Erratum
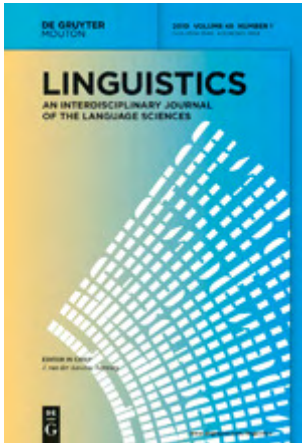
Frank Landsbergen, Robert Lachlan, Carel ten Cate, and Arie Verhagen:
A cultural evolutionary model of patterns in semantic change (pp. 363–390)

Only the first two named authors were the equal and principal contributors to the conception and implementation of the reported work and the writing of the manuscript and not all authors as previously indicated in Note 1.

A cultural evolutionary model of patterns in semantic change[1]

FRANK LANDSBERGEN, ROBERT LACHLAN,
CAREL TEN CATE, AND ARIE VERHAGEN

Abstract

Language change has been described as an unintended effect of language in use (Keller 1994). In this view, change results from the way individuals use their language; the challenge is thus to explain change and its properties in terms of factors operating on the individual level, and population dynamics. An intriguing example of such a phenomenon is the finding that language change shows some highly regular tendencies. This has recently received considerable attention in the literature (Bybee et al. 1994; Heine and Kuteva 2002; Traugott and Dasher 2002; Hopper and Traugott 2003). In unrelated languages, similar words often change in similar ways, along similar "trajectories" of development. This phenomenon is called "unidirectionality", and it is an important part of processes of grammaticalization, items changing from a lexical meaning to a grammatical function. It has been claimed that around 90–99% of all processes of grammaticalization are unidirectional (Haspelmath 1999).

This article explores several mechanisms that may lead to language change, and examines whether they may be responsible for unidirectionality. We use a cultural evolutionary computational model with which the effects of individual behavior on the group level can be measured. By using this approach, regularities in semantic change can be explained in terms of very basic mechanisms and aspects of language use such as the frequency with which particular linguistic items are used. One example is that frequency differences by themselves are a strong enough force for causing unidirectionality. We argue that adopting a cultural evolutionary approach may be useful in the study of language change.

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 "nonsubjective > subjective > intersubjective" and "premodal > deontic > epistemic", as described by Traugott and Dasher (2002). Related to this are the paths of grammaticalization as described in Heine et al. (1991), Bybee et al. (1994), Heine and Kuteva (2002), Hopper and 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 article, we 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. The advantage of using a cultural evolutionary approach is that it models patterns in complex systems such as the language of a population as a result of the interactions between individuals, and by population-level processes of selection and random drift. This approach is not new in itself: Keller (1994) proposes an approach of this kind in his "invisible hand theory". In his model, group level phenomena are to be reduced to individual behavior, and the notions of mutation on the one hand and spread or propagation of new variants on the other are distinguished. He states that language change is the unintended result of intentional individual behavior. Individuals apply strategies or "maxims" when they use their language, such as "speak in such a way that you are communicatively successful" and "talk like the others talk". Although most of these maxims lead to the creation of linguistic conventions, some maxims can — unintentionally — lead to change, such as "talk in such a way that you are noticed" and "talk in such a way that you do not spend superfluous energy". Haspelmath (1999) uses Keller's model to explain unidirectionality in grammaticalization.

Croft (2000) gives a conceptual model of language change as cultural evolution. He explicitly states that actual utterances are the units of cultural transmission, a view that we adopt in this article. He also differentiates between linguistic factors and social factors. Linguistic or cognitive factors give rise to new variants that come into existence by both intentional and unintentional mechanisms such as reanalysis, creativity and economy. Whether these mutations spread through a population depends on social factors, such as the structure of the population and the prestige of speakers.

The concept underlying Croft and Keller's linguistic theories is that of cultural evolution, an analogy with genetic evolution. The transmission of cultural traits such as ideas, traditions, values, patterns of behavior, and language occurs when one individual learns from another, not by the transmission of genetic material. The difference in the mechanism of transmission does not alter the fact that culture in this sense forms a system of information that is inherited by new generations from the previous one. Apart from this analogy with the transmission of genes during reproduction, we find analogues to mutation in the fact that copying errors might occur, or that traits can be created or altered by individuals. Again, the difference in mechanisms does not alter the fact that cultural traits can change. The existence of both mutation and reproduction results in a system of variation and inheritance, which can lead to evolutionary change in combination with drift or by selection. Mechanisms of cultural evolution in this sense have been described in detail by Cavalli-Sforza and Feldman (1981) and Boyd and Richerson (1985). In linguistics, the idea of cultural evolution is adopted in a general way in a number of recent studies (e.g., Croft 2000; Tomasello 2003), but we know of no detailed studies applying cultural evolutionary models to language change.

Agent-based computer models of language change have been used in other studies before, such as Niyogi and Berwick (1997) and Yang (2000). These models simulate syntactic change as a result of imperfect learning. A different approach, which is more comparable to the model presented in this article, is the use of "language games". In these models, change is the result of communication and adjustment of the agents' knowledge based on the input they receive, and imperfect learning plays no role (e.g., de Boer 2001). In general, most of the studies using computer simulations so far have focused on syntax and phonetics. As Steels (2003) has pointed out, grammaticalization and unidirectionality have received less attention in this line of research.

In this article, we introduce a cultural evolutionary model of semantic change. After a general introduction to the behavior of the model, we focus on two concrete examples of tendencies in semantic change. First, we discuss factors that affect (the amount of) change. Second, we look at 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 our model will be explained in Section 3, after a general discussion in Section 2 of possible mechanisms envisaged in the literature, for producing directionality in semantic change. The results from our simulations are discussed in Section 4, followed by conclusions.

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 et al. 1994: 5). Second, related to this but still an indepedent 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 and Dasher, who state that "innovations can only be minimally different from earlier meanings" (Traugott and Dasher 2002: 280). Relating this to words like *come* versus *stroll*, significant semantic innovations may be more easily accepted in words with a general meaning because, relative to the range of meanings the word already has, the innovations appear less disruptive than they would for a word with a more specific meaning.

The second asymmetry concerns the direction of changes. In the literature on grammaticalizaton, 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 and Dasher (2002) 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 and Dasher 2002: 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 et al. (1991). According to this principle, "concrete objects are employed in order to understand, explain, or describe less concrete phenomena" (Heine et al 1991: 28), 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 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, which is precisely why we chose these as a research tool in this article.

3.   The model

3.1.   *Theoretical background*

Following the linguistic theory briefly mentioned in Section 2, we 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 and Slobin 1982; Croft 2000; Croft and Cruse 2004; Slobin 2005). Also, we take word meaning to be essentially prototypical and polysemous (cf. Geeraerts 1997; Traugott and 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, we 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. We 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.

3.2.   *Properties of the model*

We 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.

We first constructed an extremely simple model, containing what we considered the bare necessities for semantic evolution. We will provide an overview of this model in this section; a more detailed description is

provided in the appendix. In order to investigate the various hypotheses to explain unidirectionality, we made a series of alterations to this simple model. These will be described in the subsequent sections.

The simple model we 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).
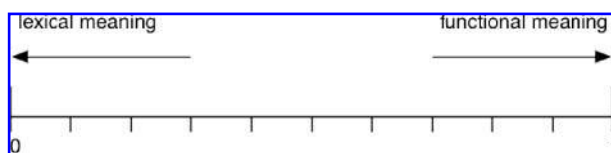

Figure 1.   *The one-dimensional semantic scale of the model*

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

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
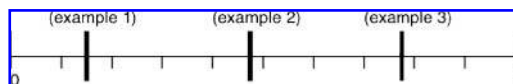

Figure 2.   *Possible positions of different senses of* while *on the one-dimensional scale*

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.)

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. The size of this set is an indication of the generality or specificity of the meaning of word w. Because the 0–1 scale represents the scale lexical-functional meaning, the *position* of the set is an indication of the grammatical status of word w (Figure 3).
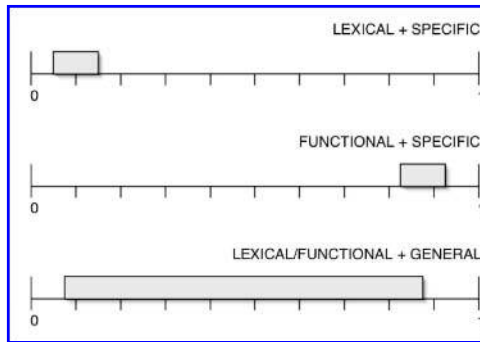

Figure 3.   *Examples of the representation of specificity vs. generality in meaning (set size) and lexical vs. specific (set position)*

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 (represented by a value) from its 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 succesfullness 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, will increase its set up to (and including) the uttered sense. 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 makes the uttered sense its new limit. As a result, speaker and hearer end up with similar limits. Figure 4 illustrates this process.
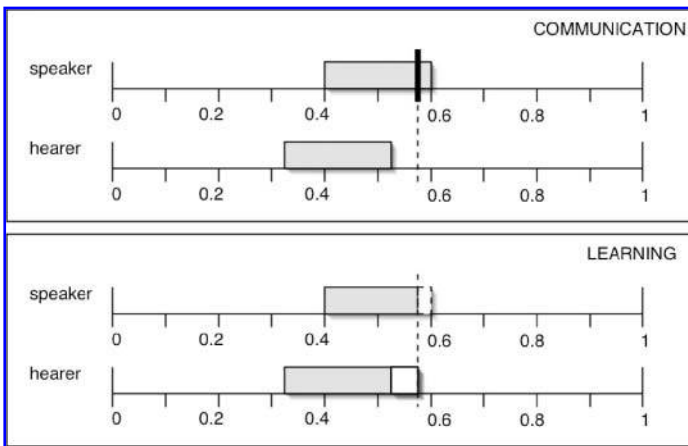

Figure 4. *An example of communication and learning. When a speaker utters a sense that is not known to the hearer, this leads to a learning process in which both speaker and hearer adjust their set of senses.*

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_r 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 we do not directly address the nature of the causes of mutations, and simply start from the assumption that they occur; but we will model certain properties of mutations that were discussed in the previous section, such as their size and their likelihood to occur in words with a more lexical or a more functional

meaning. Mutations are variable and small, and their size is determined by a Gaussian function with a standard deviation of m_s.

The model is iterated in 500 cycles we call "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.

The population consists of 100 agents, and the 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 we are able 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 and Feldman 1981).

In Section 2 we 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_r and a size m_s (eq. 1 in the appendix). The set size is an indication of the generality or specificity in meaning, and we can take the position of the set on the scale as an indication of the lexical or functional status of the word. In the next section, we first discuss general properties of our 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.

4.   Results

4.1.   *General behavior of the model*

In the standard model described in the previous section, we find 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 (Figure 5).

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 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.
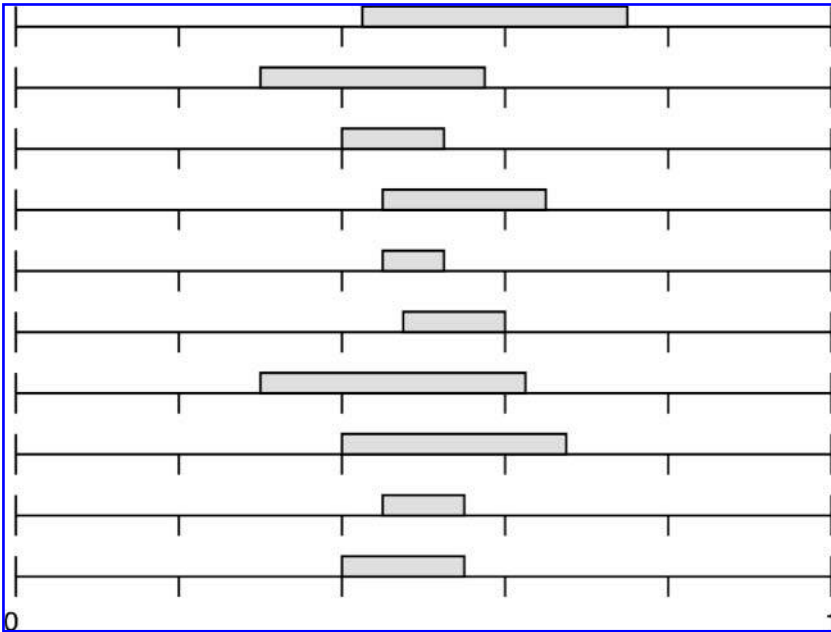

Figure 5. *Examples of random drift of the average meaning of w in 10 populations (N = 100) after 500 years, showing both drift on the 0–1 scale and drift in size. Each population started with an average knowledge with limits [0.4–0.6]. f = 500, m_r = 0.01, m_s = 0.01.*

We tested the effect of three factors on this coherency: mutation rate, frequency of use and population structure. Coherency was measured as the average amount of overlap, between agents in that population, of the sets of senses. The greater this overlap, the greater the consensus about the meaning of word *w* (eq. 3 in the appendix).

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). Changes in frequency do not affect the coherency of the population significantly (Figure 6b). This is due to the fact that, in this model, the rate of mutation is linked to the frequency of
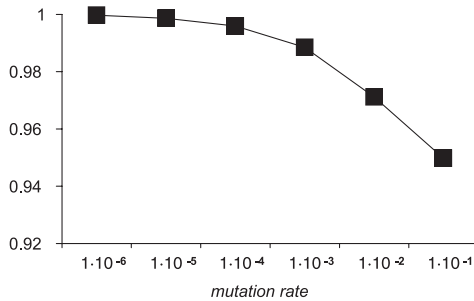
(6a)
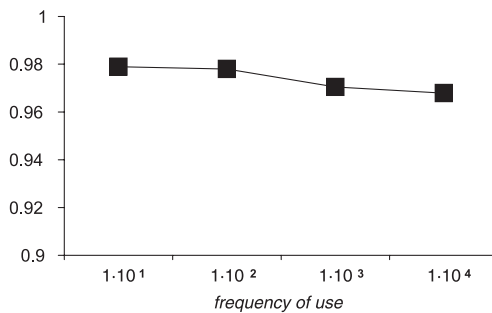
(6b)

Figure 6.   *The coherency of the population (y-axis) with different mutation rates (6a) and different frequencies of use (6b) after 500 years. N = 100, $m_s = 0.01$. For (6a), f = 500, for (6b), $m_r = 0.01$.*

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, 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 and Milroy 1992). We 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. 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 (Figure 7).
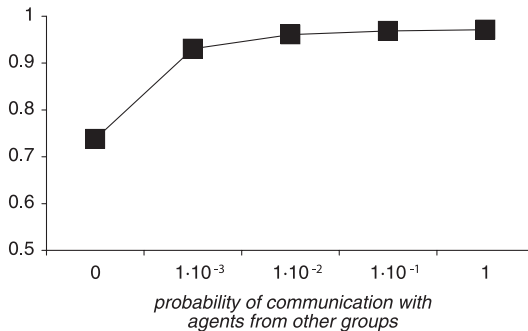

Figure 7. *The coherency of a population of $N = 2000$ divided into 20 subgroups of 100 agents, with different rates of g, the probability of communication with an agent from another subgroup.* t = 500 years, f = 500, m_r = 0.01, m_s = 0.01.

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. At the same time, it introduces a novel class of cultural evolutionary models by exploring a difference between genetic and cultural transmission that has led to disputes over the value of the cultural evolutionary analogy (cf. contributions in Aunger 2000,

i.a. Sperber 2000). Whereas genes are transmitted directly from one organism to another, and the phenotype is constructed based on the information they contain, in culture this pattern is reversed: mental representations must be constructed from observation of the phenotypes of other organisms. This difference is made explicit in our model. 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, we find that 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 "many-to-one" (cf. Cavalli-Sforza and Feldman 1981), combined with the ability to generalize to fill in the gaps.

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

4.2.　*Factors affecting the rate of semantic change*

As was 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 our model is typically rather small, and is determined by a Gaussian function with a standard deviation (m_s). 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.

We 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. 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, we take the amount of change in the position of the meaning set on the 0–1 scale after 500 years (equals
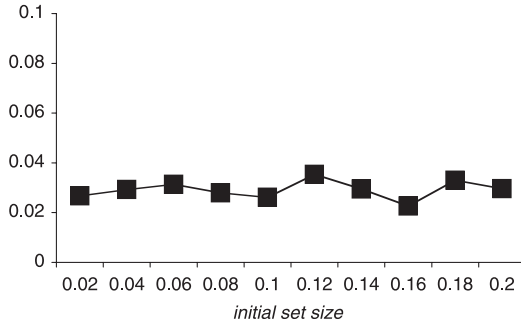
Figure 8. *The amount of drift for increasing set sizes. The y-axis shows the average amount of drift over 20 groups. N = 100, t = 500 years, f = 500, m_r = 0.01, m_s = 0.01.*
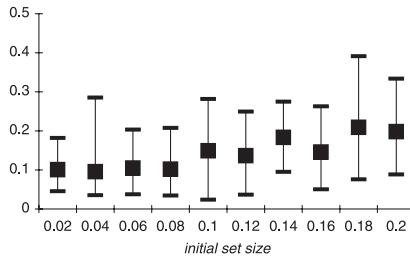

Figure 9. *Final set sizes after 500 years for 20 groups, for different initial set sizes. Apart from the averages (in squares), the minimal and maximal values are shown to indicate the large amount of variance between groups. N = 100, f = 500, m_r = 0.01, m_s = 0.01.*

4–6 in the appendix). 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 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.

Table 1.   *Word frequencies in the British National Corpus (per million words).*

	frequency (per million words)	frequency relative to *come*
come	1512	—
walk	215	0.142
roll	49	0.032
side	30	0.020
swim	25	0.017

Source: *Word Frequencies in Written and Spoken English: based on the British National Corpus.* Geoffrey Leech, Paul Rayson, Andrew Wilson (2001). Longman, London. http://www.comp.lancs.ac.uk/ucrel/bncfreq/flists.html

First, there is a minimal set size for each agent $(1 \cdot 10^{-5})$. Whenever learning or mutation in the form of contractions lead 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.[3] 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.

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 mutation $(m_r = 0.01)$. In order to use somewhat realistic relative differences in frequency, we used the relative frequencies of some English movement verbs (mentioned as an example in Bybee et al 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*. We used similar orders of magnitude in the simulation ($f = \{10, 100, 1000, 10000\}$).

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.

To test the effect of the third factor, mutation, we 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
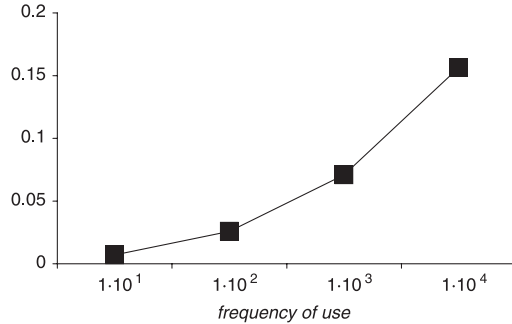
Figure 10. *The role of frequency of use in determining the rate of semantic change. The y-axis shows the average amount of drift between 20 groups for increasing frequencies of use. N = 100, t = 500 years, $m_r = 0.01$, $m_s = 0.01$.*

$f = 500$ per year, respectively. The mutation rate was kept constant at $m_r = 0.01$, but we 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. 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.

If we compare 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 we did not include in our model. On the other hand, both
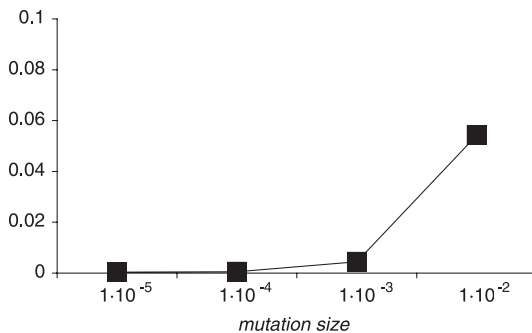

Figure 11. *The role of mutation in determining the rate of semantic change. The y-axis shows the average amount of drift between 20 groups for increasing amounts of mutation. N = 100, t = 500 years, f = 500, $m_r = 0.01$.*

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 nondirectional, which is why we have referred to them as "drift". In the next section we will discuss directionality in change.

4.3.   *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.

We 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, we kept the mutation rate constant at $m_r = 0.05$, but varied the probability of the direction of mutations 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 we 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 (eq. 2 in the appendix).

In a first simulation, we combined both factors, asymmetry in mutation and asymmetry in frequency of use. For the asymmetry in mutation, we 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, our model is more conservative than the hypothesis. For the asymmetry in frequency of use, we used $p_s = 2$, which also leads to a weak preference for more functional meanings. We ran two simulations, 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 ([0.7–0.9]). The latter was added to see whether change in the opposite direction, i.e., from functional to lexical meaning was possible.

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).

As a comparison, we ran a simulation 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.

However, the question remains whether the two factors could also cause unidirectional change when they operate alone. We first tested whether the asymmetry in frequency of use by itself can create a sufficient
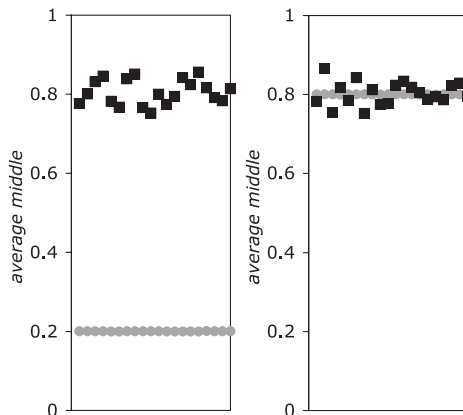

Figure 12.   *The effect of asymmetries in mutation and frequency of use. The results are shown for 20 unrelated groups (x-axis). The y-axis shows the average position of each group on the 0–1 scale. Grey circles indicate the average starting position of each group, black squares end positions after* t = 500 *years.* f = 500, m_r = 0.05, m_s = 0.05, p_m = 0.55, p_s = 2. *Left: lexical starting point; right: functional starting point.*
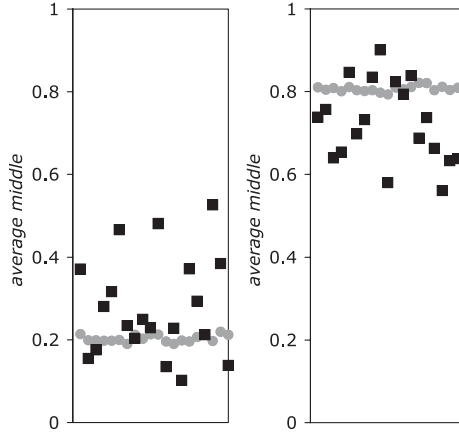
Figure 13.   *The effect of a random drift simulation. The results are shown for 20 unrelated groups (x-axis). The y-axis shows the average position of each group on the 0–1 scale. Grey circles indicate the average starting position of each group, black squares end positions after t = 500 years. f = 500, m_r = 0.05, m_s = 0.05, p_m = 0.50. Left: lexical starting point; right: functional starting point.*

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.
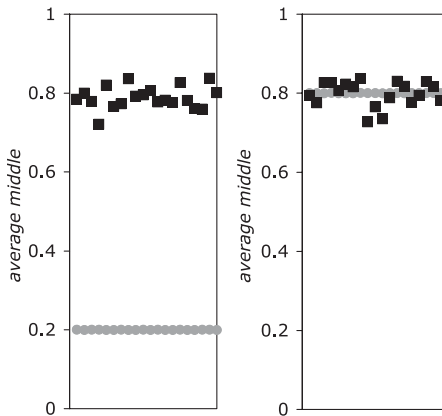

Figure 14.   *The effect of an asymmetry in frequency of use, using p_s = 2. The results are shown for 20 unrelated groups (x-axis). The y-axis shows the average middle of each group on the 0–1 scale. Grey circles indicate the average starting position of each group, black squares end positions after t = 500 years. f = 500, m_r = 0.05, m_s = 0.05, p_m = 0.50. Left: lexical starting point; right: functional starting point.*
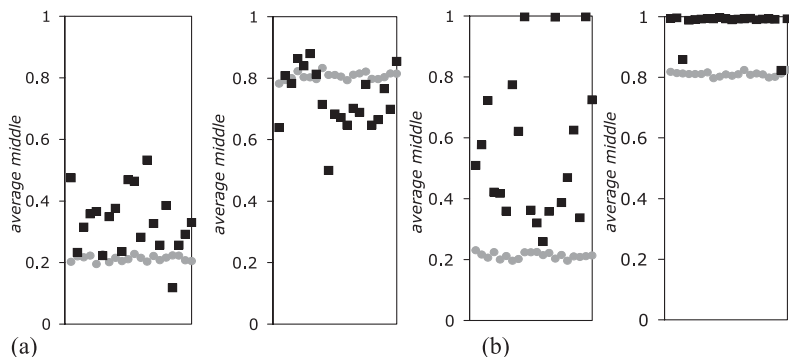
Figure 15.   *The effect of two different asymmetries in mutation. The results are shown for 20 unrelated groups (x-axis). The y-axis shows the average middle of each group on the 0–1 scale. Grey circles indicate the average starting position of each group, black squares end positions after* t = 500 *years.* f = 500, m_r = 0.05, m_s = 0.05, p_m = 0.55 *(Figure 15a),* p_m = 0.95 *(Figure 15b). Left: lexical starting point; right: functional starting point.*

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 15a). 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% of the cases. We therefore tested a much stronger asymmetry of $p_m = 0.95$, which indeed proves to be a very strong pressure for unidirectionality in change (Figure 15b).

While we 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, we 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 16 is a combination chart that shows the average distance for different settings of both an asymmetry in mutation and in frequency of use. Note that $p_m = 0.50$ and $p_s = 1$ are simulations with random drift.
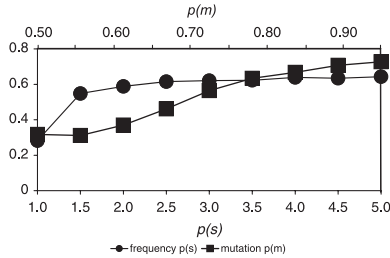
Figure 16.   *The effect of different values of asymmetries in mutation* (p_m) *and frequency of use* (p_s). *The average distance is the distance between the initial average middle and the average middle after 500 years of 20 groups.* f = 500, m_r = 0.05, m_s = 0.05.

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 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 artefact 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 and 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 our model shows that the elementary mechanism of a small difference in frequency is powerful enough to cause unidirectionality by itself.

5.   Conclusions

Our results demonstrate how a cultural evolutionary perspective may be of use in making sense of linguistic hypotheses about language change. We 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 we investigated may be operating at the same time. By deliberately keeping our model simple, we were able 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: we 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 (we 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 et al. 1994, Traugott and Dasher 2002) as possible causes and our results confirm their hypotheses. Our 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. Our 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 our 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 our model either. Related to this, the frequency of use of the word as a whole was constant in our model, while a shift towards functional meaning usually implies an increase in absolute frequency of use. Lastly, our model simulated words in isolation, disregarding influences from factors such as context, which is argued to play a major role in the grammaticalization process.

We 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 we know from historical linguistics to occur.

Perhaps the most striking result emerging from the simulations presented in this article, 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 we used here contains only one word so that issues of competition and *relative* frequency do not enter into the picture here. Nevertheless, we 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.

Appendix. Mathematical properties of the model

General

The simulation is carried out in a series of iterated sequences, each representing one "year". The simulation is run for 500 years, with each agent involved in 500 speaker events (and 500 listener events), on average per year, which is referred to

as the frequency of use *f*. Speakers and hearers were selected in random order in each year.

The population is kept at a constant size of $N = 100$ by yearly replacing agents over 70 years of age by new agents who have one randomly assigned parent (of age 15–50). New agents get an exact copy of their parent's context range. The value for N is an intermediate value of those suggested by Dunbar (1998) and Milroy and Milroy (1992).

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 $S_{min} = 0.4 + d$, with *d* a random value $d \in [0, 0.2]$.

Agent knowledge

For agent *i*, the knowledge about the meaning of *w* is defined by the minimum ($S_{min,i}$) and maximum ($S_{max,i}$) senses that it understands, with each sense represented by a double-precision number $s[0,1]$. We ensured that $S_{max,i} - S_{min,i} > 1 \cdot 10^{-5}$ by limiting any change to the limits (due to learning or mutation) that might cross this threshold.

Mutation

Mutation occurs with a fixed probability m_r each time an agent is selected as speaker in communication. The side of the set on which mutation takes place is determined by the probability p_m, which represents the probability that the mutation will change the value of $S_{max,i}$ or $S_{min,i}$. Unless otherwise indicated, $p_m = 0.5$.

The actual mutation size *m* is variable, and the probability of a certain size is determined by a Gaussian function and the parameter m_s. It is calculated by the following equation:

$$(1) \quad prob(m = x) = |m_s \cdot e^{-(1/2)x^2}|$$

Communication

In communication, a randomly selected speaker *i* utters a sense s_i to a randomly selected hearer *j*. Sense s_i is randomly selected from the speaker's set (where $S_{min,i} < s_i < S_{max,i}$), according to a uniform distribution, except in some of the simulations of unidirectionality, where senses are selected randomly according to the following distribution:

$$(2) \quad prob(s_i = x) = p_s^x$$

while ensuring $S_{min,i} < s_i < S_{max,i}$.

Communication is deemed successful if s_i is recognized by *j* (i.e., $S_{min,j} < s_i < S_{max,j}$). Successful communication has no consequence, but unsuccessful communication leads to both *i* and *j* adjusting their limits.

Learning

After an unsuccessful communication attempt, the receiver *j* increases its set of senses to include s_i (it makes the inference that s_i must be an appropriate sense,

since the speaker used it). the new minimum or maximum sense of the receiver becomes $s_i \pm 1 \cdot 10^{-6}$. This extra value is added to ensure that s_i can be used in communication later on (since the limits of the set themselves are excluded from communication). The speaker i adjusts its set in a similar way, but instead reduces its set and makes $s_i \pm 1 \cdot 10^{-6}$ the new maximum or minimum limit (it makes the inference that, since s_i is already not understood well, senses that are even closer to the minimum or maximum values will not be understood at all).

Population structure
When the population is divided into subgroups, the likelihood of communicating with someone from another subgroup is determined by a probability g.

Measurements
All measurements are averaged over 20 simulations. The coherency θ of a population is measured by calculating the overlap in meaning between agents:

$$(3) \quad \theta = \frac{1}{N(N-1)} \cdot \left(\sum_{i=1}^{N} \sum_{j=1}^{N} \frac{w_{ij}}{w_i} \right) \qquad with\ i \neq j$$

with w_{ij} the overlap in meaning between agents i and j, and w_i the set size of agent i.

Drift δ 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. For a group with i agents, the middle d of the average set of senses of all the agents is calculated and compared to the average middle d_{av}.

$$(4) \quad d = \frac{1}{N} \cdot \sum_{i=1}^{N} (S_{max,i} + S_{min,i})/2$$

$$(5) \quad d_{av} = \frac{1}{MN} \cdot \sum_{i=1}^{MN} (S_{max,i} + S_{min,i})/2$$

with M groups of N agents.

$$(6) \quad \delta = \sqrt{\frac{1}{M-1} \cdot \sum_{j=1}^{N} (d_j - d_{av})^2}$$

The standard deviation between these middles indicates drift δ. The larger the standard deviation, the greater the value for δ.

The simulation was implemented in Java, and the source code is available from the authors.

Notes

2. OED = Oxford English Dictionary.
3. 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.

References

Aunger, Robert (ed.). 2000. *Darwinizing culture: The status of memetics as a science*. Oxford: Oxford University Press.

de Boer, Bart. 2001. *The origins of vowel systems*. Oxford: Oxford University Press.

Boyd, Robert & Peter J. Richerson. 1985. *Culture and the evolutionary process*. Chicago: Chicago University Press.

Bybee, Joan L. & Dan I. Slobin. 1982. Why small children cannot change language on their own: Suggestions from the English past tense. In Anders Ahlqvist (ed.), *Papers from the 5th international conference on historical linguistics*, 29–37. Amsterdam & Philadelphia: John Benjamins.

Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar. Tense, aspect, modality in the languages of the world*. Chicago & London: Chicago University Press.

Cavalli-Sforza Luigi Luca & Marcus W. Feldman. 1981. *Cultural transmission and evolution: A quantitative approach*. Princeton, NJ: Princeton University Press.

Croft, William. 2000. *Explaining language change: An evolutionary approach*. London: Longman.

Croft, William & D. Alan Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.

Dunbar, Robin. 1998. *Grooming, gossip, and the evolution of language*. Cambridge, MA: Harvard University Press.

Geeraerts, Dirk. 1997. *Diachronic prototype semantics. A contribution to historical lexicology*. Oxford: Clarendon Press.

Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37. 1043–1068.

Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

Heine, Bernd, Ulrike Claudi & Friederike Hünnemeyer. 1991. *Approaches to grammaticalization. A conceptual framework*. Chicago & London: Chicago University Press.

Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.

Keller, Rudi. 1994. *On language change: The invisible hand in language*. London & New York: Routledge.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar, vol. I: Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Milroy, Leslie & James Milroy. 1992. Social network and social class: Towards an integrated sociolinguistic model. *Language in Society* 21(1). 1–26.

Niyogi, Partha & Robert C. Berwick. 1997. A dynamical systems model for language change. *Complex Systems* 11. 161–204.

Simpson, John A. & Edmund S. Weiner (eds.). 1989. *Oxford English dictionary*, 2nd edn. Oxford: Clarendon Press.

Slobin, Dan I. 2005. From ontogenesis to phylogenesis: What can child language tell us about language evolution? In Sue Taylor Parker, Jonas Langer & Constance Milbrath (eds.), *Biology and knowledge revisited: From neurogenesis to psychogenesis*, 255–286. Mahwah, NJ: Lawrence Erlbaum.

Sperber, Dan. 2000. An objection to the memetic approach to culture. In Robert Aunger (ed.), *Darwinizing culture: The status of memetics as a science*, 163–173. Oxford: Oxford University Press.

Steels, Luc. 1998. Synthesising the origins of language and meaning using coevolution, self-organisation and level formation. In James R. Hurford, Michael Studdert-Kennedy & Chris Knight (eds.), *Approaches to the evolution of language*, 384–404. Cambridge: Cambridge University Press.

Steels, Luc. 2003. Evolving grounded communication for robots. *Trends in Cognitive Science* 7(7). 308–312.

Tomasello, Michael. 2003. *Constructing a language. A usage-based theory of language acquisition*. Cambridge, MA & London: Harvard University Press.

Traugott, Elizabeth Closs & Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.

Werner, Heinz & Bernard Kaplan. 1963. *Symbol-formation: An organismic-developmental approach to language and the expression of thought*. New York & London: Wiley.

Yang, Charles D. 2000. Internal and external forces in language change. *Language Variation and Change* 12(3). 231–250.