction containing the bootstrapped phonological string is added to the grammar.

When we look at the relative importance of the various operations involved in the acquisition and reinforcement of lexical constructions (figure 6.1), we can see a very similar picture to Gleitman's emerging. Light-colored cells depict a high amount of applications of the learning mechanism, and dark-colored cells a low amount. I counted an application of cross-situational learning, bootstrapping and adding a most-concrete construction only if the representation with which the grammar was updated was not already in the grammar. In other words: I counted the first three mechanisms only if they gave rise to a novel representation.

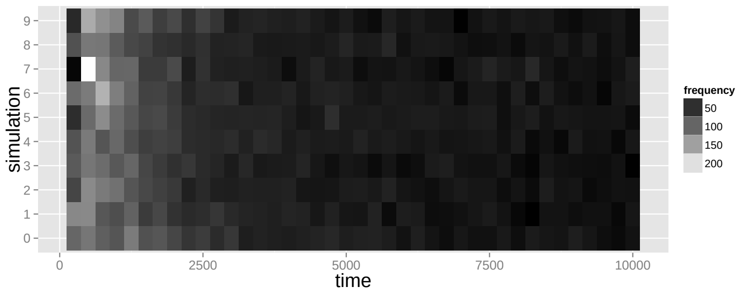
Simple, associative cross-situational learning is used only in the very early stages, up until about 250 input items, after which it completely falls out of use. After having processed very few input items, the model seems to have built up a repertoire of grammatical constructions allowing it to bootstrap novel lexical constructions. This mechanism remains being used by the model to obtain novel lexical constructions throughout development, although less frequently (recall that the model has seen almost all word types after some 1500 input items). This means that over the whole of development, most lexical constructions are obtained by bootstrapping them on the basis of the linguistic knowledge applied to the rest of the utterance rather than by a form of cross-situational learning.

The mechanism whereby the model adds a new representation on the basis of the most-concrete construction given an existing lexical item rarely occurs. This does not come unexpected: most words have a fixed set of semantic features, and hence abstractions over words are typically not very useful to the model. Hence, these abstractions are few, and so are any novel most-concrete construction **mcc**s learned on the basis of analyses involving these abstractions.

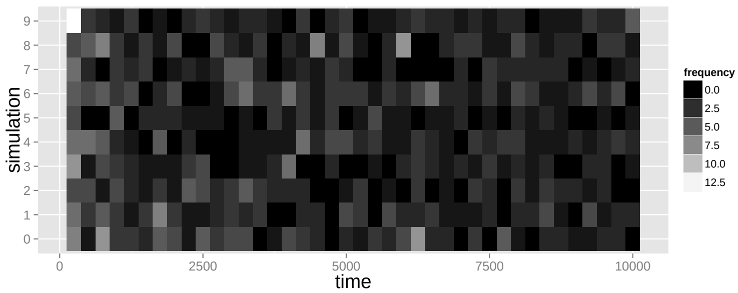
Of course, one caveat here is that I only implemented one form of cross-situational learning. Nonetheless, I believe this result provides us with an in-



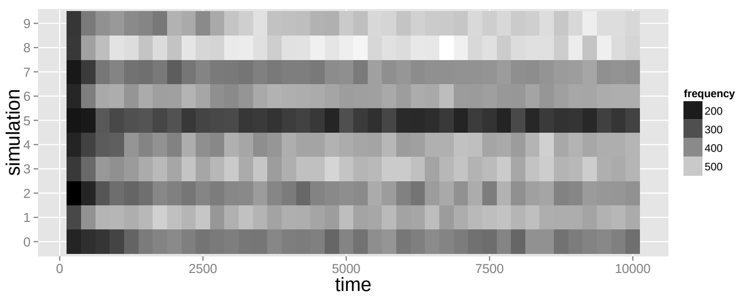
(a) Cross-situational learning.



(b) Bootstrapping.



(c) Update of a lexical most-concrete construction.



(d) Reinforcement of a lexical most-concrete used construction.

Figure 6.1: Frequency of learning mechanisms involved in the acquisition of lexical constructions over the first 1000 input items.

teresting line of further study, namely the exploration of the ways in which lexical constructions, or words and their meanings are acquired and the question which sources of information are used *over developmental time*. The results from SPL, in line with Gleitman's idea, suggest that a combination of knowledge of the rest of the linguistic structure with some form of top-down processing, may be dominant in later development, whereas associative learning may prevail earlier on.

An interesting pattern, finally, that we can glean from these graphs, is that in simulation 5, relatively few reinforcements of the most-concrete used construction are made. As we will see, this is because the model reinforces most-concrete used *grammatical* constructions instead. I postpone the analysis of this observation to the next section.

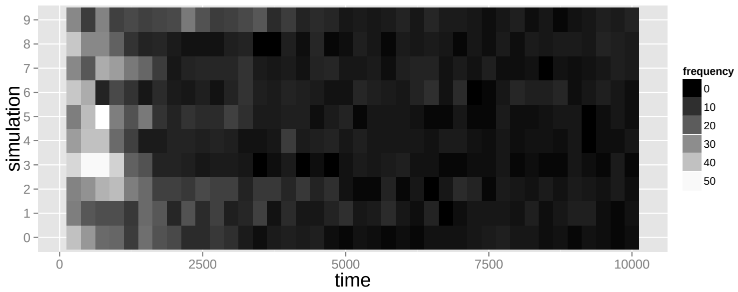
6.1.2 Grammatical learning

As for the acquisition of lexical constructions, we find variation in the frequency of use of various learning mechanisms for grammatical constructions over time (figure 6.2). Syntagmatization is mainly found in early development, after which SPL starts abstracting over the obtained grammatical representations. Later syntagmatization operations likely involve the extension of three-argument to four-argument patterns, and we will look at this more closely in the latter two sections of this chapter.

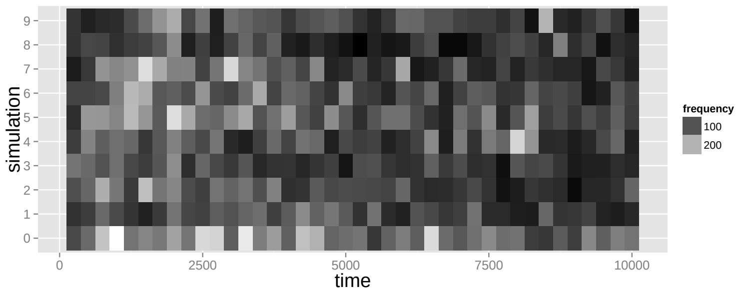
Learning from most-concrete constructions is also a learning mechanism that takes place mostly early in development, but its use over time decays slower than that of syntagmatization. Recall that with the addition of a most-concrete construction **mcc**, the model creates a trace of the processed exemplar. As novel input items (i.e., input items that – as a whole – have not been seen before) will be presented to the model throughout development, adding a trace of the analysis of that novel input item is something the model will keep doing. Of course, the number of novel utterances will decay over time, and because of that, the amount of **mcc**s.

An interesting finding for abstraction is that, unlike the other mechanisms, its application is not smoothly distributed over time. Syntagmatization and the acquisition of novel representations by most-concrete constructions are frequent early on, and gradually decay over time. Abstraction, however, seems to take place in bursts. What happens here, is that when SPL encounters an analysis with a novel grammatical construction, for instance through adding an **mcc**, this pattern may trigger a number of abstractions, with various other constructions. These bursts are suggestive of a developmental pattern Kwiatkowski (2011) models, namely, the non-gradual development of the learner's production. Similar bursts in the model's potential will be seen in chapter 7, where we discuss how the model generates utterances on the basis of a situation.

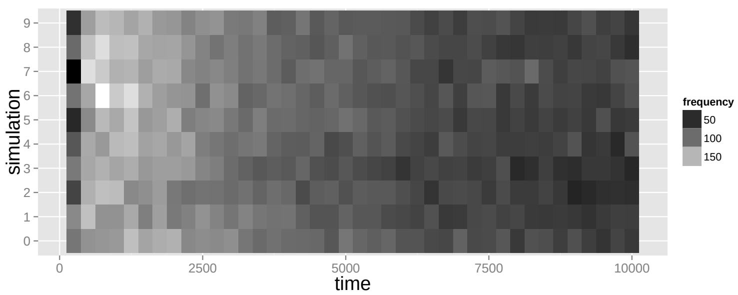
Reinforcement of the most-concrete used constructions (the **mcuc**s) is something that takes place continuously. Recall that we observed that for sim-



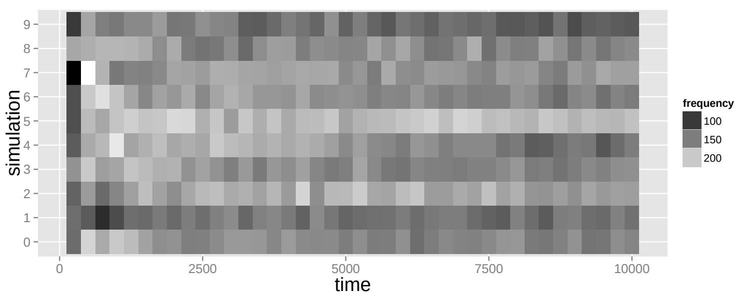
(a) Syntagmatization.



(b) Abstraction.



(c) Update of a grammatical most-concrete construction.



(d) Reinforcement of a grammatical most-concrete used construction.

Figure 6.2: Frequency of learning mechanisms involved in the acquisition of lexical constructions over time.

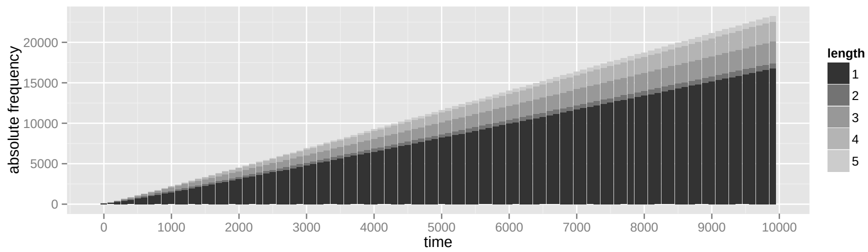
ulation 5, SPL performed fewer updates of lexical **mcuc**s than for the other simulations. Interestingly, we find the reverse for the grammatical **mcuc**s, namely that there are more reinforcements of grammatical **mcuc**s in simulation 5 than for the other simulations. What happens in simulation 5, is that the model relies more on lexically specific grammatical constructions than in the other simulations. This is merely an effect of the order of the first hundreds of input items, but it raises the interesting possibility that the order and temporal distribution of the input items may affect the kinds of representational categories that are used and reinforced, thus allowing for individual variation despite the same mechanisms and sensitivities (or parameters) of the mechanisms. Crucially, in all simulations adequate behavioral performance is achieved: the model is able to identify the target situation, analyze the full utterance and understand to what parts of the identified situation the elements of the utterance refer. This finding supports the recent insight that it may be the case that, despite behavioral near-identity in everyday behavior, language users' internal grammars may vary (e.g., Dąbrowska 2012). However, they do so through a different route: whereas in the case of Dąbrowska's results, the differences between individuals are likely a product of differences in the quantity and quality of experience, in the case of this modeling experiment, the quantity and (to a large extent) the quality of the linguistic experience are the same between simulations. This raises the interesting suggestion that the order of input items may affect the representations learned by a language user.

6.2 The representational potential

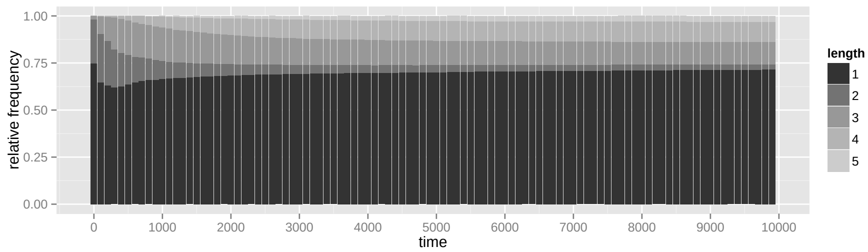
In section 5.3.4 we looked at how often constructions of various length and abstraction are used by SPL in comprehending utterances. At that point, I remarked that there may be a difference between the constructions used by the model and the potential the model has. The internal state of the model can be compared with the behavior of the model (in comprehension, for instance). This way, we can arrive at an understanding how distant the model's constructional potential is from the behavior it produces. Such insights are important, given that in many usage-based corpus studies a strong what-you-see-is-what-you-get perspective is taken, assuming that the behavior as given in a corpus does not provide evidence for a more abstract representational system, but it may be that the typically highly limited behavior of children is produced by a richer (i.e., more abstract) representational system in which, for instance, the abstract patterns never surface in behavior because they are always preempted by more concrete, slightly worse-fitting but better-entrenched ones.

6.2.1 Length of the acquired constructions

Before we look at specific cases, let us inspect some general properties of the model. Figure 6.3a illustrates how the constructional knowledge is monoto-



(a) Unsmoothed absolute frequency of constructions of various length over time.



(b) Unsmoothed relative frequency of constructions of various length over time.

Figure 6.3: Unsmoothed frequency of constructions of various length over time.

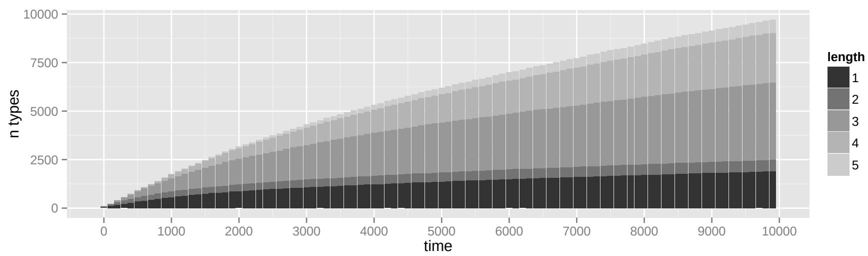


Figure 6.4: Number of unique construction types of various length over time.

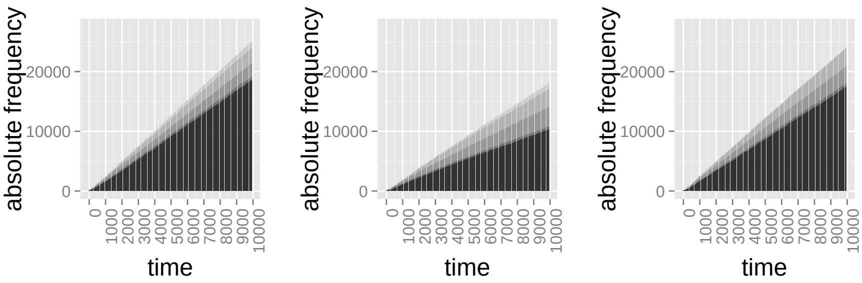
nously increasing. The height of the bars reflect the total amount of reinforcements the constructions of various length have received as **mcuc**s. As the frequency is unsmoothed, constructions with a count of zero (i.e., those that have been acquired through bootstrapping, cross-situational learning, syntagmatization, paradigmatization, or as an **mcc**, but that have never been reinforced), do not count towards the global frequency.

The figure that depicts the same data, but then as proportions of the total grammatical knowledge (figure 6.3b) shows a slightly different picture. In it, we can see that in the early stages, most of the counts are divided over lexical constructions and length-2 grammatical constructions. One by one, length-3, length-4 and length-5 constructions enter the constructicon and become reinforced.

Many constructions may be present as representations without ever having received any reinforcements, and as such figures 6.3a and 6.3b give a slightly distorted image. After all, in actual use, the counts of these constructions are smoothed, so that their probability is non-zero. An alternative way of conceiving of the absolute and relative strength of the various representations is by looking at the number of unique construction types at each time. Figure 6.4 gives this information.

One striking aspect of the number of types, when compared to the absolute or relative frequencies, is that there are many constructions of length 3 and greater that have not been reinforced. This is an effect of the blind application of the paradigmatization operation, where any and all abstractions are added to the grammar. It also points to the clear way in which SPL instantiates the idea of immanence: any overlap between any two patterns is part of the model's potential for analyzing novel utterances.

Looking at the variation between simulations, next, we can first observe that there is a difference in the absolute number of reinforcements divided over the grammar. Whereas in simulation 3, the total number of reinforcements after $10,000$ input items is around $23,000$, the number of reinforcements



(a) Simulation 3 (b) Simulation 5. (c) Simulation 6.

Figure 6.5: Absolute frequencies of constructions over various lengths over time for three simulations. Legend is the same as in figure 6.3.

in simulation 5 lies around 18, 000. Interestingly, simulation 5 also performs slightly worse on the identification of the target situation as well as the situation coverage (cf. figures 5.1 and 5.3). As we will see in the next chapter, the model also behaves slightly differently in simulation 5 than in the other simulations. Nonetheless, even in simulation 5, SPL is a relatively successful communicative agent, correctly identifying over 70% of the target situations.

Furthermore, an interesting pattern in the comparison between the simulations surfaces. Whereas simulations 3 and 5 (and all others) have constructions of length-5, in simulation 6, reinforced constructions of that length are not in the representational system most of the time, with a few emerging only at the end. The model is in this simulation nonetheless as successful as in the other simulations. What happens in simulation 5, is that various length-4 constructions of the type given in example (37) are acquired. These constructions become reinforced both by sentences of the type *you put ball in basket* as well as cases of *you put ball there*, where the model analyzes *there* as referring to the LOCATION. At some point in simulation 6, constructions of the type in (37) have been reinforced to such an extent that the final word may even be known (e.g., [SPEAKER / *me*]), but this word cannot be concatenated with the construction, as it refers to the LOCATION as well. Combining them with concatenation would constitute a violation of the isomorphy principle, and is therefore excluded. Alternative analyses (e.g., using a length-3 construction and concatenating that with the well-known word) turn out less likely than the ones on the basis of the types of constructions exemplified in (37).

(37) [[ENTITY] [PUT / *put*] [ENTITY] [LOCATION]]

(38) [[ENTITY] [PUT / *put*] [ENTITY] [LOCATION-ROLE] [LOCATION]]

At around 9000 input items, SPL has started to acquire the caused-motion construction as exemplified in (38). Upon encountering further instances of sentences like *you put ball in basket*, the model is now able to parse them with a length-5 construction, and it is likely that this construction will continue being reinforced over time. All in all, nothing is lost, but the model is simply a late learner with respect to the length-5 constructions.

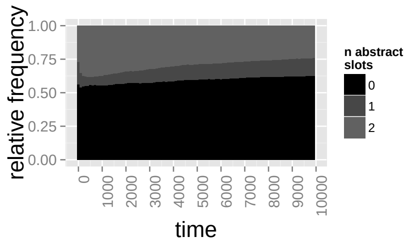
However, not all utterances that *can* be covered with length-5 constructions *are* covered with such constructions. The model has mistakenly taken up the prepositional dative construction (the construction behind sentences such as *he gave the book to Mary*) as a length-4 construction:

(39) [[ANIMATE$_i$] [GIVE / *give*] [OBJECT] [ANIMATE$_j$ / *to*]] | GIVE(GIVER(ANIMATE$_i$),GIVEN-OBJECT(OBJECT), RECIPIENT(ANIMATE$_j$))

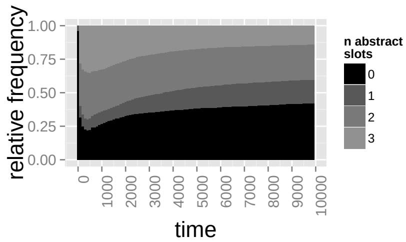
This construction involves (correctly) an animate entity in the giver-role, the verb, and a given object. It has mistakenly learned *to* to refer to the recipient entity, but only in the context of this constructions: SPL is able to analyze sentences such as *you go to school* or *you take ball to table* with a construction that involves *to* as a marker of direction. As with the earlier erroneous caused-motion construction in (37), the fact that *to* refers to the entity filling the recipient role blocks the pattern from being concatenated with a noun or pronoun following it, even if that noun or pronoun is well known.

To ascertain that this is not an effect that can be overcome with more data, I let simulation 6 continue processing input items after it was done. Even after $20,000$ input items, the model still analyzes prepositional datives with constructions such as (39). We can take this to mean that the model got stuck in a local optimum. This means that it has acquired a construction (i.e., the one in example (39)), that allows it to identify the situation correctly in most cases, but that does not cover all of the utterance and the situation. Of course, real language-learning children would never find themselves 'stuck' in such a situation: the functional relatedness of *to* in the prepositional dative to that in several motion constructions (underlying such utterances as *you go to school* and *you take ball to table*), and the fact that the application of the construction of (39) always leaves one word of the utterance unanalyzed, even if that word may be well known, should, at some point, convince the learner that the construction in (39) is not a conventional pattern of the language.

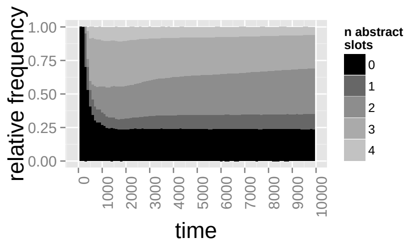
This points to a point of weakness of the model: it is not able to overcome these local optima. This constitutes a kind of brittleness that we would like a model to be able to overcome. To my mind, a crucial change in the model might be to make the 'penalty' for ignoring words proportional to how well these words are known. If the learner encounters *you give it to me*, and knows that *me* refers to the speaker, it should penalize analyses in which *me* is taken to be noise more severely than analyses in which *to* is taken to be noise (as that word likely has little reinforcements outside of the constructions in which it constitutes a fixed element).



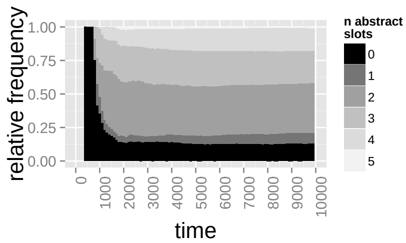
(a) Abstraction among length-2 constructions.



(b) Abstraction among length-3 constructions.



(c) Abstraction among length-5 constructions.



(d) Abstraction among length-5 constructions.

Figure 6.6: Relative frequencies of the various degrees of abstraction, per length, over time.

6.2.2 Abstraction in the representational potential

In section 5.3.4, I discussed the use of constructions of various length and degrees of abstraction over time. Being the constructions that are used in finding the best analysis, these constructions are also the ones that are reinforced over time. We can interpret the effects on the abstraction of the constructions of various lengths by looking at how much reinforcement each of these levels of abstraction has accrued over time. Figure 6.6 shows the normalized frequencies of each level of abstraction, per length, over time.

What the various figures show, is that the potential for generalization is quickly obtained by the model (somewhat later in the length-4 and length-5 constructions than the length-2 and length-3 constructions). After having found this potential, more and more more concrete patterns are learned that take up increasingly much of the relative frequency. That is: the potential for having a fitting representation for each situation becomes greater over time. Note that, unlike SPL's use of unanalyzed lexical chunks, it's increasing use of analyzed but phonologically specific constructions is in line with the findings reported by McCauley & Christiansen (2014*a*).

What are the semi-abstract longer constructions that are well-entrenched? If we look at simulation 5, and inspect the most-frequently used length-5 constructions, the following five constructions constitute the top-5:

(40) [[SPEAKER / *you*] [PUT / *put*] [PLURAL-PERSON / *them*] [CONTAINMENT-ROLE / *in*] [ENTITY]] ($count = 94$)

(41) [[PERSON$_i$] [GIVE / *give*] [THING / *it*] [GIVER-ROLE / *to*] [PERSON$_j$]] ($count = 93$)

(42) [[PERSON] [GIVE / *give*] [THING / *it*] [GIVER-ROLE / *to*] [WOMAN]] ($count = 80$)

(43) [[PERSON$_i$] [GIVE / *give*] [THING] [GIVER-ROLE / *to*] [PERSON$_j$]] ($count = 53$)

(44) [[SPEAKER / *you*] [PUT / *put*] [THING] [CONTAINMENT-ROLE / *in*] [ENTITY]] ($count = 30$)

We see both the caused-motion pattern and the prepositional dative in various degrees of abstraction among the five most-frequently used constructions. These semi-open constructions function as composite multi-word units in comprehension: multi-word units because they capture frequently occurring lexical patterns, composite because each of the parts of the construction specifies a certain role in the more global meaning. As such, these patterns are distinct from true 'chunks', that are internally not analyzed.

In all of the five most-frequently used length-5 constructions, the verb is fixed. In fact, in none of the length-5 constructions in this simulation, a pattern in which a generalization over caused-motion constructions and prepositional dative construction is made. This is a direct effect of the fact that the model has erroneously acquired *to* in the prepositional dative to refer to the GIVER, or AGENT, role. Because of this, the model cannot form an abstraction over the meaning representations of the two constructions.

As we can glean from figure 5.9h in chapter 5, there are some simulations in which the abstraction over caused-motion constructions and prepositional datives is made, judging by the small, but non-zero amount of length-5 constructions with 5 abstract slots. In simulation 2, for instance, the model has acquired a construction, given in (45), that only has a fixed subject, but no other lexically specified roles. The reason this abstraction could be made, is that in simulation 2 the model did correctly acquire the meaning of *to* in the prepositional datives as referring to the RECIPIENT role (unlike in simulation 7, where it is analyzed as denoting the PATIENT or GIVEN-THING role, and simulation 5, where it is analyzed as marking the RECIPIENT referent).

(45) [[HEARER / *you*] [CAUSE] [OBJECT] [ROLE] [ENTITY]] | CAUSE(CAUSER(HEARER),AFFECTED(OBJECT),ROLE(ENTITY))

This construction, however, is used only between 1500 and 4900 input items, and only to analyze prepositional datives. What happens here, is that the model extracts the construction in (45), and finds it to be part of the most likely analyses of prepositional datives with *give* as the verb. These analyses then are added to the grammar as maximally-concrete constructions (**mccs**), and after 4900 input items, SPL has acquired a range of these more concrete patterns to the extent that the abstraction in (45) is no longer needed.

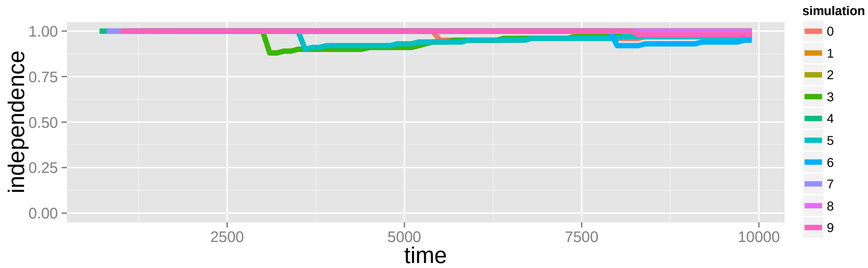
Returning to *to*, it seems that the various simulations differ in how they analyze *to* in the prepositional dative. Out of ten, only four assign the correct RECIPIENT role to the word, whereas in five cases the RECIPIENT referent is taken to be the meaning of *to*, and in one case, as we have seen, the PATIENT role. Of course, more than 40% of children acquiring English get the meaning of *to* correct (although it may be a preposition for which semantic errors could be expected).

Several aspects prevent the model from being like a child for this phenomenon. First, the input is more scarce in types of verbs and prepositions than a child receives. If various verbs and prepositions are heard, the chance of acquiring the right meaning of *to*, which contrasts with other prepositions in that position, becomes greater. Suppose various verbs are heard in length-5 constructions. An abstraction of the type (45) is then quickly acquired. Even if the meaning of *to* is erroneously acquired, a construction with an open verb slot, like the one in (45), may 'overrule' the more specific, but erroneous pattern with *give* and *to* lexically specified. As we have seen in chapter 5 that more abstract constructions may 'overrule' (which we can take to be the antonym of 'pre-empt') the use of more concrete ones if the abstract ones are well-entrenched and the lexical units filling the slots are well-entrenched as well.

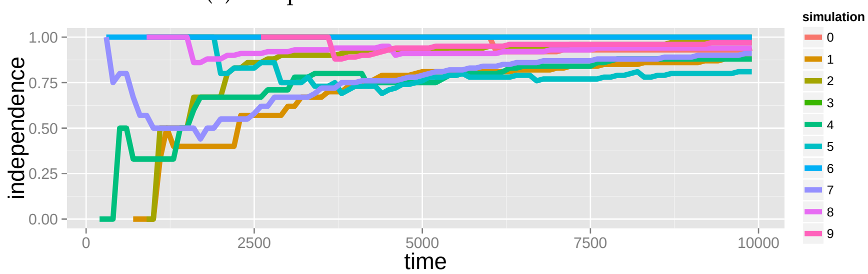
6.3 The independence of morphemes

We saw that *to* was acquired, with the correct or incorrect meaning, as part of a larger construction. In none of the simulations, a lexical construction of the type [RECIPIENT-ROLE / *to*] is well-entrenched. Of course, the representation is there, but it never gets reinforced, because it is always the larger construction in which *to* is used that is reinforced. This raises an interesting issue, namely whether the model can tell us something about the independence of the smallest units. This is an issue that touches on both (what are traditionally called) morphology and syntax: a unit is bound if it can be only used in combination with other units, whereas it is free if it can be used independently. Can the model give us any insight in the degree of independence of the various words?

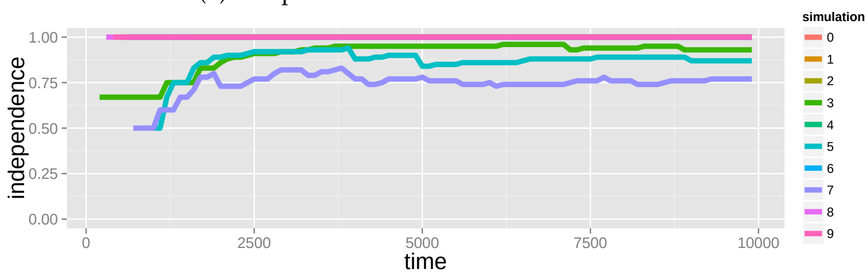
The issue of independence finds its theoretical relevance for the usage-based approach in the programmatic article on language acquisition by Langacker (2009), who argues that the independence of a unit (construction) depends on the variety of contexts it occurs in, in interplay with the frequency of the unit itself, both inside and outside of particular constructional contexts. Certain words, such as determiners, may never obtain strong independent



(a) Independence of the word *cereal* over time.



(b) Independence of the word *animal* over time.



(c) Independence of the word *aunt* over time.

Figure 6.7: Independence of various entity words over time.

status, whereas others, occurring over a variety of contexts, do get reinforced as independent units. I would like to add one aspect to Langacker's conceptual analysis, namely that, besides the token frequency and the dispersion of a word over various constructions, also the type frequency of the constructional slot (i.e., the amount of types of units filling it), plays a role in establishing the degree of independence of a unit. This last point is taken from Bybee's (2006) analysis of constructional productivity. In this section, I will show how all these effects can be seen in the model when we look at the strength of the representation of lexical constructions as opposed to grammatical constructions containing those lexical constructions.

In the following paragraphs, we look at five groups of words, corresponding roughly to nouns, adjectives, pronouns, verbs, and prepositions/spatial adverbs. We can expect the degree of independence to vary between them, as they have different quantitative values for the three properties mentioned above.

As a simple measure to operationalize the independence of a word form w, I take the relative frequency of lexical constructions out of all constructions in which a word form w is lexically specified (cf. equation 6.1, where Γ_w is defined as the subset of the constructicon Γ consisting of all constructions in which w occurs as the phonological specification of a constituent). This tells us how often the word form w is analyzed with a lexical construction. The more frequently this happens, the more we can claim that w and its meaning are free units. We call this value the **independence** score, ranging between 0 and 1.

$$\mathbf{independence}(w) = \frac{\sum\limits_{c \in \Gamma_w \wedge c = \mathbf{lexical}} c.count}{\sum\limits_{c' \in \Gamma_w} c'.count} \tag{6.1}$$

6.3.1 Entity words

Words referring to entities, typically called 'nouns', can be expected to be among the most independent words. After all, they occur as the arguments of multiple action words ('verbs') in the input generation procedure, and many other entity words fit these slots as well, making it likely that the optimal analysis involves the lexical construction involving the entity word and a grammatical construction with an open slot where the entity word is fit in. Figure 6.7 shows, for three entity words, that this is indeed the case. After $10,000$ input items, constructions involving the phonological strings *cereal*, *animal*, and *aunt* are mostly lexical.

When focusing on the developmental path, we see *cereal* being used mainly in lexical constructions from the onset of the simulations, whereas *aunt*, and especially *animal* start out as often being part of a grammatical construction early on, and gradually being used more as an independent word, and hence receiving more reinforcement as a lexical construction. The string *animal* occurs in all but a few cases as the theme argument of a caused-motion construction (in utterances such as *you put animal on table*. Because of the restricted variability, the model does not have to use the lexical construction [ANIMAL / *animal*] in any other context, and in the context of caused-motion sentences, the model has a semi-open construction of the type in example (46). Whereas the semi-open constructions in example (46) will receive reinforcement over time, the lexical construction will not. This pattern of pre-emption, however, is gradually overturned, as constructions such as (47) also receive much reinforcement. In this construction, the theme argument is open, and because many different theme arguments are encountered, this kind of construction receives much reinforcement. Over time, the best analysis is increasingly likely to involve the construction with an open theme-argument slot (example (47)) and the independent lexical construction [ANIMAL / *animal*], and more reinforcement is given to the lexical unit. The same happens, to a weaker degree, for *aunt*.

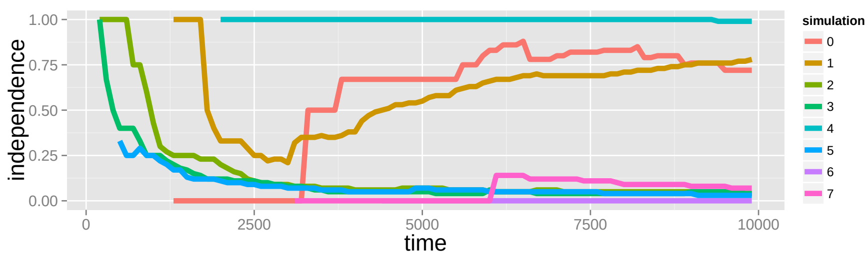
(46) [[HEARER / *you*] [PUT / *put*] [ANIMAL / *animal*] [SURFACE-ROLE / *on*] [ENTITY]]

(47) [[HEARER / *you*] [PUT / *put*] [OBJECT] [SURFACE-ROLE / *on*] [ENTITY]]

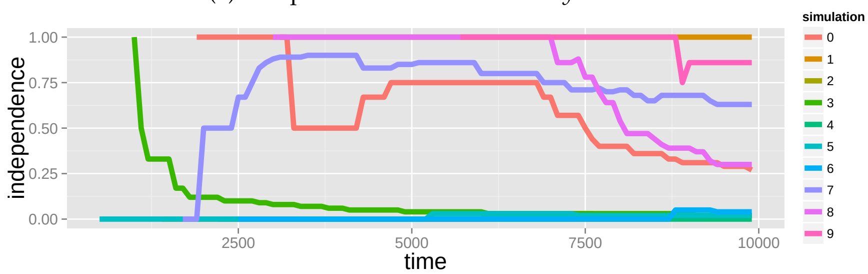
6.3.2 Attribute words

Unlike the entity words, the attribute words ('adjectives') are not used in many different constructions and the verbs that have them as arguments have a fairly restricted set of attribute words in the input generation procedure. Especially in the case of the construction [[ENTITY] [BECOME / *get*] [ATTRIBUTE]], the model moreover often acquires chunks consisting of *get* and the attribute word. Whenever attribute words are acquired, they vary in whether they are learned as a lexical construction or as part of a grammatical construction. For all three words we find the tendency that they become increasingly associated with a construction in which they are lexically specific (decreasing values on the y-axis). However, in some simulations (e.g., simulation 1 for the word *dirty*), the word starts out being used most often in lexical constructions, after which it is used as an element of a grammatical construction, and finally it is dissociated from that construction again. This effect is caused by the interaction of the fact that *dirty* is only used in the [[ENTITY] [BECOME / *get*] [ATTRIBUTE]] construction, but that the ATTRIBUTE slot of that construction is extended with new types over time, leading to increased reinforcement, and hence a greater likelihood of combining the lexical [DIRTY / *dirty*] construction with the construction in which the ATTRIBUTE slot is phonologically open.

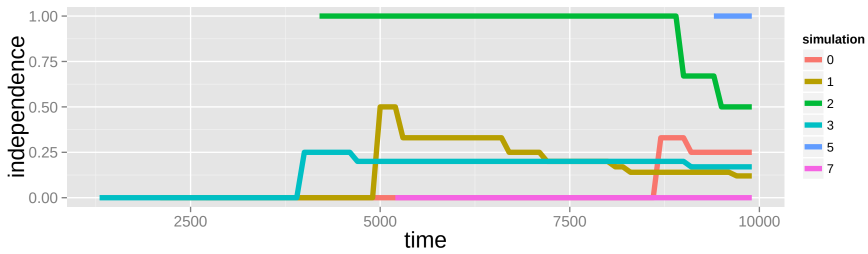
The fact that these attribute words gravitate towards being used in grammatical constructions may be partially due to the fact that there are no copula



(a) Independence of the word *dirty* over time.



(b) Independence of the word *closer* over time.



(c) Independence of the word *pretty* over time.

Figure 6.8: Independence of various attribute words over time.

constructions in the input generation procedure. If there were, the attribute words would also be used in those cases, and their reinforcement as lexical constructions would be greater.

6.3.3 Pronouns

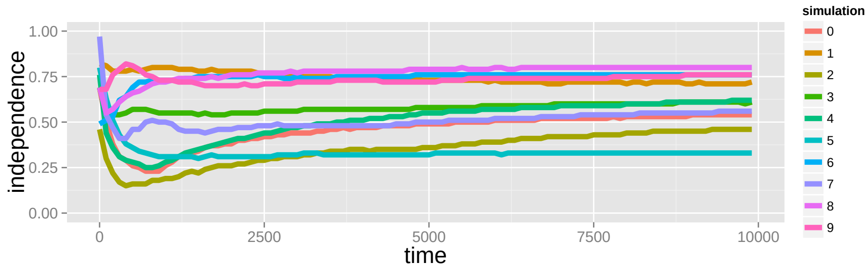
Pronouns constitute an interesting case for the test of independence. Because of their high frequency, they are expected, on a usage-based account, to be part of argument-frame constructions. On the other hand, their varying distribution (especially in a language like English where the pronouns only express two grammatical cases) makes the reinforcement of their independent forms to be expected. Figure 6.9 shows the independence for three pronouns, *you*, *I*, and *we*. As we can see, their degree of independence varies dramatically among them and between simulations. *You* is acquired in all cases both as part of a lexical construction and as part of a grammatical construction (i.e., the learner has both a [HEARER / *you*] construction, and various grammatical constructions in which *you* is used, and, crucially, reinforces all of them regularly (otherwise the relatively stable, horizontal lines of figure 6.9a would not be maintained). The variation ranges between independence scores per simulation of 0.3 and 0.8.

For *I*, the picture is different. Here, we see that there is a significant amount of between-simulation variation, but the stable state of the model in various simulations seems to be more 'polar': either *I* is most strongly represented as an independent construction, or the grammatical constructions in which [SPEAKER / *I*] is a constituent are the primary locus of the knowledge about *I*.

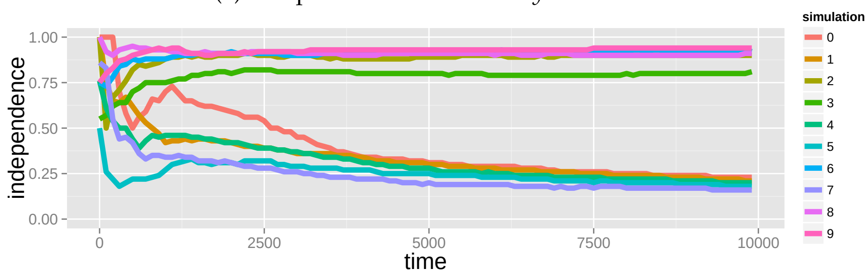
We, finally, is primarily acquired as the phonological constraint on an independent lexical construction. Unlike for the entity and attribute words, it is not easy to find an explanation for this high amount of variation: all three words are used in various constructional slots, and these slots are typically highly productive (i.e., many other items can fit in them). This difficulty of explanation, however, does point to the insight that the degree of independence of a word may be an effect of many interacting factors.

6.3.4 Event words

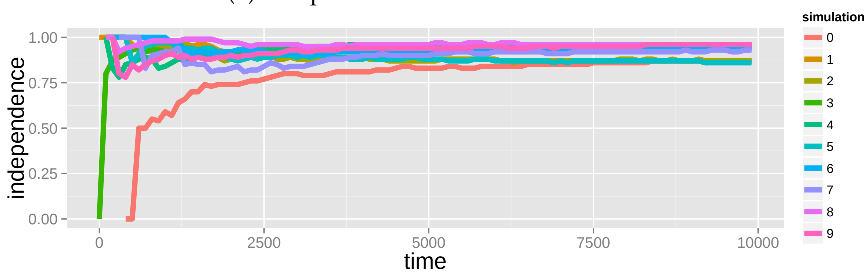
As with the pronouns, the picture of the **independence** of the event words ('verbs') is rather diverse (figure 6.10). The word *eat* is most strongly represented as an independent construction in most of the simulations. This does not come as a surprise if we bring to mind that the non-lexically-specific transitive construction (i.e., the transitive construction with an open EVENT slot) is strongly reinforced. *Put*, on the other hand, is only processed in the context of the caused-motion construction, and this construction allows for no other verbs in it in the input generation procedure. Furthermore, the abstraction over the various caused-motion constructions and the prepositional datives is



(a) Independence of the word *you* over time.

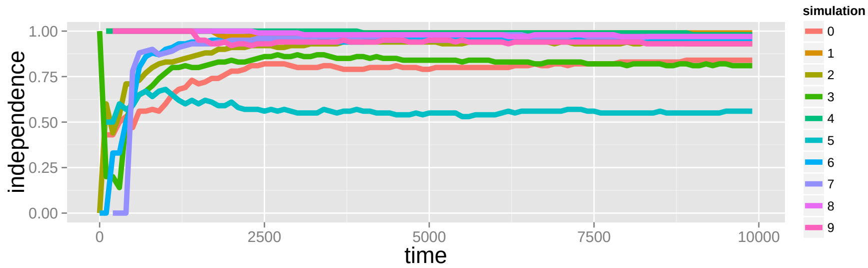


(b) Independence of the word *I* over time.

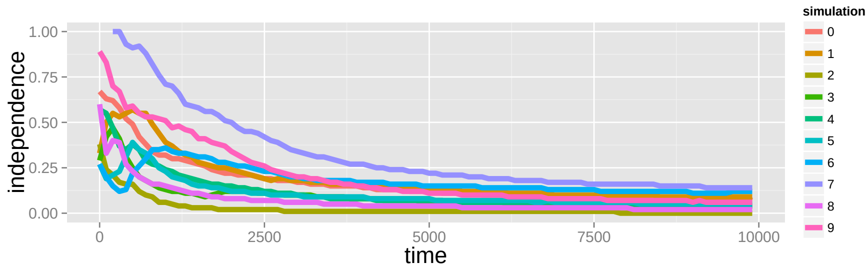


(c) Independence of the word *we* over time.

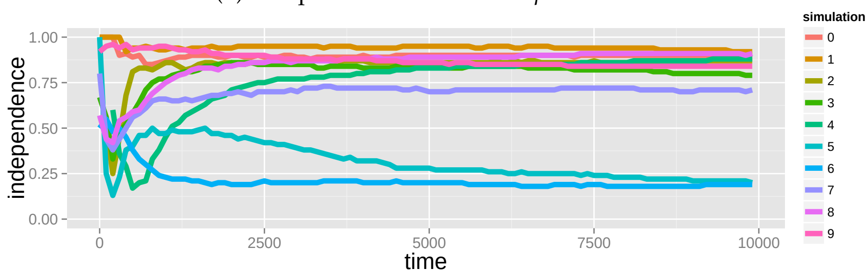
Figure 6.9: Independence of various pronouns over time.



(a) Independence of the word *eat* over time.



(b) Independence of the word *put* over time.



(c) Independence of the word *make* over time.

Figure 6.10: Independence of various event words over time.

rarely made, so that there is neither a non-verb-specific length-5 construction available. *Make*, finally, and like *you*, varies between simulations. In some simulations, the string is primarily used as the sole phonological constraint on a lexically specific construction, whereas in others, *make* is the phonological constraint of the [MAKE / *make*] slot of a larger, grammatical, construction.

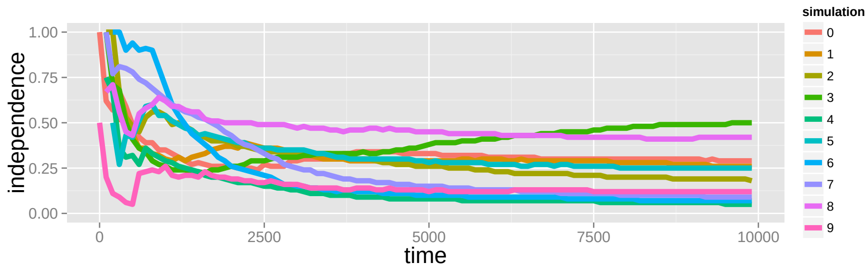
A curious phenomenon for the verbs (and to some extent for the pronouns as well) is that in some simulations, the curve displays a dip in **independence**, after which the value goes up again. The effect that causes this is again the interplay of the productivity of the EVENT slot of various grammatical constructions and the variety in grammatical constructions the verb can occur in. In some simulations, it seems to be the case that the event word starts out as an independent word (it is bootstrapped, or learned by means of cross-situational learning), after which the semi-open constructions in which it is specified amass reinforcement. As the amount of variation in the input data grows, the abstractions over the various semi-open constructions begin accruing reinforcement as well, and at a certain point the most likely analyses involving these event words consist of a grammatical construction with an open EVENT slot, combined with a lexical construction containing the event word. This moment is at the bottom of the dip: afterwards, the **independence** score starts rising again, because the lexical construction gets reinforced, but the grammatical constructions with lexically specified EVENT slots do not.

One could tentatively associate this effect with the idea that children are conservative in the generalization of early verbs (McClure et al. 2006). This very finding has been questioned (Naigles et al. 2009), but it may be that there is a lot of variation between learners, between verbs, and that the periods in which the learner behaves conservatively, or, oppositely, too progressively, may vary as well.

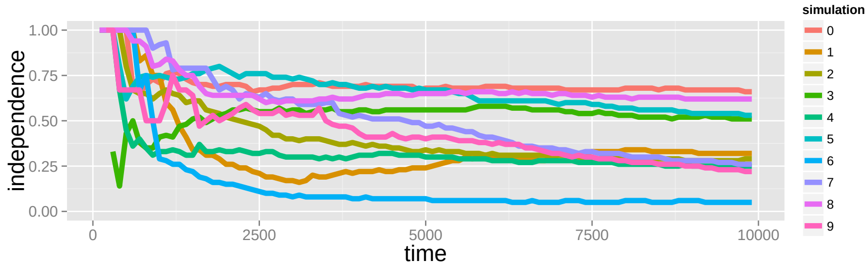
6.3.5 Role-marking words

Role-marking words, traditionally known as prepositions, are expected to be fixed elements of the grammatical constructions they occur in. However, as figure 6.11 shows, this does not seem to be (fully) the case: both *on* and *in* have a relatively strong representation as the phonological constraints on independent lexical constructions. This is due to the fact that these words do occur in multiple constructions, and contrast with other role-marking words (e.g., *to* and *out of*). Nonetheless, in most simulations, the **independence** scores are decreasing over time, meaning that more and more, the words are only used as parts of grammatical constructions.

The difference between *on* and *in* is especially striking. After all, both words occur in exactly the same constructional environments. I believe the difference is due to their varying token frequencies. We can expect, in line with Bybee (2006), that, all other things being equal, words with higher token frequencies (in particular environments) will be more entrenched in those environments. After $10,000$ input items, the average counts of all constructions



(a) Independence of the word *in* over time.



(b) Independence of the word *on* over time.

Figure 6.11: Independence of various role-marking words over time.

containing *in* and *on* is 879 and 350, respectively. This means that *in* is simply more frequent than *on* in the input generation procedure. The effect of this difference is that the more frequent word, *in*, is associated more strongly with the constructional environments it is used in, and hence that lexically specific constructions containing *in* receive more reinforcement than those containing *on*. We see here that SPL not only captures the effect of type frequency on productivity, but also the effect of token frequency on entrenchment.

6.3.6 Comparing the classes

Finally, let us take a more global look. If we group all words for which the model has any representation in at least one of the simulations according to the five-way distinction presented above, and subsequently average over all simulations and all words, per semantic class, we obtain the average **independence** values presented in figure 6.12.

The pronouns and entity words clearly have the strongest independent representations. Attribute words are mostly reinforced as part of the grammatical construction they occur in, and event and role-marking words start out relatively independent, but become more and more associated with par-

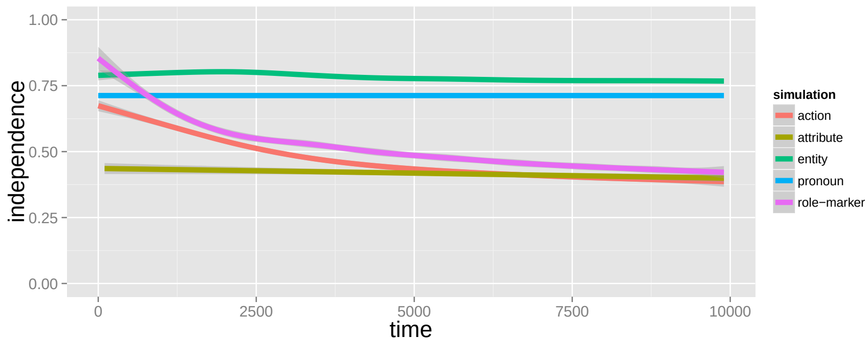


Figure 6.12: Mean independence of the five classes of words over time.

ticular grammatical constructions. This last finding is at odds with the verb-island hypothesis (Tomasello 1992), but I believe it constitutes a viable hypothesis within a usage-based framework in which the starting-small perspective is re-emphasized. Once the learner (i.c., the model) has bootstrapped or cross-situationally learned the meaning of an event word, the build-up phase consists in finding the arguments with which this word can occur. Once those have been found, abstractions are made over that event word and others occurring in similar argument-structure constructions. These abstractions only receive reinforcement if they are extended to novel event words. If that does not happen, the event words will increasingly become associated with the argument-structure constructions, leading to more lexically-specific constructions. This hypothesized developmental pathway also suggests that, in the long run, the representational knowledge of a speaker that is actually used becomes more concrete over time (cf. McCauley & Christiansen 2014*a*), whereas the potential for generalization to novel cases remains stable. Speakers get better at what they do most, without forgetting the tricks for handling novel grammatical situations.

6.3.7 Discussion

The analysis of words in various semantic classes shows us how the degree of independence, as measured by the **independence** score varies on the basis of (1) the type frequencies of the constructional slots of grammatical constructions they occur in, (2) the amount of different constructional environments they occur in, and (3) their token frequencies, much in line with Bybee's (2006) and Langacker's (2009) characterization of notion like productivity and independence. The cases discussed thus provide insight in the subtlety of the notion of productivity when applied to grammatical, as opposed to morphological constructions. Nonetheless, I believe that through careful analysis and interpretation, we can identify the factors involved in the productivity of the construction. Notably, this is not merely a study of the corpus frequencies of the words: we have to take into account that we are dealing with a learner selectively reinforcing patterns over ontogenetic time. An important difference with Langacker's account is that a short phase of independence may, in SPL, precede a higher degree of dependence. Whether this is an artefact of the model, or an actual developmental phenomenon that becomes visible once we re-evaluate aspects of the starting-small conception of language acquisition as they apply to the usage-based theory, remains to be seen. I find the latter option not inconceivable.

An important insight from the various cases is that the model, in a way, does engage in whole-to-part learning besides part-to-whole learning (D6-3), but in a quantitative way. Qualitatively, after all, the word has been established as a lexical unit. The 'dips' discussed for the event words suggest that after this establishment, the word may go through a phase of being bound to the grammatical constructions it occurs in, after which it re-establishes independence. Part-to-whole and (quantitative) whole-to-part learning thus interact in an interesting way.

A second insight from this analysis, is that SPL displays a tremendous amount of variation between the simulations. The internal representational states of the various 'speakers' differ in the independence of various words. Nonetheless, they all perform very similarly on the comprehension experiments described in the previous section, as well as, as we will see, on the production task. It seems that there is more than one representational way to Rome when grammatical behavior is concerned, a finding in line with the recent experiments of Dąbrowska (2012).

Finally, I believe this exercise supports the recent reanalysis of some old conceptions of language. Most of linguistics, even within the constructivist take on it, is committed to a perspective in which words are the atomic primitives of languages, to be combined with grammar.

The words-as-atomic-primitives perspective has led, within functional linguistics especially, to debates about the nature of word-meaning. It has long been recognized that words can have multiple related senses, a property especially true of function words, such as adpositions, auxiliary verbs and discourse particles. The discussion about word meaning mainly concerns the question whether words have a single, highly abstract meaning (monosemy), the details of which are filled in by the pragmatics, or multiple concrete and related meanings (polysemy). The Croftian perspective, in which the constructions are the primitives (but not necessarily the atoms), allows us to question the central assumption underlying this debate: the word as the locus of meaning. If we take the perspective that constructions are the non-atomic primitives of linguistic knowledge, words (as we normally conceive them as linguists) become secondary, derived realities. A word, by this token, is simply a phonological and conceptual similarity relation between the parts of various constructions. In some cases, these constructions may coincide with the word (which is what we expect for many nouns, for instance), but in others, the 'word' is the potential emanating from the use of a phonological structure and several similar functional structures across several constructions.

This perspective is much in line with suggestions of Verhagen (2006) and Boogaart (2009). Boogaart argues, for modal verbs, that there may be a third option, resolving the discussion, namely that words have certain meaning within certain constructions. Polysemy becomes, under Boogaart's analysis, a superficial effect of the same word form occurring in multiple constructions. This analysis is supported by the results of the analysis in this section: words that are strongly associated with a particular construction have weak independent representations as lexical constructions. It can be expected that modal verbs, Boogaart's case study, are strongly associated with particular constructional frames (after all, they are fairly restricted in their use across constructions, there is only a small set of them, and they have high token frequencies). If that is the case, it may well be that the lexical representation of a Dutch modal verb like *kunnen* is very weak and that the primary locus of representational strength of the word is in various constructions, each with their own meaning (e.g., deontic vs. epistemological modality).

6.4 The growth of the caused-motion construction

Besides these more quantitatively-oriented explorations of the representational potential of the model, it may also be insightful, especially for those used to doing grammatical analysis within the construction grammar framework, to see how the 'network' of constructions grows over time. Because the grammars after $10,000$ input items contain about the same number of constructions, it is not feasible to look at all of them. Therefore, we focus on a part of this network, namely where it involves events in which motion is expressed, with an external cause for that motion being presented. These are the constructions underlying such utterances as *you put it on table*. As even for this small part of the network, the number of constructions is too vast[1] to represent graphically, I focus on some interesting ones.

Figure 6.13 displays a part of the network after 100 input items. The thickness of the lines is indicative for the counts of the constructions, and constructions in grey have not been reinforced. We can see that the model has learned the action word *put*, and syntagmatized it with two entity words, *you* and *Sarah*, to form two very simple grammatical constructions. Over these constructions, furthermore, the abstraction [[PERSON] [PUT / *put*]] is made,

[1] Note that this way of framing it ('number') presents the constructions as discrete units, which they are in the implementation. As I argued earlier, we can equally well regard these as the potential for generalization the model has – a vast number of constructions in the discrete conception corresponds to a wide potential on a 'immanent perspective'.

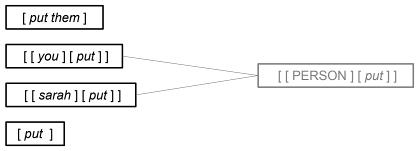


Figure 6.13: Part of the network of caused-motion constructions after 100 input items.

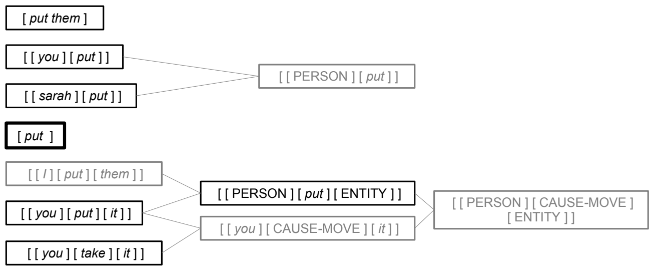


Figure 6.14: Part of the network of caused-motion constructions after 500 input items.

but this construction has not been reinforced yet. We can see that a chunk has been extracted as well, viz. [PUT(AFFECTED(THEM)) / *put them*], and this chunk has been reinforced several times.

Four hundred input items later (figure 6.14), the lexical construction [PUT / *put*] has been further reinforced. Furthermore, several length-3 constructions have been added. The various fully lexically-specific ones give rise to a small network of abstractions, even though many of the fully lexically-specific ones may not have been reinforced (not all constructions are shown here, as there are already dozens of length-3 constructions at this point). The 'old' constructions remain at the same level of reinforcement: as there is now a more useful length-3 construction, the various length-2 constructions no longer lead to optimal analyses.

Moving to the state of the constructicon after 1000 input items, we can seen that length-4 and length-5 constructions now entered the scene. For length-4 constructions, a small, but generalizable network has been built up, including a well-reinforced, highly abstract construction in which only the word *put* is specified. Note here that this abstract construction, [[PERSON] [PUT / *put*] [OBJECT] [LOCATION-ROLE]], has received more reinforcement than its daughter nodes. This is because it is the abstract construction, rather than

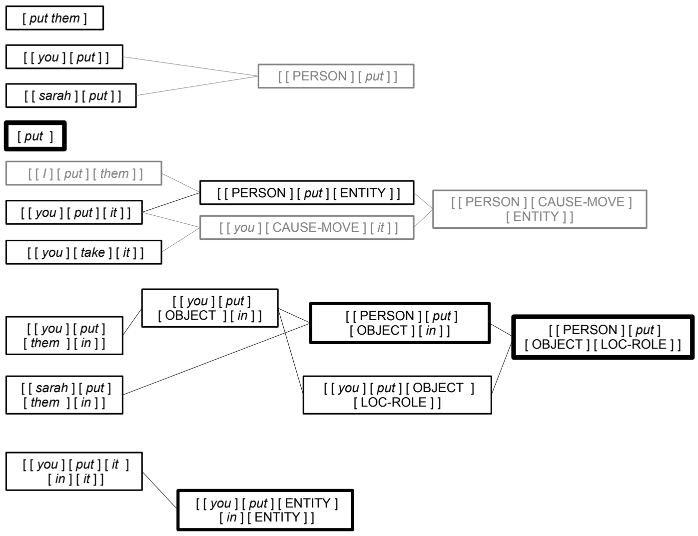


Figure 6.15: Part of the network of caused-motion constructions after 1000 input items.

its daughters that is used in processing the input items. The effect here is akin to the effect of abstract constructions obtaining unit status without the more concrete ones doing so, as described in Langacker. For the length-5 constructions, a relatively lexically-specific construction [[HEARER / *you*] [PUT / *put*] [ENTITY] [CONTAINMENT-ROLE / *in*] [ENTITY]] has been extracted, but the model has not seen any evidence for abstractions beyond this level.

After $10,000$ input items, it has seen evidence for more abstract length-5 constructions, as can be seen in figure 6.16. The network now even contains a construction in which the action word is not specified, abstracting over the constructions with *put* and those with *take* (the other verb occurring in the caused-motion construction in the input generation procedure). This maximally abstract construction has even been reinforced several times, but its more concrete daughter construction involving a phonologically specified ACTION slot (with [PUT / *put*]) has received the most reinforcement, and constitutes the prototype of this network. As we have seen in the previous chapter, it is this construction that sometimes trumps the use of more concrete constructions, because of the many different types of arguments it occurs with.

Interestingly, between 1000 and $10,000$ input items, another length-4 construction emerged as well. The pattern with a lexically specific LOCATION-ROLE, i.e., [LOCATION-ROLE(LOCATION) / *there*], is used frequently enough

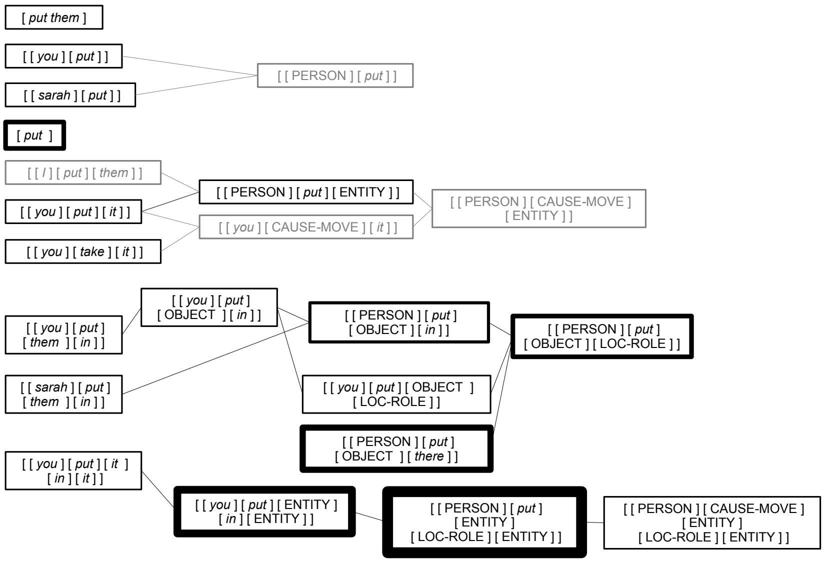


Figure 6.16: Part of the network of caused-motion constructions after 10, 000 input items.

to be strongly reinforced. A more abstract construction exists as well, but that one is less strongly reinforced. Note, finally, that all the older constructions have not received any reinforcement in the meantime. If the model involved some sort of decay function, these constructions would, by now, have withered away.

The visualization of the development of the network illustrates several important aspects of SPL. First of all, the constructions grow in length and abstraction, with each next step of length and abstraction depending on what, at that point, is available to the model. Second, we have seen how abstract units may obtain unit status, or (at least) become strongly reinforced 'prototypes' in the network. Third, the temporal, or dynamic dimension of the model becomes clear: old constructions fall out of use, while novel, and more useful, ones, take over.

6.5 Discussion

In this chapter, I looked at the learning mechanism the model employs and the representations resulting from these. I made several observations that all follow from a rigorous application of usage-based theory to the development of a model, but that may be at odds with some conceived ways of thinking.

First, we saw how the learning mechanisms are applied in section 6.1). Whereas all mechanisms are available to SPL throughout development (D6-4), the frequency of their application varies over time. Notably, the acquisition of lexical constructions is primarily done by means of bootstrapping rather than by cross-situational learning. Whereas the latter is used to get an initial set of lexical constructions, the former makes for a more reliable way of acquiring word meanings as the abstraction in the representational potential grows. In the learning mechanisms for grammatical constructions, we found that syntagmatization is applied only early on, after which the reinforcement of most-concrete used constructions and the addition of most-concrete constructions become the primary means of learning. If language is, as I suggested earlier, a set of old tools (evolutionarily speaking), used for novel purposes (i.e., language), the tools are of various use at different moments in time. A final point of interest is that paradigmatization, the process whereby novel, more abstract constructions are acquired, takes place in bursts. This observation may bring the usage-based conception in harmony with the finding that not all development is gradual.

Next, I discussed the length and abstraction of the acquired representation (section 6.2). I found that the length of the constructions in the representational potential of the model grows over time, in line with the law of cumulative complexity (D6-1). For abstraction, the first main finding was that for longer constructions, the model goes through a phase of abstraction before building up an ever growing inventory of more concrete constructions. This suggests that adult language users may operate with a large number of semi-open constructions, and that the abstractions are merely kept as a failsafe device in case the more concrete constructions cannot be applied. Nonetheless, an answer to the question when it is better to use a more concrete construction than the combination of a more abstract one and a lexical construction, depends on various quantities, viz. the degrees of reinforcement of the two grammatical constructions as well as the lexical one. As we have seen in the previous chapter, a more abstract construction may lead to a more likely analysis than a more concrete one.

The second main finding concerning abstraction was that length-3 constructions (i.e., transitives) were generally more abstract than constructions of other lengths. I argued that this effect is due to the type frequency on the EVENT slot: as many different words occur in it, the more abstract version of the transitive construction accrues more reinforcement as compared to constructions of other length.

Thirdly, I looked at the degree of independence of lexical constructions (section 6.3). Word forms may be strongly associated with lexical constructions, or with parts of grammatical constructions. In the former case, they constitute independent units, whereas in the latter, they should be considered dependent on the grammatical construction they occur in. We found that, for some items, the independence of word forms varies enormously between words, semantic word classes and even simulations. The main factors I identified were (1) the type frequency of the slot of the grammatical construction, (2) the number of constructions a word occurs in, and (3) the token frequency of the word. High values for the former two create more independent lexical constructions, whereas high values for the latter create more dependent word forms. The effect of this is that words in semantic classes that combine freely and have relatively few tokens, such as entity words, or nouns, display stronger independent representations than words in semantic classes that occur in a fixed set of environments, where the environments themselves display little variation, and the token frequencies are high, such as event words, or verbs.

An interesting development over time was found for the event words and pronouns. For both cases we saw that, in some simulations, the word was first used mainly as part of a grammatical construction, then as a free unit, and finally as part of a grammatical construction again. In other simulations, we observed only the second and third stage. Especially these latter cases are at odds with the general conception of learning in a usage-based framework, which states that the learner starts with larger units, which are decomposed over time. However, I argued that these findings do follow from the insights of a starting-small approach as applied to usage-based theory.

Despite this finding, the model does engage in some sort of whole-to-parts learning. When a word form is used mainly as a part of grammatical constructions early in development and later on, by developing strong abstract representations, the model comes to understand the word form as an independent entity, it has effectively performed part-to-whole learning, albeit in a quantitative sense. Qualitatively, the word form has already been established as an independent unit, because the blame assignment (i.e., the creation of a symbolic link to the meaning of the word form) has already been done.

The exploration of the development of the network in section 6.4 highlights several important aspects of SPL. First, the law of cumulative complexity is illustrated with the increase of length and abstraction in the network. Second, we saw how more abstract units may receive strong reinforcement despite their more concrete daughter constructions being less strongly reinforced. Finally, the temporal dimension of SPL becomes clear: some constructions may play an important role early on, but become obsolete as longer and more encompassing constructions enter the scene.

In all of the first three sections, I discussed the between-simulation variation. As a mere effect of the input, I found that (1) some learners rely more on lexically-specific grammatical constructions than others, and (2) that the degree of independence of lexical constructions varies between simulations. Despite this variation, all simulations perform similarly in the comprehension experiment, as well as, as we will see, on the production experiment. This suggests that, even without differing sensitivities to the input data, the order and dispersion of the input items may have an effect on the representations that are built up.

What the analyses in this chapter finally show, is that the models potential for linguistic behavior cannot be directly equated with its behavior itself. We could consider this a re-appreciation of the competence-performance distinction, where the competence is, of course, one that is built up through language use. Just as the strict division of competence and performance may be a false reification of an analytic principle in generative approaches to language acquisition, so may the all-too-strong reliance on behavior to understand the representational system in usage-based approaches constitute a case of the reverse. The fact that, in the usage-based framework, the potential and the use of that potential are considered to be one thing ontologically, does not imply that we can make a direct inference from the use of that potential to the potential itself. This point will be further supported by the production experiments presented in the next chapter.