Chapter 2

A cultural evolutionary model of patterns in semantic change

2.1 Introduction

When one takes language as a dynamic, continuously changing system, one of the striking features is that many changes do not seem to be arbitrary, but instead show at least some degree of regularity and directionality. Examples from semantics are tendencies such as “non-subjective > subjective > intersubjective” and “premodal > deontic > epistemic”, as described by Traugott & Dasher (2002). Related to this are the paths of grammaticalization as described in Heine, Claudi, & Hünnemeyer (1991), Bybee, Perkins & Pagliuca (1994), Heine & Kuteva (2002), Hopper & Traugott (2003). These paths describe tendencies in morphosyntactic change that are often accompanied by a semantic change and an increase in frequency. An example is the development of can (ABILITY > POSSIBILITY), in which can has changed from a full verb with lexical meaning (indicating the subject’s ability to perform some activity) to a modal auxiliary with a functional meaning (indicating the likelihood of some situation).

In this chapter, I try to explain such tendencies in semantic change by taking a cultural evolutionary perspective on language, and using an agent-based computer model of cultural evolution. After a general introduction to the behavior of the model, I focus on two concrete examples of tendencies in semantic change. First, I will discuss possible factors that affect (the amount of) change. Second, I will look at some factors that have a possible effect on directionality in changes of the kind “lexical meaning > functional meaning”. An example of such a change in English is the development of get (TO TAKE > PASSIVE), in which get has gradually lost its agentive meaning and has acquired a use as a marker of passive voice.

The workings of the model will be explained in section 3, after a general discussion in section 2 of possible mechanisms proposed in the literature for

---

1 This chapter is a rewritten version of Landsbergen, Frank, Robert Lachlan, Carel ten Cate & Arie Verhagen (in press). ‘A cultural-evolutionary model of patterns in semantic change’. 
producing directionality in semantic change. The results from the simulations are discussed in section 4, followed by conclusions.

2.2 Possible causes for asymmetries in semantic change

Directionality in semantic change and differences in likelihood to undergo such changes are examples of asymmetries in linguistic change. The division of the phenomenon of linguistic change in the two distinct processes of mutation and propagation leads to the question which of these processes is responsible for such asymmetries.

First, let us consider the asymmetry in likelihood of change. Words in particular constructions that have a general meaning in some conceptual domain, such as English *come* and *go* in the domain of movement, grammaticalize, while more specific movement words like *walk, stroll, saunter, swim, roll* and *slide* do not. There are different mechanisms that have been adduced as possible explanations for this asymmetry. First, a sole difference in contexts of use can be a cause (Bybee, Perkins & Pagliuca 1994: 5). Second, related to this but still an independent mechanism is the frequency of use, which is likely to be higher in words with a general meaning than words with a specific meaning. A third explanation is given by Traugott & Dasher, who state that ‘innovations can only be minimally different from earlier meanings’ (Traugott & Dasher 2002: 280). Relating this to words like *come* versus *stroll*, this could be taken to imply that words with a general meaning allow ‘larger’ semantic innovations than words with a specific meaning, because the relative change to the meaning of the words will be the same.

The second asymmetry concerns the direction of changes. In the literature on grammaticalization, directional change is often referred to as ‘unidirectionality’ to stress that changes take place in one direction (A > B) and not in the opposite one (*B > A). Such types of change could be caused by restrictions on the kinds of mutations people make. On the other hand, it may also be the case that any mutation can occur, but that there are constraints on the interaction between individuals which produces the spread; directionality might then result from certain types of mutations having a bigger chance of spreading than other ones, something that would require independent explanation.

In their work on semantic change, Traugott & Dasher (ibid.) focus on the restrictiveness of new mutations. They state that ‘[…] the path that the meaning of [a] form or construction takes is constrained by speakers’ tendency to recruit referential meanings to less referential functions of language’ (Traugott & Dasher ibid.: 86). In other words, speakers extend the meaning of existing referential words
by a small amount for a somewhat less referential function, and hearers pick this up from the speakers’ use of the words. Notice that in this view, these changes themselves take place in a certain direction, i.e. from more to less referential, because of speakers’ strategies. Although each change is small, large-scale unidirectionality is seen as the result of this kind of ‘directed mutation’.

This approach is similar to the ‘principle of the exploitation of old means for novel functions’ by Werner and Kaplan (1963), mentioned in Heine, Claudi & Hünnemeyer (1991: 28). According to this principle, ‘concrete objects are employed in order to understand, explain, or describe less concrete phenomena’, in which ‘concrete’ is associated with lexical items and less concrete with more functional items. Here too, the mutations themselves are directed.

Haspelmath (1999) uses Keller’s maxim of expressivity (Keller 1994: 101) to explain unidirectionality. Speakers are sometimes expressive, but in their need for expressiveness, their possibilities are limited. Haspelmath claims that in the lexicon-grammar continuum, speakers can only freely manipulate the lexical end of the continuum. This will lead to the use of a lexical item for a grammatical function, and not the other way around, since ‘functional elements cannot be used outside their proper places’ (Haspelmath 1999: 1059); so this is yet another candidate mechanism that is a case of directed mutation.

However, according to Haspelmath, this constraint on mutation is by itself not enough to cause language change. Once the mutation exists, it will spread because grammatical meanings are needed more often in language use than lexical meanings. By following Keller’s maxim of conformity (’talk like the others talk’), the use that speakers make of their language will cause this meaning to spread through the population. Thus, unidirectionality of large-scale change is explained by mechanisms in both mutation and propagation.

In summary, words with a general meaning show a stronger tendency to grammaticalize than words with a specific meaning, and this asymmetry has been explained by (i) their higher number of contexts of use, (ii) their higher frequency of use and (iii) their allowing larger mutations. The unidirectionality of the change from lexical to functional meaning has been explained by the factors mutation and frequency in different ways: (i) mutation only occurs in words with lexical meaning and not in words with functional meaning, (ii) the higher frequency of use of functional meanings.

The causal role of these different proposed factors can hardly (if at all) be investigated independently. Thus it would seem that it would be impossible to test claims such as, for example, that directed mutation is or is not necessary for unidirectional change, or that frequency of use by itself is or is not sufficient to produce unidirectionality. However, they can be studied independently, as well as in interaction, in computational models of cultural evolution.
2.3 The model

Theoretical background

Following the linguistic theory briefly mentioned in section 2, I take a usage-based approach to language change, in which individuals construct their linguistic knowledge on the basis of the input they receive in communication, in which actual utterances are the units of transmission and in which the locus of mutation is in adult communication (Bybee & Slobin 1982, Croft 2000, Croft & Cruse 2004, Slobin 2005). I assume word meaning to be essentially prototypical and polysemous (cf. Geeraerts 1997, Traugott & Dasher 2002), with words having multiple related senses. Semantic change is regarded as a change in this polysemy structure, in which new senses can be added to the established ones (possibly resulting in a shift of the prototypical meaning over time). Also, I assume a continuum of possible word meaning, reaching from lexical meaning on the one end to functional meaning on the other, rather than a sharp boundary between the two types of meaning.

Semantic mutations are small changes in the total set of existing senses, and are only minimally different. In principle, these may involve both small extensions and small contractions of an existing set of senses of an individual. These changes to the agents’ set of senses can spread through the population through communication. Extensions can introduce novel senses that may be picked up by the hearer, while contractions may spread because senses beyond a certain limit are used less. I will use the term ‘mutation’ over ‘innovation’ to stress the fact that any changes in the system are meant, and not just those that are intended and/or creative.

Properties of the model

I use a so-called ‘agent-based model’ of cultural evolution. The approach derives its name from the fact that it is a computer simulation of a group of individuals, or agents. The behavior of each agent can be independently controlled, and its effect on the population can be measured.

The first model is an extremely simple model, containing what are considered the bare necessities for semantic evolution. In order to investigate the various hypotheses to explain unidirectionality, I made a series of alterations to this simple model. These will be described in the subsequent sections.

The simple model I present here simulates the semantic evolution of a single random word $w$ in a population of speakers. The meaning of $w$ is represented by a set of senses, which represent concrete uses of $w$. These senses are positioned on a one-dimensional scale with a range of values between 0 and 1. Each value on this scale represents a specific sense of $w$ with nearby values representing similar
senses. The left end of the scale (with value 0) is arbitrarily chosen to represent lexical senses and the right end of the scale (with value 1) functional senses (figure 1).

![Figure 1. The one-dimensional semantic scale of the model.](image)

Consider the English word *while* in examples (1-3). In Middle English, *while* was used only as a noun with the meaning ‘short period of time’, as in (1). Later it came to be used in adverbial phrases with the meaning ‘during the time’, as in (2). In present day English, *while* has become a marker of co-temporality, as in (3).

1) Whether he lyf lang or short *while*. (1340)
2) I for both have wept when all my tears were bloud, the *while* you slept. (1633)
3) Mr. Montgomerie said rather gallant things to me, … *while* the girls looked shocked. (1908)

In the model, the concrete senses from examples (1-3) may be thought of as represented on the 0-1 scale shown in figure 2. The meaning of the lexical noun *while* is situated at the left end of the scale, and the meaning of the functional marker *while* towards the right end (note that the exact placement of the examples on the scale is arbitrary, and only intended to serve as an example).

![Figure 2. Possible positions of different senses of while on the one-dimensional scale.](image)

---

2 Examples 1-3 are taken from the *Oxford English Dictionary* (1989).
Because words have multiple senses, i.e. they are polysemous, the total meaning of word $w$ is represented in the model by a continuous set of senses instead of a single sense as in figure 2. On the 0-1 scale, this set can occur on different positions and in different sizes. The position of the set is an indication of the grammatical status of word $w$: if the set is positioned towards the left side of the scale, the word’s meaning is mainly lexical. By contrast, if the set is positioned towards the right side of the scale, the word’s meaning is mainly functional. Apart from this, the set can also differ in size. Since one value on the scale represents a single sense of the word $w$, the size of the set of senses is a measure of the range of the word’s meaning: the wider a set, the more senses it can be used in. I will refer to this as the generality or specificity of word $w$. Note that this means that both lexical and functional meaning can be specific as well as general. In this model, the number of senses determines specificity or generality, not the position: a functional meaning can be specific in meaning in that it can only be used in a very limited number of (functional) senses, while a lexical meaning can be general in meaning by having a wide range of (lexical) senses. Figure 3 shows some examples.

An agent’s knowledge of word $w$ is simply a continuous set of senses on the [0-1] scale. This set is defined by the values of the lower and upper limit: any value between these limits (excluding the limits themselves) is part of the meaning of $w$ in an agent’s knowledge. For example, if an agent’s limits are [0.559, 0.782], his knowledge includes all values between these limits, such as 0.55980, 0.62 and 0.78195.

Agents construct their linguistic knowledge on the basis of input they receive during communication. Communication in the model is the random selection
of two agents from the population, one of which is assigned the role of speaker and one the role of hearer. The speaker selects a specific sense \( s_i \) (represented by a value) from its\(^3\) set of senses and transmits it to the hearer. This models the evaluation by the speaker that the word \( w \) is applicable in the specific context, given the set of senses of \( w \) that the speaker knows. The hearer compares the transmitted sense to its own set of senses, i.e. it evaluates whether the word \( w \) is applicable in the context, given its set of senses of \( w \). When this sense is already part of the hearer’s knowledge of \( w \), communication is successful and the communication process comes to an end. However, the speaker can also transmit a sense that is unknown to the hearer, i.e. that is outside the hearer’s range of senses associated with \( w \). In that case, communication fails, and the fact of this failure is understood by not only the hearer, but also the speaker. Although such direct feedback might be considered unrealistic, it is plausible to assume that speakers do obtain clues about the successfulness of their utterance, e.g. from the failure to achieve a communicative goal. As such, the feedback in this model is a simplification of this process (compare the ‘language games’ in Steels 1998 and De Boer 2001).

Unsuccessful communication results in a learning process, in which both agents adjust their sets of senses of \( w \). The hearer, confronted with a new sense, makes the inference that \( s_i \) must be an appropriate sense, since the speaker used it, and will increase its set to include \( s_i \). The speaker, confronted with unsuccessful communication, realizes that any values beyond the uttered sense will lead to more unsuccessful communication and therefore decreases its set and changes it to exclude values beyond \( s_i \). However, I mentioned earlier that the set’s limits are not part of the agent’s knowledge. To ensure that an uttered sense will be the actual new ‘active’ limit of the agent’s knowledge, \( s_i \) is increased or decreased with a very small value. This value is set at \( 1 \cdot 10^{-6} \). As a result, speaker and hearer end up with similar limits that just include the uttered sense \( s_i \). Figure 4 on the next page illustrates this process. An agent’s set can never become bigger than 1, or smaller than \( 1 \cdot 10^{-5} \).

Apart from the learning process described above, agents also change their linguistic knowledge by mutation. Mutation in the model is a randomly occurring small change in set size. Agents that are selected for communication have a probability \( m \) to undergo mutation before that communication event. Mutations may be extensions or constrictions on either side of the set. In linguistic reality, possible causes for the former include the need to express something for which there is not yet a signal, and for the latter the need to redress ‘semantic overextension’ or competition by another word. However, in this model I do not directly address the nature of the causes of mutations, and simply start from the assumption that they occur; but I will model certain properties of mutations that were discussed in the

\(^3\) In this chapter and subsequent ones, I will refer to an agent in a simulation as ‘it’.
previous section, such as their size and their likelihood to occur in words with a more lexical or a more functional meaning.

The size of a mutation is based on the value of parameter $m_s$. During a particular simulation, $m_s$ will have a fixed value, but the actual size of a mutation in an agent’s knowledge can deviate from this value. However, the larger the deviation, the smaller the probability of its occurrence. The equation used for this probability is shown in equation 1.

$$prob(m = x) = m_s \cdot e^{-\frac{x^2}{2}}$$

**Equation 1.** The size of a mutation is determined by a Gaussian function with a standard deviation of $m_s$.

The model is iterated in 500 cycles called ‘years’. These years are defined in relation to the age and replacement of agents in the population, and the amount of communications between them. In such a communication, two agents are randomly selected from the population as speaker and hearer, and each year, each agent participates in 500 communications on average. This frequency is the usage frequency $f$ of word $w$. Each simulation is initiated with agents having a random age.
between [0, 70] and a set of senses with a fixed size of 0.2 and a variable minimum of $S_{\text{min}} = 0.4 + d$, with $d$ a random value $d \in [0, 0.2]$.

The population consists of 100 agents. This value is an intermediate value of those suggested by Dunbar (1998) and Milroy & Milroy (1992) for the size of linguistic communities. In the standard model, the population is considered to be a linguistic unity: there is random communication between all 100 agents in the population. However, I do investigate the effect of non-random communication within the population on the overall linguistic coherency in the next section. It seems realistic to assume that as populations get larger, they will start to become divided into several (socially based) subgroups, within which agents communicate randomly, but between which there is less frequent communication (cf. the notion of ‘social networks’ in sociolinguistic theory, e.g. Milroy & Milroy 1992). I have simulated such a structure by dividing the total population into a number of subgroups and limit communication between individuals from different subgroups. The probability of communicating with an agent from another subgroup is given by factor $g$. For example, if $g = 0.01$, there is a 1 percent chance that agents will communicate with agents from another subgroup, while there is a 99 percent chance of communicating with an agent from the same subgroup. Note that with $g = 1$, there is random communication between all 100 agents.

Agents have a maximum age of 70 years, after which they are replaced by an agent with age 0. Newborn agents start with an exact copy of the set of senses of a randomly assigned ‘parent’, after which they participate fully in the communication between agents. Note that this ‘parent’ is not the agent that is being replaced (because in such a case there would be no need to add generations in the model). Rather, the transmission of the parent knowledge is a simplification of the acquisition process. This means that any evolution displayed by the model is not due to imperfect learning situations in child language acquisition, but to variation coming about and spreading in adults; in this way it is possible to test whether such variation can by itself lead to semantic change. Note that this does not mean that transmission in the model is completely horizontal (i.e. within peer groups only); communication is random between all agents regardless of their age, and therefore transmission can be said to be both horizontal and oblique (Cavalli-Sforza & Feldman 1981).

In section 2, I introduced several factors that have been proposed by linguists to play a role in the (uni)directionality of the change “lexical meaning > functional meaning”. These factors can now be linked to the model. Frequency is the number of times agents communicate with each other each year ($f$). Mutation is the change in set size, with a rate of $m_t$ and a size $m_s$. The set size is an indication of the generality or specificity in meaning, and the position of the set on the scale can be taken as an indication of the lexical or functional status of the word. In the next
section, I will first discuss general properties of the most simple model, and then the effects of these factors on simulations of semantic change, and the way they relate to asymmetries in such changes.

### 2.4 Results

**General behavior of the model**

In the standard model described in the previous section, it is found that stable, coherent meanings for word \( w \) develop within populations, and then gradually change over time. The simulations show slightly different behaviors each time they are run, with fluctuations in the average meaning size as the result: specialization and generalization both occur. Basically, the simulations exhibit random drift in the direction of both the upper and lower limit of the meaning set. With meanings drifting in both directions along the scale, there is evolution, but no unidirectionality. In most cases, the coherency of populations remains high regardless of the amount of drift. Nevertheless, over 500 years, the meanings of \( w \) in different populations can diverge to the extent that communication between them would be seriously limited. This can be seen in figure 5 on the opposite page, which shows the average meaning sets of 10 populations after a simulation of 500 years: while some of these average sets have relatively similar values (e.g. those of populations 1 and 8), most of these sets have very different upper and lower limits. For example, the sets of populations 5 and 8 only show a relatively small amount of overlap.

When focusing on the coherency of the population, note that there is no direct transmission of the ‘total’ meaning of \( w \) between agents; they are only exposed to single senses in utterances, and shape their meaning of the word on the basis of this information. The model shows that such indirect transmission does not lead to an incoherent population, when certain conditions are met.

I have tested the effect of three factors on this coherency: mutation rate, frequency of use and population structure. Coherency (represented by \( \theta \)) was measured as the average amount of overlap of the sets of senses between agents in that population. This is done by first calculating the overlap of the sets of senses of each pair of agents in the population and then calculating the average of all these values (equation 2 on the opposite page). The greater this overlap, the greater the consensus about the meaning of word \( w \).
First, the mutation rate in the population should not be too high. A certain amount of communication is needed for a single mutation to spread through the entire population and to even out the emerged variation between the agents. When the number of communications relative to the mutation rate becomes too low, the individual variation caused by mutation is not transmitted to other individuals often enough, thus causing a lower coherency (figure 6a on the next page). Changes in frequency do not affect the coherency of the population significantly (figure 6b, also on the next page). This is due to the fact that, in this model, the rate of mutation is linked to the frequency of use, and therefore the number of communications relative to the mutation rate remains the same.

Second, the population structure involves random communication between all agents. This might be realistic for small groups (of $N = 100$), but not when populations are much larger. In the latter case it seems more realistic to assume a population divided into several (socially based) subgroups. As I mentioned in the previous section, I have simulated such a structure by dividing the total population
into a number of subgroups and limit communication between individuals from different subgroups and the probability of communicating with an agent from another subgroup is given by factor $g$.

Not surprisingly, the less communication there is between the subgroups of the total population, the less coherent this population becomes. However, only a very limited amount of between-group communication ($g = 0.01$) is needed to create considerable coherency in the total population. This can be seen on figure 7.
In summary, populations are basically coherent unless there is a great deal of mutation or virtually no communication between groups of agents. At the same time, word meaning gradually evolves within populations over time. Therefore, the model, simple as it is, behaves in a linguistically realistic way, and demonstrates the benefits of a cultural evolutionary approach to language change. Recently, this cultural evolutionary analogy has been the topic of dispute (cf. contributions in Aunger 2000). Sperber (2000) questions the value of the analogy by stating that there is a fundamental difference between genes and memes, the units of cultural inheritance. Genetic replication involves copying, but in the replication of cultural items (such as linguistic structure), copying ‘will only ever be a small proportion of cultural learning. It is but the limiting case of a much more complex process involving multiple steps of inferencing’ (Aunger 2000: 19). However, the results of the model in this chapter seem to indicate that the cultural evolutionary approach can also be valuable when replication involves more than simple copying and requires multiple steps of inferencing on the language user’s part. Individuals must generate their own concept of the range of contexts in which a word can be used, yet they are not directly informed of the limits of the ranges used by others, and they do not hear all of the possible contexts. Nevertheless, the ranges of contexts can be maintained very conservatively throughout the population over time. This is a consequence of the mode of transmission, which might be best characterized as

---

4 As such, the question whether the replicator is a mental representation (as is claimed by Sperber 2000) or a token of linguistic structure in an utterance (as Croft 2000 claims and which is also the view adopted in this study) becomes less urgent, since the nature of the replicator does not immediately affect the inferencing that takes place when linguistic structure is passed on in communication.
“many-to-one” (cf. Cavalli-Sforza & Feldman 1981), combined with the ability to generalize to fill in the gaps.

In the next section, I will take a closer look on this behavior with regard to change, and discuss factors affecting patterns of change.

Factors affecting the rate of semantic change

As I described in section 2, a major restriction on grammaticalization seems to be that words are not equally liable to grammaticalize: words with a more general meaning do, while more specific words do not. Three possible explanations for this relationship were discussed: words with a general meaning are applicable in a wider range of contexts (factor 1), they will have a higher frequency (factor 2) and they allow wider mutations (factor 3). As to the third factor, recall that the size of an individual semantic mutation in the model is typically rather small, and is determined by a Gaussian function with a standard deviation ($m_i$). However, it is conceivable that different meanings allow different sizes for one-step extensions; if so, then it is natural to assume that general meanings will allow larger extensions than specific meanings, rather than the other way around. I have carried out a series of simulations in order to test the feasibility of these three explanations.

The effect of factor 1, the number of contexts, was simulated by initiating different populations with different sizes of the meaning sets. As I explained in section 3, a small meaning set represents a limited number of contexts a word can be used in, and therefore represents words with a specific meaning. A large meaning set, i.e. a wide range of senses, represents a word with a general meaning. Frequency of use, mutation rate and size were kept constant.

As a measure for the liability to change, I take the amount of change in the position of the meaning set on the 0-1 scale after 500 years. In this case, the change involved is ‘drift’, because no selection pressures have been added to the model, and therefore changes can occur in any direction. The amount of drift $\partial$ is measured as the distance between different groups after 500 years, since they start off in a similar position in the ‘space’ of possible senses. This distance can be calculated by comparing the average middle values of the different groups. For example, an agent with a set of senses [0.20, 0.46] has a middle value of 0.33, and the average middle of an entire group of agents can be calculated by taking the sum of all these middle values and by dividing them with the number of agents in the group. Subsequently, the average middle value of all groups can be calculated. Finally, by comparing this value to the average middle value of each group, one gets a measure of the distance between the different groups. This is shown in equations 3-5: equation 3 shows the calculation of $d$, the average middle of one group, equation 4 shows the calculation of $d_{av}$, the average middle of all groups, and equation 5 shows the calculation of the
standard deviation between \( d \) and \( d_{av} \), indicating drift \( \partial \). The larger the standard deviation, the greater the value for \( \partial \).

\[
d = \frac{1}{N} \cdot \sum_{i=1}^{N} (S_{\min,i} + ((S_{\max,i} - S_{\min,i})/2))
\]

Equation 3.

\[
d_{av} = \frac{1}{MN} \cdot \sum_{i=1}^{MN} (S_{\min,i} + ((S_{\max,i} - S_{\min,i})/2))
\]

Equation 4.

\[
\delta = \sqrt{\frac{1}{M-N} \cdot \sum_{j=1}^{M} (d_{j} - d_{av})^2}
\]

with \( M \) groups of \( N \) agents.

Equation 5.

Figure 8. The amount of drift for increasing set sizes. The y-axis shows the average amount of drift over 20 groups. \( N = 100, t = 500 \) years, \( f = 500, m = 0.01, m_i = 0.01 \).

The results of this simulation are shown in figure 8. As the figure shows, there are some fluctuations in the amount of drift found, but there is no clear link between set
size and drift. This seems to indicate that a word’s generality in meaning per se does not make the word more liable to change (and thus grammaticalization).

The set sizes themselves also change over time. There are large fluctuations in the final set sizes for each parameter setting: both specialization and generalization occur, which leads to averages that are almost equal to the initial set sizes (figure 9). However, the smaller initial set sizes (≤ 0.08) behave somewhat differently, showing more generalization than specialization, and therefore showing an average increase in set size over time. This effect is probably due to the way the model is constructed. First, there is a minimal set size for each agent (1 \cdot 10^{-5}). Whenever learning or mutation in the form of contractions leads to a zero set size, the agent is assigned this minimal value instead. Because such a limitation is absent in the case of an extension, this will lead to a slight ‘benefit’ for extensions versus contractions. Second, the variance in set size among agents will be smaller when set sizes are relatively small, and in such an environment, extensions spread more easily than contractions. As figure 9 shows, an average set size of ~ 0.1 seems to be a ‘minimal’ set size for the populations with the current parameter settings.

![Figure 9](image)

**Figure 9.** Final set sizes after 500 years for 20 groups, for different initial set sizes. Apart from the averages (in squares), the minimal and maximal values are shown to indicate the large amount of variance between groups. N = 100, f = 500, m = 0.01, m_s = 0.01.

The next step is to consider the effect of frequency on change in the model (factor 2), by looking at the effects of different frequencies of use, simulated by manipulating the number of communications per year. In these simulations, the initial set size was kept equal in all cases (width = 0.2), as was the amount of

---

5 Limitations to extensions do exist when set sizes would become larger than 1. However, such large set sizes do not occur in the settings used in these simulations.
mutation ($m_r = 0.01$). In order to use somewhat realistic relative differences in frequency, I have used the relative frequencies of some English movement verbs (mentioned as an example in Bybee, Perkins & Pagliuca 1994: 5) as a basis. These are shown in table 1. The frequency of the very general word *come* is more than 7 times higher than the more specific *walk*, and almost 60 times higher than *swim*. I have used similar orders of magnitude in the simulation ($f = \{10, 100, 1000, 10000\}$).

<table>
<thead>
<tr>
<th></th>
<th>frequency (per million words)</th>
<th>frequency relative to <em>come</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>come</td>
<td>1512</td>
<td>-</td>
</tr>
<tr>
<td>walk</td>
<td>215</td>
<td>0.142</td>
</tr>
<tr>
<td>roll</td>
<td>49</td>
<td>0.032</td>
</tr>
<tr>
<td>side</td>
<td>30</td>
<td>0.020</td>
</tr>
<tr>
<td>swim</td>
<td>25</td>
<td>0.017</td>
</tr>
</tbody>
</table>


Figure 10 shows that the amount of drift is strongly correlated with the frequency of use of word $w$: high frequency words show more drift than low frequency words, as a result of a difference in frequency alone. This is an indication that frequency of use is an important factor in the differences in likelihood to change (hence also to grammaticalize) that are found between these types of words.

![Figure 10. The role of frequency of use in determining the rate of semantic change. The y-axis shows the average amount of drift between 20 groups for increasing frequencies of use. $N = 100$, $t = 500$ years, $m_s = 0.01$, $m_r = 0.01$.](image-url)
To test the effect of the third factor, mutation, I have looked at different values for mutation size $m_s$ and measured its effect on the amount of drift. Both set size and frequency of use were kept constant at 0.2 and $f = 500$ per year, respectively. The mutation rate was kept constant at $m_r = 0.01$, but I varied the mutation size between $m_s = 0.00001$ and $m_s = 0.01$ to test its effect on the amount of drift. The results in figure 11 show a strong increase of the amount of drift for increasing mutation sizes.

![Figure 11](image)

Figure 11. The role of mutation in determining the rate of semantic change. The y-axis shows the average amount of drift between 20 groups for increasing amounts of mutation. $N = 100$, $t = 500$ years, $f = 500$, $m = 0.01$.

This suggests that, if indeed words with a general meaning allow for wider extensions than words with a specific meaning, this leads to a higher amount of drift in the former than in the latter type.

When comparing the results of the three factors in the model, it shows that pure differences in the ‘size’ of meaning do not cause differences in the rate of drift by themselves. It is possible that the generality of a word influences the rate of semantic evolution through some other correlated effect that was not included in the model. On the other hand, both differences in frequency and differences in the size of mutations are factors that do lead to a dramatic difference in drift, even if they occur alone. In the basic model, the average knowledge of the populations shifts over time as a result of drift. An increase in either frequency of use or size of mutations allows for larger amounts of drift over the one-dimensional space, without any extra forces added to the model.

All changes described above were non-directional, which is why I have referred to them as ‘drift’. In the next section I will discuss directionality in change.
Factors causing unidirectional semantic change

In section 2, two different kinds of possible explanations were given for unidirectionality in change: first, speakers may only be able to freely manipulate lexical meanings of a word and second, functional meanings are used more frequently than specific, lexical meanings. Haspelmath (1999) argues that the combination of both factors leads to a unidirectional change from lexical meaning to functional meaning.

I tested these two factors in the following way in the model. The first hypothesis is equivalent to an asymmetry in mutation: words with lexical meaning can be adapted to express functional meaning, but not the other way around. To simulate this difference, the mutation rate was kept constant at $m_r = 0.05$, but the probability of the direction of mutations was varied with a parameter $p_m$.

The second hypothesis concerns an asymmetry in the frequency of use: senses with a functional meaning have a higher chance of being used in communication than senses with a lexical meaning. Individuals must select a sense of $w$ for communication from within their set of meanings, but here I varied how likely they were to pick different senses from within that meaning. In all simulations up to this point, individuals picked a sense according to a uniform random distribution. In the present set of simulations, senses were picked according to an exponential distribution. In this type of distribution, the probability of selecting a certain sense increases with increasing sense values. The strength of this increase can be altered with a parameter $p_s$. For example, if $p_s = 2$, the probability of an agent selecting $s = 1$ is twice as big as selecting $s = 0$ (provided the agent has both senses in its set of meanings), while with $p_s = 100$, the difference in probability is 100 (equation 6).

$$\text{prob}(s = x) = p_s^x$$

Equation 6.

In a first simulation, I combined both factors, asymmetry in mutation and asymmetry in frequency of use. For the asymmetry in mutation, I used $p_m = 0.55$, which meant a probability of 0.55 for mutations to occur on the functional end of the agent’s set (and a probability of 0.45 for mutations to occur on the lexical end of the set). Because this is only a small asymmetry, the model is more conservative than the hypothesis. For the asymmetry in frequency of use, I used $p_s = 2$, which also leads to a weak preference for more functional meanings. Two simulations were run, one in which the agents in the population started with a lexical meaning (with an average set of [0.1 - 0.3]), and one in which they started with a functional meaning.
The latter was added to see whether change in the opposite direction, i.e. from functional to lexical meaning was possible.

Figure 12 is a representation of the average sets of senses in 20 unrelated groups. In this figure, the average middle of each group is shown twice, at the beginning of the simulation (as a grey circle) and after 500 years (as a black square). Its values are plotted on the vertical scale. I will use this representation in subsequent figures as well. In figure 12a, it can be seen that all groups start off with similar starting average middle values of 0.2 (grey circles), while there is some variation in their average middle values after 500 years (black squares).

Both factors combined indeed create a selection pressure that drives the average set of senses of a population from the lexical side of the spectrum to the functional side, even if both factors are weak (figure 12a). Also, the selection pressure blocks any change in the opposite direction (figure 12b). In this simulation, the starting middle value of all groups was 0.8 (indicated by the grey circles), but the average values after 500 years of all groups (indicated by the black squares) do no differ significantly from these starting values.

As a comparison, a simulation was run in which neither factor was operative. As figure 13 shows, changes from lexical to functional meaning can occur as a result of random drift, but changes in the opposite direction are also possible, and no strict unidirectional change is taking place.

**Figures 12-13.** (12) The effect of asymmetries in mutation and frequency of use together, and (13) the effect of a random drift simulation. The results are shown for 20 unrelated groups (x-axis). The y-axis shows the average position of each group on the 0-1 scale. Grey circles indicate the average starting position of each group, black squares end positions after \( t = 500 \) years. \( f = 500, m = 0.05, m_a = 0.05, p_m = 0.55, p_s = 2 \). Left: lexical starting point; right: functional starting point.
However, the question remains whether the two factors could also cause unidirectional change when they operate alone. I first tested whether the asymmetry in frequency of use by itself can create a sufficient selection pressure. As figure 14 shows, this is the case. The same value for $p_s$ as previous but now acting by itself creates a selection pressure for functional meaning, thus causing unidirectionality in change.

The next question was whether the small asymmetry in mutation ($p_m = 0.55$) would have a similar effect on the direction of change. It turns out that this factor creates a much weaker selection pressure when compared to the previous factor (frequency). Although change from lexical meaning to functional meaning takes place, changes in the opposite direction are not completely blocked (figure 15).

![Figure 14](image1.png)

**Figure 14.** The effect of an asymmetry in frequency of use, using $p = 2$, $f = 500$, $m = 0.05$, $m_r = 0.05$, $p_m = 0.50$. Left: lexical starting point; right: functional starting point.

![Figure 15-16](image2.png)

**Figures 15-16.** The effect of two different asymmetries in mutation. $f = 500$, $m = 0.05$, $m_r = 0.05$, $p_m = 0.55$ (Fig. 15), $p_m = 0.95$ (Fig. 16). Left: lexical starting point; right: functional starting point.
However, the force proposed by Haspelmath (1999) and discussed in section 2 may be a very strong one; his formulation suggests an absolute constraint: ‘functional elements cannot be used outside their proper places’ (Haspelmath 1999: 1059). In the model using $p_m = 0.55$, mutations from functional to lexical still occur in 45 percent of the cases, so I tested a much stronger asymmetry of $p_m = 0.95$, which indeed proves to be a very strong pressure for unidirectional change (figure 16).

While I have compared the effects of some, more or less randomly chosen parameter settings for both factors, their effect can be compared more precisely by measuring the amount of change that takes place. For this, I measured the average distance between the initial meaning and the meaning after 500 years of all agents in the population. This value can be taken as a measure for the strength of the unidirectional change.

Figure 17 is a combination chart that shows the average distance for different settings of both an asymmetry in mutation ($p_m$) and in frequency of use ($p_s$). Note that $p_m = 0.50$ and $p_s = 1$ are simulations with random drift. In these simulations, a variable amount of change takes place in the 20 groups as a result of mutation, and these variable amounts results in an average change of around 0.30.

As expected, larger values of both $p_m$ and $p_s$ lead to a higher amount change. A larger $p_m$, giving a higher probability for mutation on the functional end of an agent’s set, leads to more (unidirectional) change, and for $p_s$, the strength of the bias for expressing functional meaning, we see a similar pattern. For $p_s$, the

![Figure 17](image-url)
amount of change quickly increases with $1 < p_s < 1.5$. However, the increase seems to halt around 0.60 while this is not the case for greater values of $p_m$. This turns out to be an artifact of the way the model was constructed. The meaning scale in the model is limited to $[0, 1]$, and when a possible (extending) mutation exceeds one of these limits, another mutation is tried. This means that the closer an agent’s set comes to one of the limits, the higher the chance a contracting mutation will be created. In turn, this will lead to smaller average set sizes, and these smaller set sizes can move closer to one of the limits than larger set sizes.

These results seem to indicate that asymmetries in both mutation and frequency might not have to be working together to create a unidirectional pressure. Small asymmetries in frequency and somewhat larger asymmetries in mutation already lead to clear unidirectional change in the model. However, as noticed above, a large asymmetry in mutation requires a fairly strict distinction between lexical and functional meanings, and this may be at odds with the generally observed gradualness of semantic change, including shifts from lexical to functional (Hopper & Traugott 2003); it may therefore be considered a relatively implausible cause of unidirectionality on its own. In this respect, it is of course interesting that the model shows that the elementary mechanism of a small difference in frequency is powerful enough to cause unidirectionality by itself.

2.5 Discussion and conclusions

The results demonstrate how a cultural evolutionary perspective may be of use in making sense of linguistic hypotheses about language change. I have given a concrete example of this by investigating the phenomena of unidirectionality in the change from lexical to functional meaning and differences in the liability of words to grammaticalize. By using cultural evolutionary simulations it was possible to study several hypotheses independently, something that is more difficult to do with empirical data, where all of the different factors that were investigated may be operating at the same time. By deliberately keeping the model simple, it was possible to elucidate the mechanisms underlying some of these hypotheses. The model presented here shows that with indirect transmission alone (individuals inferred the overall meaning of a word from multiple instances of hearing that word being used), it is possible to maintain a linguistically coherent population, provided that there is sufficient communication between agents, and the mutation rate is not too high. Whether this is in fact the case is an interesting topic for future research, and this is directly linked to the question what order of magnitude the general mutation rate in language actually has.
As for the rate at which semantic change takes place, it seems that the generality of a word’s meaning does not have a direct effect: I found no effect on the amount of change when frequency of use and mutation rate are kept at a constant level. Rather, linguistic consequences of generality (I investigated two examples: a higher frequency of use and a greater ease with which individuals can use a word in new contexts) are more likely to be the direct causes of higher rates of change. Both these factors are mentioned in earlier studies (Bybee, Perkins & Pagliuca 1994, Traugott & Dasher 2002) as possible causes and the results confirm their hypotheses. The results also show that both frequency of use and mutation size have similar effects when operating without the other, and that therefore they do not have to occur together.

A similar point can be made on the basis of the results regarding directionality in change. The results seem to confirm that unidirectionality in semantic change can be understood as a result of different usage properties of words with a lexical meaning versus those with a functional meaning. On the one hand, the fact that functional meanings are more general and abstract and can therefore be used in more contexts than lexical meanings is by itself a force that produces unidirectionality. On the other hand, the relative ease with which lexical meaning can be manipulated, that is, with which mutations can take place, acts as a force for unidirectionality as well. These findings are in accordance with the predictions made by Haspelmath (1999). However, the results of the simulations suggest that each of the two factors alone may already lead to unidirectionality, with only relatively small asymmetries in either mutation or frequency of use.

Several factors that are often associated with change in general and grammaticalization in particular were not included in the model. First, the knowledge of agents was restricted to a continuous meaning set. Old meanings can only remain in use next to newly developed meanings when all the meanings in-between remain in use as well, and this may not always be the case. Second, the notion of entrenchment, connected to the frequency of use of particular senses (e.g. Langacker 1987: 59), was not included in the model either. Related to this, the frequency of use of the word as a whole was constant in the model, while a shift towards functional meaning usually implies an increase in absolute frequency of use. Third, the model simulated words in isolation, disregarding influences from factors such as context, which is argued to play a major role in the grammaticalization process. Lastly, of the non-linguistic factors, I only investigated the effect of the coherency of the population, while it is possible that other factors such as population size and time (as in duration of the simulation) could affect change as well. I do not claim that these factors do not play a role in the process of linguistic change, but instead hope to have shown that the presence of such more
complicated factors is not a necessary condition for basic types of semantic changes that are known from historical linguistics to occur.

Perhaps the most striking result emerging from the simulations presented in this chapter, is that the very basic, ‘mechanistic’ factor of frequency of use is a recurrent, dominant factor producing regularities in semantic change, even independently of (considerations about) other signals: the model I have used here contains only one word so that issues of competition and relative frequency do not enter into the picture here. Nevertheless, I have been able to reproduce some general properties of processes and products of semantic change, and to indicate more or less plausible factors producing specific patterns of change.