Chapter 5

Simulating the semantic change of *krijgen* with an exemplar model of language

5.1 Introduction

In the previous chapter, I discussed how the verb *krijgen* has developed since the Middle Dutch period. Its main changes have been the loss of subject agentivity and the development of several auxiliary uses. Around 1300, *krijgen*’s prototypical meaning was ‘to obtain’, with an agentive and therefore animate subject. In its present day use, *krijgen* has two main prototypical uses, one that can be best described as ‘to receive’ (e.g. *De omroep kreeg veel reacties* ‘The broadcasting company received many reactions’) and one with a ‘change of state’ meaning (*Krijgen we ooit nog een strenge winter?* ‘Are we ever going to get another cold winter?’). In both of these cases, the subject is no longer the agent, and does not have to be animate. This bleached meaning of *krijgen* subsequently enabled the development of different auxiliary uses.

I argued that the main mechanism operative in these changes was the extension of the set of objects used with the verb. The characteristics of this set gradually changed, and in turn this led to a changing interpretation of the meaning of *krijgen*.

In this chapter, I will continue the study of *krijgen* with a series of computer simulations. I will present a model of the semantics of *krijgen* that is based on existing exemplar models that were developed for phonology and syntax. Meaning is particularly hard to model because it is not transmitted as directly as other aspects of language like morphology, phonology and even morphologically marked aspects of syntax such as case: while sounds and forms can be perceived and imitated, meaning generally has to be inferred. Of course, there is no absolute difference between the transmission of meaning on the one hand and the transmission of sounds and forms on the other hand. Rather, it can be argued that these aspects of language can be placed on a continuous scale of mode of transmission, in which sounds and forms would be positioned on one end, that of
direct transmission, and meaning would be positioned on the other, that of indirect transmission. The transmission of meaning itself can also occur more directly in some situations than in others; in the case of explicit instruction at school, new meanings are transmitted more directly than when they have to be inferred from context.

In the model I present in this chapter, I will represent this arguably continuous scale of transmission in a simplified way, with form transmitted directly, and meaning indirectly. In this system, the recipient agent in communication needs to infer the meaning of a particular utterance. For this, it will use both its present knowledge (which has been constructed in previous communications) and clues from the utterance itself. I will argue that in such a system, certain regularities and tendencies appear as well, and that these can be directly linked to semantic phenomena such as those observed in the case of *krijgen*. In other words, even a relatively ‘fuzzy’ system like semantics behaves in mechanistic ways that can be modeled, which is very much in line with the view on semantics as proposed in Traugott & Dasher (2002) and various studies on grammaticalization (e.g. Hopper & Traugott 2003). The model of transmission of meaning is in line with the model proposed by Croft (2000).

In this chapter, I will argue that the indirect transmission of meaning leads to a system that is relatively stable, but also easily prone to change. I will discuss factors that affect the amount of change, and factors that affect the preservation or loss of original meaning after a change has occurred. These factors include parameters that affect the individual behavior of agents – such as the likelihood to create new exemplars and the way agents link exemplars to abstract categories – as well as ‘external’ parameters such as population size and the run time of the model. I will argue that an exemplar-based model can produce a relatively realistic simulation of the development of *krijgen*.

### 5.2 Exemplar models and usage-based approaches to language

The main idea behind the model presented here is that a speaker’s knowledge of verb meaning consists of both specific instances, or exemplars, of use of the verb in context, and abstractions that are based on these exemplars. This approach is similar to that proposed by Bybee (2006), who argues for an incorporation of exemplar models (as used in e.g. Pierrehumbert 2001, Bod 2006 and Wedel 2006) into the usage-based view of language (e.g. Bybee 1985, Barlow & Kemmer 2000). Also, the model is in many ways similar to that of De Boer’s model of the vowel system, in which agents construct vowel exemplars based on ‘phonetic’ input (De Boer
SIMULATING THE SEMANTIC CHANGE OF KRUGEN

2001), and shares similarities with the model introduced in Baxter et al. (2006, to appear), which uses Croft’s evolutionary framework for language change (Croft 2000).

Exemplar theory is based on the notion that speakers are continuously sensitive to linguistic experience (Wedel 2006: 250 and references cited there). Linguistic categories (e.g. vowels or syntactic structures) are represented in memory by a large cloud of remembered tokens of that category (Pierrehumbert 2001: 140). These representations are ‘exemplars’: encoded versions of the perceived tokens. Each exemplar has a particular ‘strength’ or ‘activation’ level, which is based on both the frequency with which the exemplar has been perceived, and the recency of the last perception. The assumption is that memory decays over time, and that therefore exemplars are transitory. They can only remain part of the speaker’s knowledge if they are regularly activated by ongoing use.

The notion that linguistic knowledge is directly linked to linguistic experience is also one of the main tenets of usage-based approaches to language, such as construction grammar (Croft & Cruse 2004). Here too, we find the assumption that individual linguistic expressions that are perceived in communication are stored in memory, and that abstract categories are formed on the basis of generalization over these specific units (Bybee 1985, 2001, 2006). Bybee also introduces the notion of ‘lexical strength’, which is roughly similar to that of ‘strength’ or ‘activation’ in exemplar theory: ‘Each time a word in processing is mapped onto its lexical representation it is as though the representation [is] traced over again, etching it with deeper and darker lines each time. Each time a word is heard and produced it leaves a slight trace on the lexicon, it increases in lexical strength. The notion of lexical strength allows us to account for the various effects that frequency has on the behavior of words’ (Bybee 1985: 117, 2006). Another aspect that is shared by both construction grammar and exemplar models of language is the fact that both specific instances (e.g. an utterance like the older, the wiser) and abstract constructions based on these specific uses (the X-er, the Y-er) are part of an individual’s linguistic knowledge. Abstract categories, called schemas by Bybee (2001: 22), are constructed on the basis of similarities between specific exemplars. The number of exemplars of a category, the token frequency of that category, determines its strength. The token frequency will also have an effect on the productivity of the category, in that the higher the number of exemplars that belongs to a category, the greater the likelihood that this category will be used to form new items.

However, not all exemplars contribute equally to the category: the strength of an exemplar’s contribution is inversely linked to its frequency. A high frequency of a specific exemplar only strengthens its own representation, not that of its superordinate category, which is supported by the number of different exemplars.
Exemplars with a high frequency are more autonomous than those with a lower frequency (Bybee 2006: 715), and this autonomy reduces their contribution to the superordinate category (figure 1).

Both ‘layers’ of linguistic knowledge are included in the model I present here: ‘concrete’ exemplars representing actual utterances and abstract categories constructed on the basis of similar exemplars.

Figure 1. A representation of the different role of exemplars on the strength of a particular category. Exemplars 1 and 2 are constructed on the basis of two utterances, and contribute equally to their superordinate category. Exemplar 3 is constructed on the basis of much more utterances: this higher frequency strengthens its own representation, but weakens its relative contribution to the superordinate category.

5.3 Basic structure of the model

In this section, I will give a general description of the computer model, while I will discuss the mathematical details in the next section.

The computer model presented here simulates changes in the transitive use of *krijgen*. The design of the model is based on the model I discussed in chapter 2. In the present model, a group of agents ‘communicate’ with each other by exchanging utterances, which represent sentences containing the verb *krijgen*. The agents base their knowledge of the verb on the utterances they perceive in communication, and the way this occurs is based on two main assumptions that I discussed in the previous section. First, agents are continuously sensitive to linguistic experience: they alter their knowledge of the verb after each communication. This key aspect of the usage-based approach to language is also incorporated in the model presented by Baxter et al. (2006, to appear). Second, agents have no direct access to the meaning of the verb in an utterance, and therefore have to resort to other ‘strategies’ to reconstruct this meaning. In the model, there are two such strategies: (1) agents use their existing knowledge of the meaning of the verb, and (2) agents use contextual information from the utterance.
they perceive. In reality, ‘contextual information’ can be any information that is provided in the utterance that is directly transmitted. In the model, ‘context’ is limited to the direct object. At a later stage in the chapter, I will also discuss an elaborated version of the model, in which this ‘context’ is extended to both the direct object and the kind of subject.

Another key feature of the model which I also discussed in the previous section is that an agent’s knowledge of the meaning of *krijgen* consists of two ‘layers’: specific instances, or exemplars, of use of *krijgen*, and abstractions that are based on these exemplars.

Utterances in the model are represented as [verb - direct object] combinations:

1) **example sentence:** De jongen krijgt een cadeautje
   ‘The boy gets a present’

   **representation in the model:** [V O]

   ![Diagram of agentivity scale](https://example.com/diagram.png)

   **Figure 2.** The representation of the meaning of *krijgen* on a [-1, 1] scale of agentivity. The example sentences serve as indications of the value of *krijgen* associated with the particular uses.

In the model, I will represent both the meaning of the verb (‘*V*’ from here on) and the kind of direct object (‘*O*’ from here on) on a one-dimensional scale, which leads to a two-dimensional space if the two are combined. The meaning of the verb is represented on a one-dimensional scale of agentivity, which ranges from ‘maximal subject agentivity’ on one end to ‘no agentivity’ in the middle, to ‘maximal other participant agentivity’ on the other end (figure 2). This range is based on the different meanings of transitive *krijgen* that I presented in chapter 4.
In chapter 4, I also discussed how the different direct objects that are used with *krijgen* over time could be described in terms of ‘controllability’. The nature of some objects (like *sword* or *present*) is such that they allow for a controlling agent, while other objects (like *fever* or *good weather*) cannot easily be combined with any agent. Note that, in principle, controlling agents can be either the grammatical subject, or another participant (as in *Hij kreeg een cadeau van zijn vader* ‘He got a present from his father’, in which the father is the agent performing the action). In the model, objects that can be used in utterances with *krijgen* are represented on a one-dimensional scale of controllability, ranging from ‘maximally controllable’ to ‘maximally uncontrollable’ (figure 3).

**Figure 3.** The representation of the objects used in utterances with *krijgen* on a [0, 1] scale of controllability.

**Figure 4.** Conceptual space of the possible uses of utterances of *krijgen*. Each coordinate on the map represents a single possible utterance. The area within the dotted line shows the most probable uses.
As I mentioned above, the reduction of utterances to \( V-O \) combinations, in which both items are represented as values on a one-dimensional scale, makes it possible to position each utterance as a point on a two-dimensional map, with verb meaning on the horizontal axis, and object controllability on the vertical axis. The result is a conceptual space that can be used in the analysis of the development of the verb’s meaning (Gärdenfors 2000: 159), as shown in figure 4.

The simulation consists of the iteration of ‘communications’ in which agents produce and perceive \( V-O \) utterances. For each communication, two random agents are selected from the population, one speaker and one hearer. The speaker produces an utterance and the hearer perceives it. Of this utterance, the direct object \( O \) is transmitted directly from the speaker to the hearer: its properties in terms of controllability are directly clear to the hearer. This is, of course, an idealization, but one that is legitimate for the purposes of this simulation: to model the hypothesis that knowledge of a verb’s meaning is (partly) derived from independent information about the objects it is combined with, and that this is a cause of semantic change. In the model, the form of \( V \) is thus directly transmitted, but its meaning is not; the hearer has to reconstruct the verb’s meaning before it can store the utterance as an exemplar. I will explain how this reconstruction works in the next section.

When the hearing agent has reconstructed the value of \( V \), it stores the perceived utterance as an exemplar in its memory. In addition, it will also create or adapt its knowledge of the abstract categories of \textit{krijgen} in its memory. This means that an agent’s knowledge of \textit{krijgen} is present in both stored exemplars of specific utterances and in the abstractions.

It is the speaking agent’s task to produce a \( V-O \) utterance. This production is either the selection of an exemplar from its memory, or the creation of a new exemplar on the basis of an abstract category in the speaker’s memory. The former can be interpreted as conventional language use by means of a cliché, with the transmission of an existing exemplar (retrieved from memory only). The latter can be said to represent conventional but ‘novel’ language use, because it concerns the creation of a new linguistic item according to some general rule.

### 5.4 Mathematical details of the model

As I mentioned in the previous section, the model consists of a population of agents who produce and perceive \( V-O \) combinations. This population consists of 20 agents. Generational turnover, in which agents die and are replaced by newborns, is left out of the model. This way, the effects of the addition of generations can be studied at a possible later stage. The standard simulation consists of 100,000 iterations, and
during each iteration, each agent is involved in 1 communication act as a speaker on average. Agents are randomly assigned the role of speaker and hearer.

**Representation of knowledge of exemplars and abstract categories**

Before explaining how the communication process works, I will first discuss how utterances are represented in the model, and how an agent’s knowledge is constructed. Agents have knowledge of both specific utterances, exemplars, and of abstract categories that are formed on the basis of similar exemplars. Utterances are V-O combinations, and both V and O are represented on a one-dimensional scale. This means that each utterance can be represented by its coordinates \((x, y)\), with \(x [-1, 1]\) as the value for V on the scale of agentivity, and \(y [0, 1]\) as the value for O on the scale of controllability.

Agents have a memory in which they store both the utterances they perceive in communication as exemplars, and the abstract categories that they form on the basis of similar exemplars. This memory is not limited to a particular size. However, an agent’s memory is updated after each communication and exemplars that have not been used or perceived for a particular time are removed. I will discuss this process in more detail later on.

In memory, the coordinates \(e_i \) of each exemplar \(e_i\) are stored, together with the frequency of the utterances of which the exemplar is a token \(f(e_i)\) and its most recent use \(r(e_i)\). Recent use, or recency, keeps track of the number of iterations that have passed since the exemplar was last perceived in communication. For example, \(r(e_i) = 3\) means that the particular exemplar was last perceived 3 iterations earlier. \(r(e_i) = 0\) means that the exemplar has just been perceived.

Agents also have knowledge of abstract categories that are formed on the basis of similar exemplars in memory (Gärdenfors 2000: 109-110). This knowledge consists of the coordinates of the abstract category \(c_i \) and its token frequency \(f_{token}(c_i)\). The token frequency is the number of different exemplars that together make up the abstract category. I will discuss how exemplars are linked to a particular abstract category later on.

The coordinates of the abstract category are calculated by taking the average coordinates of the exemplars that belong to that category. As I discussed in section 5.2, exemplars with a high frequency contribute less to an abstract category than exemplars with a low frequency. Therefore, each exemplar’s contribution to the coordinates of the abstract category is weighed by using the inverse of its frequency. The equations for this calculation are shown in equation 1, and figure 5 gives a graphical representation.
$w_i = \frac{1}{f(e_i)}$

$$c_i(x) = \frac{\sum_{i=1}^{N}(w_i \cdot e_i(x))}{\sum_{i=1}^{N}w_i}$$

$$c_i(y) = \frac{\sum_{i=1}^{N}(w_i \cdot e_i(y))}{\sum_{i=1}^{N}w_i}$$

*Equation 1.* Calculation of the $x, y$ coordinates of exemplar $c_i (c(x) \text{ and } c(y))$, based on the coordinates of a set of exemplars ($e_i(x) \text{ and } e_i(y)$) and the weight $w_i$ of each exemplar. $N$ is the number of relevant exemplars.

![Diagram](image)

*Figure 5.* Representation of exemplars $e$ (black circles) and its abstract category $c$ (open circle). The $x, y$ coordinates and frequency of each exemplar are given, and the coordinates of the abstract category are calculated based on these coordinates, using equation 1. The thickness of the lines represents the strength with which the exemplars contribute to the category.

**Production of an utterance in communication**

An agent who is selected to be a speaker in communication will either select an existing exemplar from its memory, or create a new exemplar based on one of the abstract categories in its memory, which represents novel use. The probability that a particular existing exemplar is selected is proportional to its frequency, a feature that is also present in the utterance selection model of Baxter et al. (2006, to appear). The probability of novel use is given by the parameter $m$. For example, if $m$ is set at 0.1, a speaker has a 10 percent chance of creating a new exemplar in communication.

In the case of novel use, an agent first selects an abstract category from its set of categories, using the token frequencies of the categories as probabilities. This
abstract category is a ‘prototype’ of a cluster of exemplars that have similar characteristics; the coordinates of the category are the center of this prototype, and the further away one gets from this center, the less strong the connection to the prototype becomes (cf. Geeraerts 1997, (Gärdenfors 2000: 5). The coordinates of the new exemplar are based on the coordinates of the abstract category. The new exemplar’s coordinates can deviate from the abstract category’s coordinates, but as the deviation gets bigger, the probability of these coordinates get smaller. For both \(x\)- and \(y\)-coordinates, the deviation \(s\) from the abstract category’s coordinate is calculated with equation 2. In this equation, \(\alpha\) is the maximum allowed distance of a new exemplar from its prototype. In the basic model, \(\alpha = 0.05\), but I will also discuss the effect of different values.

\[
\text{prob}(s = x) = \alpha \cdot e^{-\frac{1}{2}s^2}
\]

Equation 2.

Perception and reconstruction of an utterance in communication

I mentioned above that the hearing agent stores the perceived exemplar \(e_i(x, y)\) in its memory. However, before it can do so, the hearer will have to reconstruct the coordinates of this exemplar. The value \(y\) is directly clear to the hearer: it is the value of the object \(O\) that is directly transmitted from speaker to hearer. On the other hand, the value \(x\), which represents the agentivity of \(V\), has to be reconstructed. In the first step of the perception process, the hearer will try to reconstruct the agentivity of \(V\) with the help of the object \(O\) from the utterance. It assumes that utterances with similar objects will have similar meanings. Therefore, the hearer will scan its memory and compare the stored objects to the uttered object \(O\). First, it will scan the exemplars in its memory. If it finds an exemplar with an object that is similar to that of \(O\), it will assume that the meaning of \(V\) of exemplar and utterance are also similar. It will then store the utterance as this particular exemplar, add 1 to the exemplar’s frequency and set the recency to 0.

If the hearer finds no matching exemplar, it will start to compare the uttered object \(O\) with the objects of the abstract categories in its memory. It will find the most similar object by calculating the distance between \(O\) (the \(y\) value of the utterance) and the \(y\)-coordinate of each abstract category. Only abstract categories for which the distance is smaller than \(\alpha\) are considered, since this is the maximum distance of a newly created exemplar from its prototype.

If a closest category is found, the hearer will assume that the utterance belongs to the particular category. It will then reconstruct the value of \(V\) using the \(x\)-
coordinate of the category as a basis. \( V \) will deviate from this \( x \)-coordinate with a variable amount \( s \). The amount of deviation from the category’s coordinate is in principle unlimited, but the probability of the deviation is linked to its value: as the deviation gets bigger, the probability gets smaller. This is calculated by using equation 2, which was shown earlier.

If no category is found, the hearer assumes that it is being exposed to an unfamiliar utterance. It will then construct a new exemplar \( e \) based on the utterance. The utterance’s object \( O \) is directly clear (see above), and this object will become the object of the new exemplar. Verb meaning is not transmitted directly, and the hearer will use the characteristics of the object to reconstruct it. This reconstruction takes place using a complex procedure. It is based on the characteristics of the object, but the hearer’s present knowledge of verb meaning and the amount of controllability of the object also play a role. Basically, not all objects can be combined equally well with all meanings of \( \text{krijgen} \). For example, uncontrollable objects like \( \text{luck} \) or \( \text{bad weather} \) combine better with non-agentive than with agentive verb meaning. Similarly, objects like \( \text{sword} \) or \( \text{present} \) combine better with agentive than non-agentive meaning (cf. chapter 4). The hearer uses this knowledge in his reconstruction of the meaning of the uttered verb.

A complicating factor in the reconstruction is that there is no exact one-to-one relationship between object and verb meaning. When objects with some controllability are used (e.g. \( \text{sword} \)), both an ‘obtain’ and a ‘receive’ meaning are possible. These meanings are very different in terms of agentivity: in the former, the subject has the role of agent, while in the latter, another participant has this role. In this case, the hearer cannot determine the kind of agentivity on the basis of the object alone. It will therefore use its existing knowledge of verb meaning to determine which of the two kinds of agentivity is most likely. In other words, agents are conservative in meaning construction, as seems realistic.

For example, let us say that a hearer is exposed to an utterance with \( O = 0.8 \) (which e.g. could represent an object like \( \text{sword} \)) and needs to determine the utterance’s verb meaning \( V \). On the basis of \( O \) alone, the verb could be either low on agentivity (representing a ‘receive’ sense) or high (representing an ‘obtain’ sense). The former will have an \( x \)-coordinate of \(-0.8\), the latter of \(0.8\). (in other words, the value of \( O \) determines the value of \( V \), like I discussed in §5.3). The hearer will choose between one of these values by checking which meaning is more likely based on its present knowledge of verb meaning. It will divide its agentivity scale, which runs from \(-1\) to \(1\), in two parts, with \( V \geq 0 \) representing an ‘obtain’ sense, and \( V < 0 \) representing a ‘receive’ sense. For both sides, it will then take the sum of the type frequencies of all abstract categories. Say that the hearer has two abstract categories in its memory. The coordinates of these two categories are \([-0.8, 0.7]\), with a token frequency of 25, and \([0.8, 0.9]\), with a token frequency of 5. In that
case, these sums are simply 25 for $V < 0$, and 5 for $V \geq 0$. The probability that the hearer will select a particular side is based on the sum of the side’s token frequency: $p = 25/30$ for $V < 0$ and $p = 5/30$ for $V \geq 0$.

The choice between one of the two main senses (‘obtain’ for $V \geq 0$ and ‘receive’ for $V < 0$) works well when the uttered object $O$ is relatively controllable (representing objects like present or sword). However, for less controllable objects like luck, a clear agentivity of either ‘obtain’ or ‘receive’ is generally absent, and it is therefore more difficult to determine between one of the two main senses. In other words, the boundary between these two senses is often blurry for less controllable objects, and this aspect is also represented in the model. Before the hearer will select one of the two senses based on the procedure I explained above, there is a probability it will select one of the two at random. This probability is dependent on the controllability of the uttered object $O$: the lower its value, the higher the probability. This is calculated using equation 3. The object’s controllability value $O$ (the $y$-coordinate of the utterance) is compared to a value $v$, which is generated using a Gaussian distribution. This means $v$ has a large chance of being close to 0, and a decreasing chance of deviating from 0. If $O$ is smaller than $v$, verb meaning $V$ will be selected at random. For example, if $O = 0.1$, verb meaning $V$ will be selected at random. For example, if $O = 0.1$ and there is random meaning selection, the choice between $V = 0.1$ and $V = -0.1$ is random.

$$prob(v = x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2}$$

Equation 3.

Finally, once the hearer has successfully reconstructed the coordinates of the new exemplar, it will store this exemplar in its memory with type frequency $f(e_i) = 1$ and recency $r(e_i) = 0$.

Additional updates of the agent’s knowledge after communication

As mentioned earlier, in exemplar theory it is assumed that memory decays, and that exemplars can disappear from memory if their activation level becomes too low. This transitory notion of exemplars is also part of the model presented here. After each communication, the knowledge of both speaker and hearer is updated. This means a review of all exemplars in memory, and a removal of those that have not been used (either in production or perception) for a considerable amount of time. More specifically, the probability of removing an exemplar is inversely proportional to its recency $r$: the less recent an exemplar has been used, the greater the chance of removal, with a threshold of 100 iterations as a minimum age for removal. If a
category has lost all its exemplars, the category itself will be removed from memory as well: empty categories do not remain in existence. Although frequency is not used as a factor in this removal procedure, it is a side effect of a high frequency that the exemplars involved are protected from removal because their high frequency will generally lead to a more recent use.

Apart from the possible removal of exemplars, the knowledge update has two other operations. First, categories that are too close to each other merge and become one new category, a feature that is also present in De Boer’s model of vowel systems (De Boer 2001: 54). The maximum allowed distance between two categories is set at $2\alpha$, in which $\alpha$ is the maximum possible distance an exemplar can have to a prototype for an agent to consider it to be belonging to that prototype.

The newly created category consists of the exemplars of the two merged categories, and its coordinates are the averages of the two original categories, weighted with their respective type frequencies.

The second operation that can occur after communication is a split off of an exemplar with a high frequency from its category, and the formation of a new category. This feature is in line with the assumption that high frequency elements contribute less to their abstract categories than low frequency ones, due to their relative autonomy (Bybee 2006). Whether the frequency of a particular exemplar is high enough for a split off is determined by taking the product $d$ of its distance to the abstract category’s coordinates and its frequency $f(e)$ (equation 4). If this value $d$ is higher than a threshold level (set at 10,000), the exemplar will split off from the category and form a new category. The coordinates of this new category are simply the coordinates of the exemplar. This means that both the exemplar’s relative and absolute frequency play a role in the split off procedure: the distance to the abstract category is a measure of its relative frequency.

$$d = f(e) \cdot \sqrt{(x_e - x_c)^2 + (y_e - y_c)^2}$$ with abstract category $c$ and exemplar $e$.

Equation 4.

Initialization of the simulation

The agents start with an initial knowledge of 10 exemplars, with values randomly distributed between $[0.85 - 0.95]$ for the verb meaning, and $[0.85 - 0.95]$ for the objects, $r = 0$ and $f_{\text{token}} = 1$. This represents the use of *krijgen* in the fourteenth century, as shown in chapter 4, with controllable objects and strongly agentive subjects. This initial setting is then run for 1000 iterations, to allow for new exemplars and abstract categories to form and to prevent the initial 10 exemplars from playing a disproportionately big role in the actual simulation. After this
initiation period, the frequency $f$ of all exemplars of all agents is reset to 1, and their recency is set to 0. Figure 6 shows the knowledge of a random agent at the start of the simulation, after the initiation period, represented in the conceptual space that was shown in figure 4.

![Figure 6](image)

**Figure 6.** Knowledge of a random agent at the beginning of the simulation. Black circles indicate exemplars, the grey circle indicates the abstract category.

**Measurements**

At initiation, the agents are assigned different exemplars within a certain range, and throughout the simulation, this variance in linguistic knowledge remains present because no two agents are exposed to exactly the same linguistic input. The meaning of *krijgen*, therefore, is slightly different for each agent in the population. At the same time, the variation between agents is limited because they use each other’s linguistic input to construct their knowledge. In order to get a good insight in the exact ‘coherency’ of a population in this respect, I calculate the standard deviation of the knowledge of a population of agents. Standard deviation is a measure of variation among the agents: a low standard deviation means the agents’ knowledge differs only little, a high standard deviation means the variation is big. In the presentation of the first results in the next section, I will describe in more detail how exactly I have applied this measure.

Of course, terms like ‘little’ and ‘big’ do not have much relevance unless they are taken relatively: in a model like this, it is not possible to determine that a particular value of the standard deviation represents an ‘incoherent’ population, and another value a ‘coherent’ population. What the measure can give, however, is a means to compare the results of different settings.
Investigated parameters

In the next section, I will discuss the behavior of the model and the effects that different parameters have on the results. I will start with the basic model, whose outline I presented in this current and the previous section. For this basic model, I will show the effect of different settings for parameter $m$, the probability that a speaking agent creates a new exemplar for production. I will then discuss the effect of different alterations to the model. These alterations include ‘agent-internal’ aspects, such as the way input is processed and utterances are produced, and ‘agent-external’ aspects, such as the number of iterations and the size of the population.

5.5 Results

The main goal in this chapter is to try and find simulation parameters that would cause the model I present here to behave like the documented change in the transitive use of *krijgen* (as I presented in the previous chapter). A ‘successful’ simulation of the verb would therefore cause an extension in meaning, followed by a loss of the original meaning. This change should also happen over a realistic time span.

The basic model

In a first series of simulations, the probability of creating a new exemplar in production is set at $m = 0.1$, which means that 1 out of 10 times, a speaking agent will create a novel exemplar instead of selecting one from its memory.

Figure 7 on the next page shows the development of one agent’s knowledge of *krijgen* over 100,000 iterations with this parameter setting. As a matter of fact, the lack of development is the most obvious, because the knowledge at the end of the simulation is very similar to that at the beginning. That is, more abstract categories have formed over time, but there is no significant change in the kind of combinations of objects and verb-meanings. The growing number of categories is due to the fact that over time, more exemplars obtain a frequency that is high enough to become a new category. This, in turn, is due to the relatively high probability of selecting existing exemplars instead of creating new ones.

In figure 7, only the development of one agent from the population is shown; the interesting question is, however, how the knowledge of the population as a whole develops over time. Figure 8 on the next page is a plot of the knowledge of all agents in the population after 100,000 iterations; it shows the range of knowledge of categories of each agent in the population on the meaning scale.
Figure 7. Development of knowledge of a random agent over 100,000 iterations, with $m = 0.1$. Black circles indicate exemplars, grey circles indicate categories.

Figure 8. Representation of the knowledge of a population of $N = 20$ agents. Each grey bar indicates the range of knowledge of categories on the meaning scale for one agent.
As figure 8 shows, there is variation in the linguistic knowledge of the agents, but this variation is not too large. The amount of variation on both sides differs: there is more variation on the minimum side than on the maximum side.

This is due to the fact that the agents’ knowledge is limited to values between -1 and 1. The maximum value of the knowledge of many agents has reached its upper limit, while this is not the case for the minimum value. In any case, the population can be said to be linguistically quite coherent.

As I mentioned in the previous section, the amount of linguistic coherency in the population can be measured by the standard deviation of the agents’ knowledge: the smaller the deviation, the greater the linguistic coherency between the agents in the population. This is shown in figure 9. This figure is a graph in which both the linguistic knowledge and the standard deviation are shown. For the linguistic knowledge, the average minimum and maximum meaning values of the population are calculated and plotted as two dots on the x-axis. The minimum values are represented by squares, the maximum values by circles (note that in figure 9, the circles are all on the bottom right corner). The standard deviation of this knowledge for the population is plotted on the y-axis. Thus, one pair of dots in figure 9 represents one population or run, and the figure shows the result of 20 runs with the same settings as the previous graphs. The position of a dot on the y-axis is in indication of how much the knowledge of the agents within a particular population differs: the greater the value, the greater the difference among the agents.
Figure 10a, b. Examples of the development of knowledge after 100,000 iterations for different settings of $m$. From top to bottom, $m = (0.3, 0.5, 0.7, 0.9)$. For each setting, the result of one agent in a random run is shown in figure (a), and the result of 20 independently run populations is shown in figure (b), with the average knowledge of the meaning extremes on the x-axis and the coherency on the y-axis.
As figure 9 shows, this particular setting leads to a system in which, after 100,000 iterations, the linguistic knowledge of the population has consistently but marginally expanded, while the original knowledge, as was shown in figure 6, remains present. There is some variation among the agents in the population (with the standard deviation ranging between 0.08 and 0.16 for the minimum meaning value) but overall, the linguistic knowledge of all agents is relatively consistent.

The question is with what differences in parameter settings a system can be obtained that shows more change. An obvious candidate for this is the amount of novel use in the system. A higher creation rate of new exemplars \( (m) \) will lead to more variation, and variation is a necessary source for change.

Figure 10a shows the results of increasing values for \( m \) (0.3, 0.5, 0.7, 0.9) after 100,000 iterations for one agent. Figure 10b shows the results of 20 different populations in a similar way as in figure 9, with the average minimum and maximum meaning on the x-axis and the standard deviation on the y-axis.

By looking at the results in figures 10a and 10b, it is obvious that the amount of novel use correlates with the amount of change. With \( m = 0.3 \), categories develop that have change of state meaning (in the bottom middle of the graph), and with increasing values for \( m \), the number of categories with this meaning grows. At the same time, this development usually does not go any further for \( m = 0.3, m = 0.5 \) and \( m = 0.7 \); it is only with \( m = 0.9 \) that in most of the simulations, a ‘receiving’ meaning comes into existence (with values on the meaning scale between -1 and 0).

A striking aspect about the development of the verb in all these cases is that it is extending and never shifting: the original meaning of the verb remains present in all the performed runs in all settings. Apparently, there is not enough pressure for these original meanings to disappear.

The graphs also show that a higher amount of novel use (and a bigger amount of extension of the linguistic knowledge) seems to correlate with a lower linguistic coherency of the population: the standard deviation of the agents’ knowledge, particularly of their minimum meaning value, increases with higher values for \( m \). The average standard deviation of the minimum value over 20 runs for \( m = 0.1 \) is 0.11, while it is 0.21 for \( m = 0.3 \) and 0.33 for \( m = 0.9 \). This suggests that a higher rate of novel use correlates with bigger differences in individual linguistic knowledge. One probable reason for this is that an increased rate of \( m \) leads to more novel exemplars created on the basis of individual knowledge instead of on the basis of communication between agents; this will decrease the coherency between the agents. An agent might have created a novel use with a stronger ‘receiving’ sense and use this in communication, upon which this use spreads to the other agent.

However, the high amount of novel use limits the spread of this new use, because the use of conventional exemplars is low by definition; this will lead to a decrease in coherency between the agents. The high values in standard deviation can
be explained by the fact that an extension in linguistic knowledge (towards the ‘receiving’ sense) has not spread to all agents in the population. To give an example, in one of the runs with \( m = 0.9 \), 4 of the 20 agents had a minimum meaning value between 0.66 - 0.89, while all other agents had a minimum meaning value between 0 - 0.2. This seems to be a typical outcome for this setting: a majority of 14 - 16 agents with an extension in their linguistic knowledge towards the ‘receiving’ sense, and a small group of agents who do not share this extension. This particular outcome might also well be the result of the limited amount of exchange of conventional exemplars between agents.

In summary, an increase in the relative amount of novel use does lead to an extension of the use of *krijgen*. At the same time, a ‘receiving’ sense only comes into existence with a very high rate of novel use \( (m = 0.9) \). Change occurs, but only in the form of extension; a loss of the original meaning of *krijgen* cannot be obtained with these settings. For all settings, the changes occur in the majority of the agents in the population.

### Different values of parameter \( \alpha \)

In the model, \( \alpha \) is a parameter that affects the way exemplars are connected to their prototypes: it gives the maximally allowed distance of an exemplar from its prototype. Also, it is a measure of the maximal distance between two abstract categories. In the basic model, \( \alpha = 0.05 \). I now look at how two different rates of \( \alpha \) affect the development of meaning, with different values for \( m \). Figure 11 shows the results for \( m = 0.9 \) for both \( \alpha = 0.025 \) and \( \alpha = 0.075 \). I have left out the results for other values of \( m \) (0.1, 0.3, 0.5, 0.7, 0.9), because, despite the differences in numbers of prototypes and exemplars per prototype, the different values of \( \alpha \) do not lead to significant changes in the amount of change after 100,000 iterations, compared to the results of the basic model (figures 7, 9 and 10). For both values of \( \alpha \), an increase in the value of \( m \) leads to an extension of the use of the verb, comparable to that of the basic model.

With \( \alpha = 0.025 \), the maximally allowed distance of an exemplar from its prototype is twice as small as in the basic model. On the one hand, this means that in interpretation, exemplars will sooner have exceeded this maximally allowed distance and as a result, an interpreting agent could form a new prototype on the basis of this prototype. On the other hand, exemplars will differ less from their prototype in the process of exemplar production.
For $\alpha = 0.075$, we would expect the opposite: exemplars will be interpreted as belonging to a particular existing prototype easier, while in production, exemplars can differ more from their prototype than in the basic model. Also, the parameter $\alpha$ is used in the possible merging of prototypes: if prototypes are less than $2\times \alpha$ away from each other, they will become one new prototype. Therefore, with a smaller $\alpha$, more prototypes will remain in existence, while with a bigger $\alpha$, less prototypes will remain in existence.

As expected, the runs with a different value for $\alpha$ do show a difference in the number of exemplars per prototype and the number of prototypes: a low $\alpha$ leads to relatively many prototypes, with relatively few exemplars per prototype. A high $\alpha$ leads to the opposite: relatively few prototypes, with relatively many exemplars per prototype. Also, a high $\alpha$ leads to more coherency in the population, which can also be prescribed to the lower number of prototypes.
Differing the value of $\alpha$ does not seem to have an effect on the amount of change that takes place in the model, with different values of $m$. As figure 11 shows, $m = 0.9$ leads to an extension of meaning for both $\alpha = 0.025$ and $\alpha = 0.075$, similarly to the outcome for $\alpha = 0.05$ in the basic model. What is also similar is the outcome that in both cases, no loss of the agentive meaning occurred.

**Allowing the frequency of use to vary**

In the model, frequency of use plays a key role in shaping the linguistic knowledge of the agents: exemplars that are perceived frequently become entrenched more strongly and will become a stable part of an agent’s knowledge. Also, frequency is the crucial factor that determines whether an exemplar should be separated from an existing category and form its own new one.

In the basic model, the frequency of use was set at 1 communication act per agent per iteration on average. In the present alteration to the basic model, the frequency of use per iteration is allowed to vary, meaning that the average number of communications between agents is no longer fixed, as it is in the basic model. Arguably, the frequency of use of a word is at least partially linked to its range of meanings, since these meanings reflect the range of contexts in which the word can be used.

A variable frequency of use $\phi_{use}$ is added to the basic model by linking the number of communications per iteration to the number of exemplars of the agents in the population. The total number of exemplars at $t = 0$ is taken as the standard, from which the relative number of exemplars at any time is calculated (equation 5). For example, let us say that the number of exemplars of all agents in the population at $t = 0$ is 100, and that, at the present moment ($t = T$), this number is 250. This means that the original frequency of use is now multiplied with $250/100 = 2.5$. The frequency of use at the present moment is therefore 2.5 times higher than at $t = 0$, due to the fact that the agents have more exemplars. The general form of this calculation is shown in equation 5.

$$\phi_{use} = \frac{\sum_{t=0}^{T} e_i^{t}}{\sum_{t=0}^{T} e_i^{0}}$$

with $N = 20$ agents

**Equation 5.**

---

1 Examples of other factors that could possibly affect the frequency of use are the word’s register and the presence of (near-)synonyms in the language.
Figure 12a, b. Examples of the development of knowledge after 100,000 iterations in which the frequency of use is allowed to vary. From top to bottom, $m = (0.1, 0.3, 0.5, 0.7, 0.9)$. For each setting, the result of one agent in a random run is shown in figure (a), and the result of 20 independently run populations is shown in figure (b), with the average knowledge of the meaning extremes on the x-axis and the coherency on the y-axis.
Figure 12 on the previous page shows the results. It is obvious from these results that they are very similar to the ones in the basic model: there are no significant differences between them. There is an extension of the use of *krijgen* with higher rates of novel use, but a loss of the original meaning does not occur.

If we look more closely at the average values of the minimum meaning in figure 12b, it seems though as if the flexible frequency has a slight impeding effect on change in the population. In the basic model, the average minimum value got smaller with an increasing value for \( m \): 0.52 (\( m = 0.1 \)), -0.22 (\( m = 0.5 \)), -0.66 (\( m = 0.9 \)). The current model, with a variable frequency of use, shows a similar tendency, but the average minimum value does not decrease as much as in the basic model: 0.50 (\( m = 0.1 \)), -0.10 (\( m = 0.5 \)), -0.49 (\( m = 0.9 \)). This ‘slower’ pace is accompanied by a slightly higher coherency of the population. In the basic model, the average standard deviations are 0.11 (\( m = 0.1 \)), 0.21 (\( m = 0.5 \)) and 0.33 (\( m = 0.9 \)). In the current model with a variable frequency of use, these values start off lower for low values of \( m \), but ‘catch up’ when \( m \) gets higher: 0.10 (\( m = 0.1 \)), 0.11 (\( m = 0.5 \)) and 0.35 (\( m = 0.9 \)).

Overall, a flexible frequency does not seem to lead to significant changes in the results of the simulations. If there is any difference with the basic model, it is that it actually has a small impeding effect on change. However, a combination of both extension and loss of meaning does not seem to occur in this particular extension of the basic model.

**Using a skewed distribution in the frequency of use**

When trying to obtain a more realistic simulation in terms of frequency of use as I did in the previous section, the question is whether there is an even distribution in the frequency of use over all different senses of *krijgen*. One reason to argue against this is that strongly agentive uses only allow for intentional and thus animate subjects, while non-agentive uses allow for both animate and inanimate subjects, because the subject does not have to be intentional. This means that the non-agentive use, if it exists, has a relatively higher probability of being used in communication than the agentive use, i.e. has a bigger chance of being selected by a speaker than an exemplar of agentive use, even when the frequencies are the same. I will discuss this feature in more detail in this section.

In a subsequent extension of the model, the probabilities of using agentive and non-agentive senses of *krijgen* differ: non-agentive senses are given a greater chance of being used in communication than agentive senses. This simulates the idea that non-agentive senses have an advantage over agentive senses, in that they can be used with a wider range of subjects: animate and inanimate. In the basic model, the probability of selecting an exemplar in conventional use is determined by
its frequency, and the probability of selecting a category for novel use is determined by its token frequency. In the extended model, these probabilities are made partially dependent on the position of the exemplars or categories on the meaning scale. That is, both an exemplar’s frequency and its \( x \)-value determine its chance of being selected for communication. Similarly, both a category’s token frequency and its \( x \)-value determine its chance of being used for the creation of a novel exemplar.

This dependency on the position of an exemplar or category on the meaning scale is represented by the parameter \( s \); an exemplar’s frequency or a category’s token frequency is multiplied with this value, which is dependent on the position of the exemplar, or category, on the meaning scale (\( V \)). I will discuss two functions to calculate \( s \).

**Function 1: linear**

In this function, the value of \( s \) is linked to the agentivity scale by a simple linear function, as figure 13 and equation 6 show. With this function, the chance of selecting or creating an exemplar decreases linearly with increasing meaning values, with \( s \) ranging between \( s = 1.0 \) for agentive meanings and \( s_{\text{max}} \) for non-agentive meanings.

\[
s = \frac{(1 - s_{\text{max}}) \cdot V}{2} + 0.5 \cdot s_{\text{max}} + 0.5
\]

**Equation 6.**

Next, I will explore the consequences of different values of \( s_{\text{max}} \). The higher this value, the stronger the advantage for non-agentive use. I will use a fixed value for the rate of novel use and vary the value of \( s_{\text{max}} \) between 1.1 and 1.4 (note that \( s_{\text{max}} = 1.0 \) is equal to not using this parameter at all). For the rate of novel use, I use a value of \( m = 0.5 \), which means that existing exemplars are used in utterances in 50 percent of communication, and newly created exemplars in the other 50 percent. As I have shown earlier, this value leads to an extension of the use of *krijgen* to a change of state meaning in the runs with both a fixed and a variable frequency of
use. In a few cases, a ‘receiving’ sense started to develop, but this sense never became fully grown. Also, the original agentive meaning remained present in all runs.

Figure 14a, b. Examples of the development of knowledge after 100,000 iterations with different settings for $s_{max}$ using a linear function, with $m = 0.5$. From top to bottom, $s_{max} = (1.1, 1.2, 1.3, 1.4)$. For each setting, the result of one agent in a random run is shown in figure (a), and the result of 20 independently run populations is shown in figure (b).
Figure 14 shows the results after 100,000 iterations for four different settings for $s_{\text{max}}$ (1.1, 1.2, 1.3, 1.4). The results of one random agent from a population in the graphs are shown in 14a, and the average minimum and maximum meaning values of 20 populations from 20 runs are shown in 14b. Quite surprisingly, the addition of the parameter $s$ to the model does not seem to lead to significantly different development for any of the values tested when compared to the basic model. Although $s$ adds a preference for non-agentive meaning to the model, this preference is apparently not strong enough to lead to a strong development of non-agentive meaning, let alone a loss of agentive meaning. Also, the value of $s$ does not seem to make a significant difference either: for all four tested values, the results are comparable to the runs with the same value for $m$ in the basic model (shown in figures 7, 9 and 10).

**Function 2: a fixed $s$ for non-agentive meaning**

In this function, there is no linearly increasing preference for non-agentive meaning. Instead, the value of $s$ is fixed at $s = 1$ for strongly agentive meanings and fixed at $s_{\text{max}}$ for all non-agentive meanings. The value of $s$ decreases from $s_{\text{max}}$ to 1 between $V = 0$ and $V = 0.5$. This function can be argued to be a more realistic representation of preferences for non-agentive meaning versus agentive meaning than the previous function. Non-agentive senses have a bigger chance of being used than agentive senses, but there is no difference in preference within the non-agentive senses themselves: since all non-agentive senses can be used with similar animate and inanimate subjects, their chance of being used is the same. Figure 15 and equation 7 show the function of $s$, figure 16 on the next page shows the results.

**Figure 15.** The value of $s$ depends on the position of the exemplar on the verb agentivity scale and $s_{\text{max}}$.

\[
\begin{align*}
  & \text{if } V \leq 0 \quad s = s_{\text{max}} \\
  & \text{if } V > 0 \text{ and } V < 0.5 \quad s = s_{\text{max}} - 2 \cdot s_{\text{max}} \cdot V + 2 \cdot V \\
  & \text{if } V \geq 0.5 \quad s = 1.0
\end{align*}
\]

Equation 7.
Figure 16a, b. Examples of the development of knowledge after 100,000 iterations with different settings for $s_{\text{max}}$ using a non-linear function, with $m = 0.5$. From top to bottom, $s_{\text{max}} = (1.1, 1.2, 1.3, 1.4)$. For each setting, the result of one agent in a random run is shown in figure (a), and the result of 20 independently run populations is shown in figure (b), with the average knowledge of the meaning extremes on the x-axis and the coherency on the y-axis.
Let us first look at the results of one random agent in figure 16a. In all previous settings, the runs with \( m = 0.5 \) did not lead to a strong development of the receiving sense. For example, in the basic model, (shown in figure 10), the minimum value that developed after 100,000 iterations was -0.22, averaged over 20 runs with 20 agents each. In the current setting, the receiving sense develops much stronger, with -0.39 (\( s_{\text{max}} = 1.1 \)), -0.47 (\( s_{\text{max}} = 1.2 \)), -0.67 (\( s_{\text{max}} = 1.3 \)) and -0.89 (\( s_{\text{max}} = 1.4 \)). Just as higher values of \( m \) lead to a stronger development of the receiving sense, so do higher values of this function of \( s_{\text{max}} \).

Contrary to the linear function of \( s_{\text{max}} \), the current function of \( s_{\text{max}} \) also leads to an interesting development of the original agentive meaning for similar values of \( s_{\text{max}} \). In the basic model and in the linear function of \( s_{\text{max}} \), this meaning generally remained present in the linguistic knowledge of the agents in the population. This is also true for \( s_{\text{max}} = 1.1 \) and \( s_{\text{max}} = 1.2 \). Yet, as figure 63a shows, the original meaning already starts to become weaker for \( s_{\text{max}} = 1.3 \) and has disappeared from the linguistic knowledge of the agent shown for \( s_{\text{max}} = 1.4 \) after 100,000 iterations. In the latter case, a shift in verb meaning has taken place, instead of just an extension. The parameter \( s_{\text{max}} \) for this function might therefore be interpreted as a pressure not only favoring the receiving sense, but also against the agentive use. The mechanism underlying this pressure is the frequency of use: the fact that it can be used more for one sense than for another will act as a pressure for change.

Still, figure 16a only shows the knowledge of one randomly chosen agent, and it is important to see whether this shift in knowledge also occurs throughout the population. In figure 16b, let us first look at the minimum meaning. For \( s_{\text{max}} = 1.1 \) and \( s_{\text{max}} = 1.2 \), the standard deviation of the populations is quite high (0.23 for \( s_{\text{max}} = 1.1 \) and 0.26 for \( s_{\text{max}} = 1.2 \)), which means the linguistic coherency in the population is not that high. In the previous settings, high values in standard deviation were not merely a sign of little linguistic coherency. Rather, they were the result of a difference in knowledge between a small minority in the population and the majority. Such a clear divide is not that obvious in the present settings for \( s_{\text{max}} = 1.1 \) and \( s_{\text{max}} = 1.2 \). For \( s_{\text{max}} = 1.1 \), the minimum meaning values of the agents in one random population after 100,000 iterations range between -0.01 and -0.78, without a clear grouping of agents on either ‘side’ of these two extreme values. Rather, they were the result of a difference in knowledge between a small minority in the population and the majority. Such a clear divide is not that obvious in the present settings for \( s_{\text{max}} = 1.1 \) and \( s_{\text{max}} = 1.2 \). For \( s_{\text{max}} = 1.1 \), the minimum meaning values of the agents in one random population after 100,000 iterations range between -0.01 and -0.78, without a clear grouping of agents on either ‘side’ of these two extreme values.

The results for \( s_{\text{max}} = 1.3 \) and \( s_{\text{max}} = 1.4 \) in figure 16b might show an explanation of this lack of coherency in the population. For \( s_{\text{max}} = 1.3 \), the standard
deviation is also still relatively high, but decreasing (0.21), and for $s_{\text{max}} = 1.4$, it is very low (0.08): the agents have reached the limits of their knowledge as far as the model is concerned, with -1 being the minimum meaning value that is possible. The higher values for the standard deviation in the other settings of $s_{\text{max}}$ are probably caused by the fact that the linguistic knowledge is shifting towards the minimum, but that this shifting takes place at a different rate for each agent. When we look at the maximum meaning, the agentive senses, we see a similar pattern: as $s_{\text{max}}$ gets bigger, the pressure towards the receiving sense gets stronger and the coherency between the agents gets weaker.

As an agent’s knowledge gradually shifts from ‘agentive’ towards more ‘receiving’, the original agentive meaning slowly weakens and finally disappears. In this process, a few exemplars with a high frequency can resist disappearing for quite a while, due to their high frequency. This can be seen in figure 17. Such survival of a very specific use of a once much broader knowledge is in fact quite realistic. For example, as I discussed in the previous chapter, the original meaning of *krijgen* is still used in present-day Dutch, but only in a few highly specific uses.

![Figure 17](image.png)

**Figure 17.** Example of a run in which the prototypical use of the verb has shifted after 100,000 iterations, but in which some of the original use remains in existence. $s_{\text{max}} = 1.4$, $m = 0.5$.

**Allowing the model to run for 500,000 iterations**

In all previous settings, the simulations have been run for 100,000 iterations. However, this number is a rather random limit, primarily chosen for reasons of time and computing power. It might well be possible that interesting developments in the model need more time and may occur beyond the current limit of 100,000 iterations. To check this, I have therefore also run the model for 500,000 iterations, in the settings of the basic model, with different values for $m$. Figure 18 shows the results.
Figure 18 a, b. Examples of the development of knowledge after 500,000 iterations with different settings for $m$. From top to bottom, $m = (0.1, 0.3, 0.5, 0.7, 0.9)$. For each setting, the result of one agent in a random run is shown in figure (a), and the result of 20 independently run populations is shown in figure (b), with the average knowledge of the meaning extremes on the $x$-axis and the coherency on the $y$-axis.
In the runs with the basic model and 100,000 iterations (shown in figures 7, 9 and 10), there was virtually no development of non-agentive meaning for low values of $m$, while higher values of $m$ lead to an extension to non-agentive meaning. As is obvious from figure 18, the results after 500,000 iterations are very similar to this. For example, with a low amount of creation of novel uses ($m = 0.1$), there is only a slight extension of knowledge, but this extension seems to come to a stop instead of continuing slowly but steadily: if the extension would keep on occurring, we would expect a bigger extension after 500,000 iterations than after 100,000 iterations, but this has not happened. An explanation for this is that a low amount of novel use automatically means a high amount of conventional use (with a rate of 0.9). This means that the already existing exemplars and categories are used much more frequently in communication than novel exemplars are created, which will only increase their entrenchment (in the form of high frequencies for the exemplars). Simply said, novel exemplars cannot ‘compete’ with this, and therefore will never survive long enough in order to lead to a significant change in the population’s knowledge. The small extension that has taken place has probably occurred at the beginning of the simulation, when the frequencies of the existing exemplars are still relatively low. Also, we do not see any differences with regards to the loss of agentive meaning when the model is run for an extended number of iterations. For higher values of $m$, the development of meaning after 500,000 iterations does not differ significantly from that after 100,000 iterations: in both cases there is development of non-agentive meaning, while agentive meaning remains present.

Before, I showed that the use of parameter $s_{\text{max}}$ (in the non-linear function) could lead to a loss of the original agentive meaning with a relatively high value for $s_{\text{max}}$. Figure 19 shows the results of using this parameter for 500,000 iterations. In general, increasing the number of iterations does not seem to lead to very big changes in the results of the model. With increasing values of $s_{\text{max}}$, we see an extension towards non-agentive meaning. As in the case of 100,000 iterations, we also see a loss of the original agentive meaning for $s_{\text{max}} = 1.3$ and $s_{\text{max}} = 1.4$. In these runs, there were differences between the agents’ meanings: while some showed a loss of agentive meaning, this meaning remained present for others, albeit in the form of a few abstract categories. The main difference between $s_{\text{max}} = 1.3$ and $s_{\text{max}} = 1.4$ was one of degree: the number of agents that showed a loss of agentive meaning was higher for $s_{\text{max}} = 1.4$ than for $s_{\text{max}} = 1.3$.

The main difference between the results after 100,000 and 500,000 iterations is also one of degree. When given more time, more agents in the population show a loss of agentive meaning for $s_{\text{max}} = 1.3$: the average meaning range of the agents is $(-0.82 - 0.85)$ while it is $(-0.67 - 0.92)$ after 100,000 iterations. For $s_{\text{max}} = 1.4$, the results of the two settings are almost similar: the average meaning ranges for 100,000 iterations and for 500,000 iterations are $(-0.90 - 0.65)$ and $(-0.89 - 0.63)$. 
Figure 19 a, b. Examples of the development of knowledge after 500,000 iterations with different settings for $s_{\text{max}}$ using the non-linear function, with $m = 0.5$. From top to bottom, $s_{\text{max}} = (1.1, 1.2, 1.3, 1.4)$. For each setting, the result of one agent in a random run is shown in figure (a), and the result of 20 independently run populations is shown in figure (b).
Summarizing, an extended run time of 500,000 iterations does not lead to results that differ significantly from those after 100,000 iterations. This can be seen as an indication that a run time of 100,000 iterations is sufficient to catch the main developments and effects of investigated parameters in the model.

**Using a population of N = 50**

In the previous runs, the population size has been set at \( N = 20 \). This value was chosen as a compromise between a sufficiently realistic group size and computing power and speed. However, group size might have an effect on the behavior of the population in its development of linguistic knowledge: in a bigger population, it could be harder for innovations to spread because the relative amount of communication between agents decreases. This latter factor might also lead to a lower coherency of the population.

For \( N = 50 \), I have tested the settings of the basic model, and the settings for the parameter \( s_{\text{max}} \) (in the non-linear function, with \( m = 0.5 \)). Figure 20 on the next page shows the result for the ‘basic’ setting. As the figure shows, using a larger population size of \( N = 50 \) leads to a development for different values of \( m \) that is very similar to using a population of \( N = 20 \) (which is shown in figures 7, 9 and 10): with increasing values of \( m \), there is an increasing extension of the meaning range towards non-agentive meaning. For example, for \( m = 0.1 \), the average minimum meaning over 20 populations is 0.52 for a population of \( N = 20 \) and 0.51 for \( N = 50 \) and for \( m = 0.9 \), the average minimum meaning for a population of \( N = 20 \) is -0.66 while it is -0.68 for a population of \( N = 50 \). This similarity also shows in the coherencies of the two settings. The average coherency measured of the minimum meaning and \( m = 0.1 \) over 20 populations is 0.11 for both \( N = 20 \) and \( N = 50 \). For \( m = 0.9 \), these values are 0.33 (\( N = 20 \)) and 0.30 (\( N = 50 \)). However, a difference between the two settings is the variation in the amount of coherency, when measured over 20 runs. In this model, coherency is measured as the standard deviation of the average (minimum and maximum) meanings of the agents in a population. That is, a high standard deviation means that there is a high degree of difference between the agents’ meanings within the population: one agent could have developed a strong non-agentive meaning while the other has not. A low standard deviation means the agents’ meanings are more alike. When we compare the coherency measurements from figure 20b (\( N = 50 \)) to those of figures 9 and 10b (\( N = 20 \)), it turns out that over 20 runs, there is less variation in coherency for \( N = 50 \) than for \( N = 20 \). If we measure the standard deviation of the 20 measures of coherency for \( m = 0.1 \), this is 0.022 (\( N = 20 \)) and 0.015 (\( N = 50 \)). For \( m = 0.9 \), it is 0.045 (\( N = 20 \)) and 0.028 (\( N = 50 \)). A larger population in this model thus does not
Figure 20 a, b. Examples of the development of knowledge after 100,000 iterations with a population of $N = 50$ and with different settings for $m$. From top to bottom, $m = (0.1, 0.3, 0.5, 0.7, 0.9)$. For each setting, the result of one agent in a random run is shown in figure (a), and the result of 20 independently run populations is shown in figure (b), with the average knowledge of the meaning extremes on the $x$-axis and the coherency on the $y$-axis.
lead to a difference in the amount of change or differences in the coherency in the model, but it does seem to lead to a more coherent pattern in development. This could be explained by the fact that in larger populations, the effect of one individual with knowledge that differs from that of the rest of the population, is less than in smaller populations.

I also tested the use of parameter $s_{\text{max}}$ with a non-linear function for a population of $N = 50$ (figure 21). Again, the main development with the use of this parameter is similar to that for a population of $N = 20$: an increasing value of $s_{\text{max}}$ leads to a shift from agentive to non-agentive meaning. For $s_{\text{max}} = 1.4$, the original meaning has disappeared in the knowledge of most individuals in the 20 tested populations. The main difference is that for $s_{\text{max}} = 1.3$, there is a stronger shift for $N = 50$ than for $N = 20$. The average minimum and maximum meanings over 20 runs is (-0.87 - 0.74) for $N = 50$, while it is (-0.67 - 0.92) for $N = 20$. In other words, the same value of $s_{\text{max}}$ leads to higher amount of change with a larger population size.

This finding runs contrary to the expectation I mentioned at the beginning of this section, that a larger population size might show less change than smaller populations, because it is more difficult for innovations to spread through a larger population. The rationale behind this is that there is relatively less contact between agents when population size is large. Still, that this effect does not show in this model is not surprising. A larger population size has no effect on the amount of communication between agents, because, no matter what the size of the population is, agents still communicate with each other on average one time during each iteration cycle. This could also explain the higher rate of shift that we see for $N = 50$: more agents create new exemplars that have non-agentive meaning and because the relative amount of communication remains the same, there is a higher chance that these ‘innovations’ can spread and that, as a result, there is a general shift in knowledge throughout the population.

5.6 Discussion and conclusions

In this chapter, I have presented a series of computer simulations of the semantic development of the Dutch verb *krijgen*. These simulations were carried out with a computer model that was based on existing exemplar models and usage-based approaches to language.

The main point of this chapter was to show that the historical development of meaning, too, can be modeled in terms of exemplars. Although the transmission of meaning might be occurring indirectly, the interpretation of meaning can be linked to aspects of transmission that take place more directly. That is, language
Figure 21 a, b. Examples of the development of knowledge after 100,000 iterations with a population size of \( N = 50 \) and with different settings for \( s_{\text{max}} \) using the non-linear function, with \( m = 0.5 \). From top to bottom, \( s_{\text{max}} = (1.1, 1.2, 1.3, 1.4) \). For each setting, the result of one agent in a random run is shown in figure (a), and the result of 20 independently run populations is shown in figure (b), with the average knowledge of the meaning extremes on the x-axis and the coherency on the y-axis.
users are not completely clueless in their interpretation of utterances, but use their knowledge from previous communications and cues from the utterances themselves. As I have shown in this chapter, this makes it possible to construct a model that simulates semantic development in a relatively realistic way. In such a model, linguistic knowledge consists of both stored exemplars of specific uses of the verb, and abstract categories constructed on the basis of these exemplars. Although this means that no two agents have exactly the same linguistic knowledge (as no two agents are exposed to exactly the same input), the overall knowledge of the agents in the population remains highly similar. ‘Change’ in the model occurs when new uses of the verb are introduced and spread through the population.

The construction of this model shows how the implementation of linguistic theory in a model forces one to be precise. This is particularly true for the notion of frequency and the autonomy of exemplars. As I discussed in section 5.4, exemplars in the model can split off from their abstract category and form an abstract category of their own if their frequency becomes very high. This feature is in line with the assumption that high frequency elements contribute less to their abstract categories than low frequency ones (Bybee 2006).

One striking result of the simulations was that semantic extension could be easily obtained by allowing either more novel use in the system or by using a skewed distribution of the frequency of use. However, semantic shift, which actually occurred in the case of *krijgen*, turned out to be much harder to simulate. This kind of development only took place with the presence of a skewed distribution for the frequency of use of different uses of the verb. On the one hand, this added feature can be said to be realistic: it is based on the observation that some uses of the verb can be applied more frequently than other uses. Also, this outcome seems to support that of the model I discussed in chapter 2, in which differences in the frequency of use also served as a pressure for directional change.

On the other hand, it is also interesting to note that without the skewed frequency, only extension of knowledge took place and not shift. This indicates that an exemplar approach predicts that, in the absence of any selection pressures, extension of linguistic knowledge, and not shift, is the standard development.

Apart from this dichotomy of either preservation or loss of the original meaning, the model also captured an aspect of semantic change that may actually be occurring more frequently than either one of the above ‘extreme’ scenarios: that of semantic shift with the preservation of certain highly concrete original uses.

In the model of *krijgen* I have presented in this chapter, out of all conditions and parameter settings, one setting generated results that were comparable to the actual changes that occurred in the verb. This, however does not necessarily have to mean that there could be other ways to achieve similar results: I have only been able to
investigate the role of a limited number of conditions and parameters, which meant excluding other possible interesting factors. For example, the model is a one-word model, but words are commonly not used in isolation. It might therefore be feasible that competition between krijgen and other words (particularly words with comparable semantics) has played an important role in the verb’s development as well. In this model, agents communicated with each other a fixed number of times in each iteration, disregarding the size of the population. This led to the unexpected effect that, for some settings, there was more change in larger populations than for smaller populations (actually due to a higher frequency of innovations in a larger population). However, this set-up of communication might not be realistic with larger population sizes: a more complex social structure and non-random communication between agents are therefore potentially interesting parameters.

Also, the simulations suggest that the coherency of the population is tightly linked to the amount of change that takes place in the population: the higher the amount of change, the smaller the coherency. However, it is difficult to compare the coherency of the agents in this model to that of actual language users in the real world, in order to see what coherency values are actually realistic. To formulate the same point more positively: here the development of the model clearly raises new questions that invite further theoretical and empirical research.

Summarizing, the verb krijgen can be said to be combining two dimensions: that of object controllability and that of verb agentivity. In this chapter, I have demonstrated how such a constellation leads to change (in the form of either extension or shift) in a particular direction. Although the model was specifically built to simulate krijgen, it is conceivable how this result can also be interpreted on a more general level. That is, any verb with the combination of verb agentivity and object controllability might show the same kind of development. I mention two examples suggesting that such a scenario is plausible. First, the English equivalent of krijgen, get, shows a development that is very similar to that of krijgen (Givon & Yang 1994, Gronemeyer 1999, Landsbergen 2006b). Second, the Dutch verb hebben ‘to have’ also has both dimensions and seems to have developed along similar lines, from the concrete, agentive meaning ‘to hold in one’s hand’ to the very general, non-agentive meaning ‘to have’ (WNT s.v. hebben).