Computational Studies of Language Evolution*

William S.-Y. Wang, Jinyun Ke and James W. Minett

Language Engineering Laboratory, City University of Hong Kong

The study of language evolution has revitalized recently due to converging interests from many disciplines. Computational modeling is one such fruitful area. Various aspects of language evolution have been studied using mathematical modeling and simulation. In this paper we discuss several computational studies in language change and language emergence.

1. Introduction

The origin and evolution of language, the most distinctive aspect of our species, has intrigued the human mind since ancient times. Earlier speculations on these questions were seldom fruitful because there was virtually no empirical foundation to build upon. It is well known that the linguistic societies in Paris and London banned such discussions in the 19th century. By the middle of the 20th century[1], however, many of the disciplines relevant to these questions had began to come together. Our ability to deal scientifically with these questions has been increasing at an accelerated pace.

These disciplines ranged literally from A to Z, from anthropological concern with the physical development of our remote ancestors, to zoological interest in animal communication and culture. More central here are the discoveries by linguists of universal tendencies found in all languages (Greenberg 1963), by psycholinguists of the dynamics of language acquisition and loss (Jakobson 1941), and by neuroscientists of how language is organized in the brain (Deacon 1997).

Over the last several decades, the range of disciplines has broadened in two major steps. First, genetics has come on board with important hypotheses regarding the age of our Most Recent Common Ancestor, and regarding the correlation between groups of peoples and groups of languages. This development started with the so-called classical markers, and has been suc-

---

*This paper is based on a keynote address presented by the first author at the 2002 COLING in Taipei. We thank the organizers of the conference, particularly Churen Huang, for their kind invitation. Our research is supported in part by grants from the City University of Hong Kong [#9010001], and from the Hong Kong Research Grants Council [#9040781].

[1] Some good landmarks for the return to respectability of the discussion of these issues include the well-known paper by Hockett (1960), and the large conference anthologized by Harnad et al. (1976).

cessively refined to gender-specific materials, first mitochondrial DNA for the maternal line and then the Y-chromosome for the paternal line. Evidence is gradually emerging that although anatomically modern humans first appeared over 100,000 years ago, our Most Recent Common Ancestor may date to only some 50,000 years ago. Such a date correlates well with the sudden burst of cultural achievements at many sites in the world, including art, ritual, and the navigational skills to sail across large expanses of water.

It is reasonable to think that language evolved to its modern form around this date, since it is most likely that the power of language facilitated these cultural achievements. A more recent emergence date certainly makes the question of emergence more tractable, since there has been less time to obscure the traces of our primordial language(s)[2]. Indeed, some bolder scholars have been prospecting for words that may have existed in the primordial language which have been preserved in most branches of the world's languages today. And other scholars have been exploring the possibility that the unique click consonants still extant in South Africa were indeed part of the primordial language phonology which had become lost in the branch of humans that left Africa to populate the rest of the world.

Fascinating as these explorations are, the fact remains that most of the pieces of evidence collected from the various disciplines are circumstantial[3], and that it is not possible to directly reconstruct the stages whereby our ancestors invented language dozens of millennia ago. This leads us to the second major step after genetics — the use of computational linguistics in the study of language evolution, which for convenience we shall refer to as CSLE: computational study of language evolution. This is an area which has burst upon the scene with great vitality, attracting exciting research from a variety of viewpoints. This vitality can be seen from the many anthologies which have become available since 1998, including those by Hurford et al. (1998), Knight et al. (2000), Cangelosi and Parisi (2001), Briscoe (2001) and Wray (2002).

Few proponents of CSLE take the innatist position that there is literally an autonomous organ for language, that language requires a special bioprogram, or that language is based on any instinct exclusive to it. Obviously, a very wide array of abilities must have been in place before our ancestors were ready for language, ranging over sensory, motoric, memorial and cognitive dimensions, as well as social skills in courtship, forming alliances, collaborating in group activities, and strategizing against enemies. Many of these abilities are present to various extents in our ape relatives, although it is clear our ancestors must have had more language readiness than apes do. It is encouraging that some recent studies are beginning to give us hints on the neurobiological bases of some of these abilities, such as the discovery of the so-called mirror neurons and their implications for the ability to imitate.

---

[2] It is still an open question whether language was invented only once (monogenesis), or several times independently (polygenesis). Freedman and Wang (1996) present some arguments to support the latter view.

[3] Writing was invented much later, after the advent of agriculture some 10,000 years ago.

The basic assumptions that CSLE makes are that numerous interactions among members of a community, as well as among members across communities, over a long span of time can result in behaviors and structures which are quite complex. Furthermore, the bottom-up paths leading to such complex structures often involve phase transitions, points in time at which there are abrupt non-linearities where the change seems to be more qualitative than quantitative.

We see such phase transitions in the physical world, for instance, when ice changes abruptly to water, and then abruptly to steam, even when heat is added gradually and by a constant amount. Similarly, we can perhaps identify some phase transitions in the cultural evolution of language, as in the emergence of segmental phonology, the invention of hierarchic morphology and syntax, the use of recursion in sentence construction, etc. The points in time for such non-linearities and the driving forces for change are not nearly as well-defined and uniform as in physical systems, of course.

The linguistic analog to the addition of heat driving the phase transition in water would be the increasingly complexity of the communicative needs of early hominids as a result of their own expanding consciousness as they interacted with the environment (Schoenemann 1999). Furthermore, given that by 50,000 years ago there were numerous communities scattered in many parts of the Old World in diverse environmental niches, it is very likely that the evolution of language proceeded at different rates in these communities, each community crossing the various linguistic thresholds in its own way and at its own pace.

We will now consider three distinct approaches to computational studies of language evolution: modeling of lexical diffusion and the snowball effect from a dynamical systems perspective, modeling the evolution of universal grammar, also using dynamical systems, and modeling the emergence of the lexicon from a multi-agent system perspective.

2. Modeling lexical diffusion and the snowball effect

In the middle of the 20th century, the dominant view of sound change was that the unit of change is the phoneme. This was a view that linguistics had essentially inherited from the influential Neogrammarians of the 19th century, who emphasized the doctrine that sound changes can have no exceptions. Taking cues from evolutionary theory in biology, the counter-proposal was that the unit of change is the word (Wang 1969). Wang suggested that a change proceeds by variation, often partitioning the relevant words into three classes: unchanged (U), variation (V) and changed (C). This view of language change was termed *lexical diffusion*, since the change diffuses itself across the population one word at a time.

An early study of lexical diffusion was conducted by Don Sherman (1975), who investigated the growth of diatones in the history of English, that is, the increase in the number of noun-verb pairs like "permit"/"permit", "contract"/"contract", etc. The earliest pronouncing dictionary he could find, that of 1570, listed only 3 such pairs, the next dictionary listed 8 such pairs, and so on, up to 1934. Plotting the growth in numbers of diatones against time, the graph which results, shown in Figure 1, suggests that we may have the beginning of an S-curve. Apparently, such

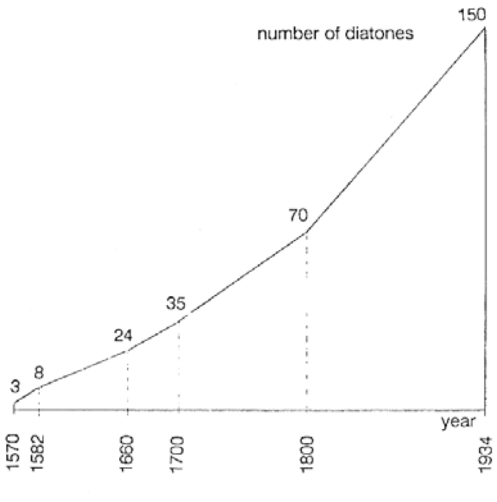

Figure 1. The chronological profile of diatone formation in English, after Sherman (1975).

curves are widely found in diffusions, both for cultural events and biological events. The biologists Luca Cavalli-Sforza and Marc Feldman (1981:29–30) wrote:

> The new word that becomes part of a language, … , is an innovation and can be considered as an analog of mutation in biology. … When the process of diffusion of an innovation is followed for a sufficiently long time, the frequency of use of the innovation almost always follows an S-shaped curve. At the beginning the number of acceptances rapidly increases. Then follows an approximately linear increase, and finally the increase slows down and is barely perceptible.

Two centuries earlier, the poet Alexander Pope expressed the same idea in more social terms:

> In words, as fashions, the same rule will hold
> Alike fantastic, if too new or old
> Be not the first by whom the new are tried
> Nor yet the last to lay the old aside

With the realization that lexical diffusion of a single word has an S-shaped trajectory, the question naturally follows as to whether the words which are cohorts in a given change influence each other, and on the nature of this influence. One answer to these questions may be summarized in the term *snowball effect*. The term takes its name metaphorically from a snowball rolling down a snowy mountainside. The further down it rolls, the faster it goes, and the more snow it picks up along the way. So if the S-curve for the first word has a particular gradient, then the curve for the second word has a steeper gradient. Furthermore, the time delay between the first pair of words will be greater than that between the second pair of words, and so on.

So far, there have been two empirical studies on the snowball effect. The study by Ogura & Wang (1996) deals with the development of the *-s* suffix in the third person singular present indicative in the history of English, starting from the Early Modern English of the mid-15th

century. The *-s* suffix competed with the *-th* suffix in this function for several centuries until it completely replaced the latter in the end.

The other, more detailed study is by Shen Zhongwei (1997) on the merger of two nasalized vowels in modern Shanghai. Shen used the ages of his informants as virtual time. Assuming that a person's habits of pronunciation are largely fixed by age 15, say, then an informant who is 60 years old may reflect the speech of 45 years ago. Such a method is far from fail-safe, of course. But it is nevertheless very useful for shorter term changes which run their courses over several decades. Shen's study has the merit of being based on a large number of informants — almost 400. There are 28 relevant words in Shen's list, each of which is pronounced with a nasalized low front vowel by some older speakers. One by one, the vowel in these relevant words moves back in its articulation, and the words become homophones with words which have back vowels. In other words, this is a classic case of vowel merger, of which there are numerous examples in language change. Our questions have to do with how the relevant words influence each other in the process.

2.1 A dynamical system model of lexical diffusion — one word

We begin our discussion of modeling lexical diffusion and the snowball effect by describing a dynamical system, first derived by Shen (1997), that models a sound change which effects a single word only[4]. The model applies to a group of homogeneous language users who can each adopt one of two possible forms for the word that is undergoing the change, either the unchanged form, U, or the changed form, C; the model does not allow for free variation, V. The state of this system at any time instant, t, can be described in terms of the proportion, or *frequency*, of individuals who use the unchanged form, $u(t)$, and the frequency of individuals who use the changed form, $c(t)$. Note that since each individual must adopt either U or C, $u(t) + c(t)$ must sum to 1 for all t.

It is assumed that the frequencies u and c at some time instant can be calculated from the frequencies at an earlier time instant. In particular, it is assumed that use of the changed form is propagated by contact between pairs of speakers, one of whom uses the unchanged form, the other uses the changed form; thus the increase in the frequency of changed forms is proportional to the product $c(t) \times u(t)$. The increase in the frequency is also proportional to the *rate of effective contact* (Shen 1997), α, and the length of time over which the sound change is observed, δt. Hence the frequencies of changed and unchanged forms at time $t + \delta t$ can be written in terms of the frequencies at time t as

$$c(t+\delta t) \;=\; c(t) \;+\; \alpha\, c(t)u(t)\delta t\,, \qquad (1a)$$

[4] We would particularly like to thank Jeff Chasnov of the Mathematics Department of the Hong Kong University of Science of Technology for his assistance in re-deriving Shen's model for lexical diffusion of a sound change effecting a single word and deriving the models for a sound change effecting an arbitrary number of words — any errors in the presentation of these models are our own.

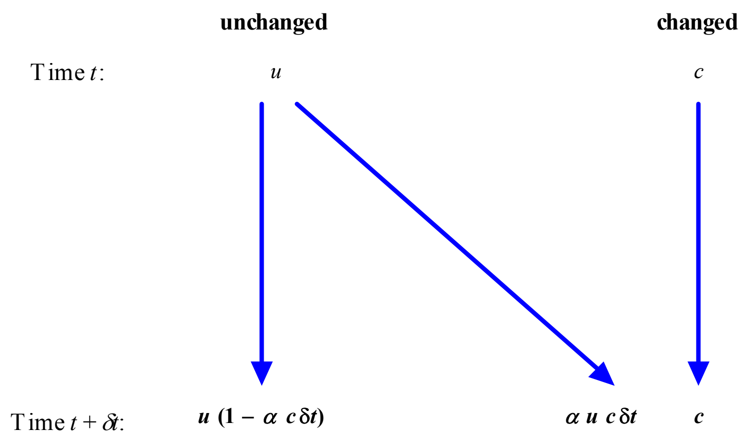

Figure 2. Diffusion of unchanged and changed forms over a time interval of duration δt — one word.

$$u(t+\delta t) \;=\; u(t) - \alpha\, c(t) u(t) \delta t\,. \tag{1b}$$

The parameter α can also be interpreted as representing phonetic, social and other pressures on individuals to adopt the changed form.

Figure 2 summarizes the rates at which the frequencies of changed and unchanged forms vary over a time interval of duration δt. Taking the difference in the values of c at time t and time $t + \delta t$, and dividing by the duration δt gives the rate of change of the frequency of changed forms:

$$\frac{c(t+\delta t) - c(t)}{\delta t} \;=\; \alpha\, c(t) u(t)\,. \tag{3}$$

Taking the limit as δt tends to zero produces the differential equation

$$\frac{dc}{dt} \;=\; \alpha\, c(t) u(t). \tag{4}$$

Recalling that $u + c = 1$, we obtain the differential equation

$$\frac{dc}{dt} \;=\; \alpha\, c(t)[1 - c(t)]. \tag{5}$$

The general solution to (5) is the well-known Logistic equation,

$$c(t) \;=\; \varepsilon \frac{\exp(\alpha\, t)}{1 + \varepsilon[\exp(\alpha\, t) - 1]}, \tag{6}$$

where ε is the initial frequency of changed forms at time $t = 0$.

A plot of $c(t)$ is given in Figure 3 for an initial value of $\varepsilon = 1\%$ and the rate of effective contact $\alpha = 0.1$. The plot exhibits the characteristic feature of the S-shaped logistic curve — a slow initial increase, followed by a period of more rapid, almost linear increase, which quickly drops

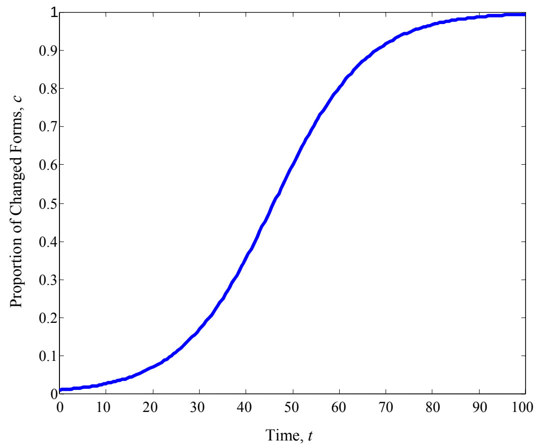

Figure 3. The logistic curve. ($\varepsilon = 1\%$, $\alpha = 0.1$)

off again as the frequency approaches 100% — and indicates the gradual diffusion of the changed form throughout the entire population.

2.2 A dynamical system model of lexical diffusion — multiple words

Having derived an expression equivalent to (6), Shen (1997) went on to apply the model to the diffusion of a sound change across a group of words. He assumed that the frequency of changed forms for each word could be described independently by the model just discussed, determining parameter values with the best fit to data collected for 28 words in Shanghainese that exhibit the merger of /ɑ̃/ and /ã/. We take a different approach, however, extending the model just described by explicitly accounting for coupling among the words themselves, in addition to the coupling between speakers that has already been modeled; that is, we assume that the rate of diffusion of the sound change in one word may affect the rate of diffusion in other words.

Given a group of n words that are effected by a sound change, we denote the frequency of unchanged forms of word i at time t by $u_i(t)$ and the frequency of changed forms of that word by $c_i(t)$, where $u_i(t) + c_i(t) = 1$. We extend the definition of Shen's rate of effective contact (α above) by specifying for each pair of words the rate, α_{ij}, at which adoption of the changed form of word i is induced by the frequency of changed forms of word j — we call this the *coupling rate* of word j on word i, referring to α_{ij} for distinct i and j as *cross-coupling*, and to α_{ii} as *self-coupling*.

The rate of increase of changed forms of word i is assumed to depend on $u_i(t)$, as in Shen's model. However, due to the coupling that we assume to exist between words, we propose that the rate of increase is proportional to the combined effect on word i of the frequencies of changed forms of all the words participating in the sound change. The frequencies of changed and unchanged forms at time $t + \delta t$ can therefore be written in terms of the frequencies at time t as

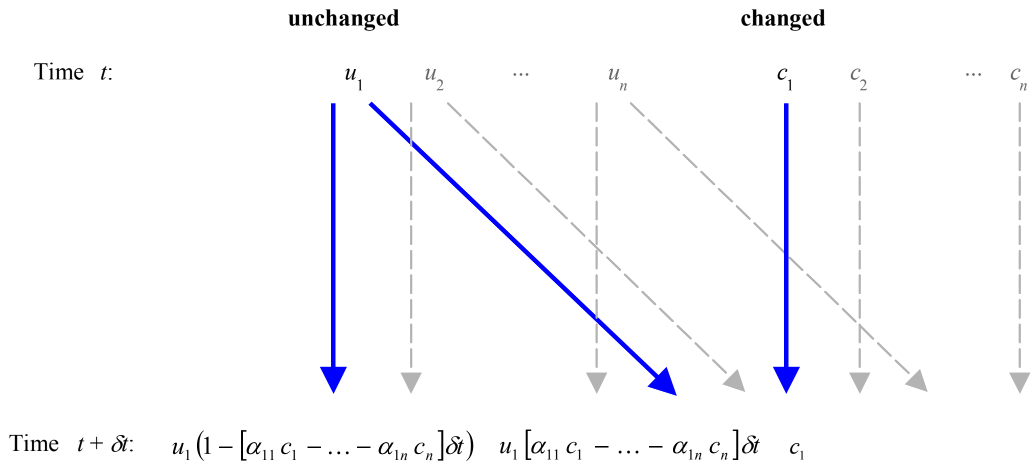

Figure 4. Diffusion of unchanged and changed forms over a time interval of duration δt — multiple words.

$$c_i(t+\delta t) \;=\; c_i(t) \;+\; u_i(t)\sum_{j=1}^{n}\alpha_{ij}\,c_j(t), \tag{7a}$$

$$u_i(t+\delta t) \;=\; u_i(t) \;-\; u_i(t)\sum_{j=1}^{n}\alpha_{ij}\,c_j(t). \tag{7b}$$

Figure 4 summarizes the rates at which the frequencies of changed and unchanged forms change over some time interval δt.

Letting δt tend to zero, as previously, we obtain the following system of differential equations for the frequencies of changed forms:

$$\frac{dc_i}{dt} \;=\; (1-c_i)\sum_{j=1}^{n}\alpha_{ij}\,c_j\,. \tag{8}$$

We have found no analytic general solution to (8) but we note, in particular, that the frequency of changed forms of any word does not follow a logistic curve, as assumed by Shen, unless the cross-coupling is zero, i.e., $a_{ij} = 0$ for all $i \neq j$. We can however characterize the behavior of the system using numerical methods.

Figure 5 shows the frequencies of changed forms predicted by the model for four words with the coupling rates and initial values given in Table 1. The system is initiated with only a single word having undergone any change. The other three words are distinguished by their having different coupling rates. In order to simplify the system somewhat, we have set the cross-coupling rate of each word with respect to all other words to a constant, although the value of the constant differs for each word.

At first, Word 1 grows at a faster rate than the other words. This is due to the frequency of changed forms of Word 1 far exceeding those of the other words — growth due to the self-

Table 1. Coupling rates and initial frequencies of changed forms of four words undergoing lexical diffusion.

Coupling: (α_{ij})	Word 1	Word 2	Word 3	Word 4
Word 1	20%	1%	1%	1%
Word 2	4%	14%	4%	4%
Word 3	3%	3%	16%	3%
Word 4	2%	2%	2%	18%
$c_i(0)$	1%	0	0	0

coupling of Word 1 therefore exceeds that due to cross-coupling. Because the frequencies of changed forms of the other words are initially zero, their growth is initiated by cross-coupling between themselves and Word 1. As the sound change progresses, the growth rates gradually increase as self-coupling is strengthened, although the rate of increase is initially greater for those words with higher cross-coupling. For example, at about $t = 10$, the rate of increase of the frequency of changed forms of Word 2 exceeds that of either Word 3 or Word 4. The rate of increase of the earlier words then begins to fall, allowing the later words to catch up. Eventually, the words that commenced the sound change later overtake the words that preceded them. For example, soon after time $t = 30$, Word 2 has attained a higher frequency of changed forms than Word 1; some time later, Word 3 has also changed more than Word 1 (although this is not visible in the figure). This feature is not unreasonable: as Ogura and Wang (1996) observed in their study of the sound change *-th* to *-s* in English, most of the words that commenced the sound change later actually completed the change before the earlier words. In our example, the frequency of changed forms in each language eventually approaches 100%.

Even for this simplified system, the relative progress of the sound change in each word does not behave in a transparent manner. For example, we do not know what combination of values of self-coupling and cross-coupling for each word cause certain words to attain a higher frequency of changed forms despite having commenced changing more slowly. We therefore simplify the system further such that the behavior becomes more transparent. We do this by setting both the self-coupling and the cross-coupling for each word to be constant. The coupling rates for a set of n words can therefore be represented by two parameters: the self-coupling, denoted by β, and the cross-coupling, denoted by γ, i.e.

$$\alpha_{ij} = \begin{cases} \beta : & i = j \\ \gamma : & i \neq j \end{cases}. \tag{9}$$

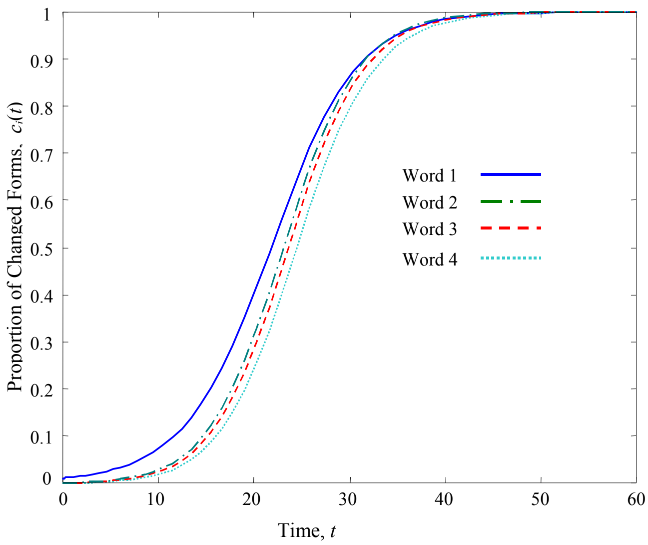

Figure 5. Predicted frequencies of changed forms for four words.
(parameter values given in Table 1)

Note that we require that $\beta \geq \gamma$ since we expect the influence of a word on itself to be at least that of other words. The differential equation (8) that describes the evolution of the frequency of changed forms of each word reduces to

$$\frac{dc_i}{dt} = \left(1 - c_i\right)\left[\beta c_i + \gamma \sum_{\substack{j=1 \\ j \neq i}}^{n} c_j\right]. \tag{10}$$

Figure 6 shows the frequencies of changed forms predicted by this simplified model for four words with the coupling rates and initial values given in Table 2. Note that, since each word has the same values of self-coupling and cross-coupling, only different initial values of the frequency of changed forms, $c_i(0)$, generate different curves. For this reason, we have encoded each word with a different initial value of c, which emulates the words' starting to participate in the sound change at slightly different times. The figure clearly shows that initially the frequencies of changed forms of words that participate in the sound change earlier grow at a faster rate than that of words that participate later. At about time $t = 10$, however, the rates of growth of the

Table 2. Coupling rates and initial frequencies of changed forms of four words undergoing lexical diffusion — simplified model.

Coupling:	Word 1	Word 2	Word 3	Word 4
All words	… $\beta = 20\%$ $\gamma = 2\%$ …			
$c_i(0)$	5%	2%	0.5 %	0.1 %

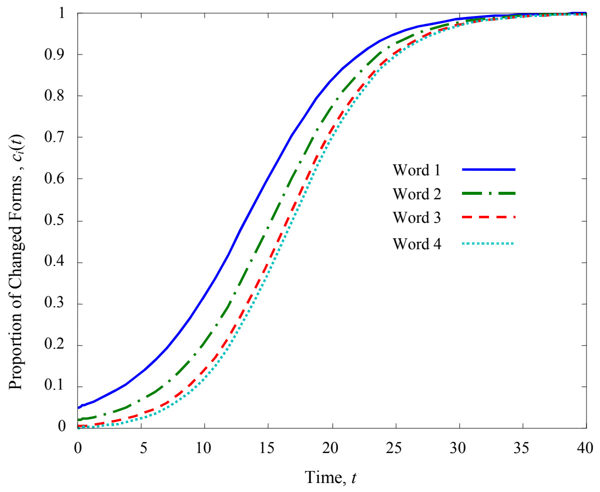

Figure 6. Predicted frequencies of changed forms for four words — simplified model. (parameter values given in Table 2)

words are approximately equal. Later, the words that participate in the change later progress at a faster rate than the earlier words; in other words, we observe a snowball effect. Eventually, the frequency of changed forms in each language approaches 100%. At no time, however, does a later word attain a higher frequency of changed forms than an earlier word. The simplified model therefore does not replicate the behavior of the full model for which later words do sometimes attain a higher frequency of changed forms than earlier words (as Ogura and Wang (1996) observed). Nevertheless, we are able to demonstrate that the snowball effect — by which we mean that later words eventually adopt changed forms at a faster rate than earlier words — is inevitable under the simplified model (10), as follows.

Suppose that a sound change effects a group of n words. To demonstrate the snowball effect we must show that when the frequency of changed forms of one word exceeds that of another word, the rate of change of the frequency of the latter word exceeds that of the former. We therefore specify the constraint

$$c_1(t) \; > \; c_2(t) \tag{11}$$

at some time instant t and determine the condition under which

$$\frac{dc_2}{dt} \; > \; \frac{dc_1}{dt} \, . \tag{12}$$

By equation (10), (12) holds when

$$(1 - c_2)(\beta c_2 + \gamma c_1 + \gamma k) \; > \; (1 - c_1)(\beta c_1 + \gamma c_2 + \gamma k), \tag{13}$$

where k is defined in terms of the frequencies of changed forms of the other words:

$$k \; = \; \sum_{j=3}^{n} c_j \, . \tag{14}$$

Thus (12) holds when

$$\beta\left(c_1^{\,2}-c_2^{\,2}\right)-\beta\left(c_1-c_2\right)+\gamma\left(c_1-c_2\right)+\gamma\, k\left(c_1-c_2\right)\;>\;0\,,$$

$$\Rightarrow\quad \beta\left(c_1+c_2\right)-\beta+\gamma+\gamma\, k\;>\;0\,,$$

$$\Rightarrow\quad \beta\, c_1+\beta\, c_2+\gamma\, k\;>\;\beta-\gamma\,. \tag{15}$$

Note that c_1, c_2 and k are all increasing functions of t, while $\beta-\gamma$ is a constant. Hence equation (15) indicates that when the frequencies of changed forms of the two words, c_1 and c_2, are large enough, the frequency of the later word grows more quickly than that of the earlier word, thereby demonstrating a snowball effect. Note also that since the rates of increase of the frequencies of later words eventually exceed those of earlier words, the tendency is for the curves to converge over time. Whether or not the distance between successive curves is less for later words than for earlier words will depend on the initial values of c_i.

In Figure 7, we show the snowball effect for the four words with the parameter values specified in Table 2, drawing attention to the time instant at which later words adopt the changed form at a higher rate than the word immediately preceding them. The first such transition occurs at about time $t=13$, when Word 4 begins to adopt the sound change at a faster rate than Word 3. Soon after, the rate of growth of Word 3 overtakes that of Word 2, followed by Word 2 overtaking Word 1. The actual order of the transition depends on the initial values of c_i.

2.3 Discussion

The dynamical system just described appears able to capture a number of features that are typically observed among a group of words that undergo a sound change. However, its utility as a realistic model of the diffusion of a sound change is yet to be established. In order to make

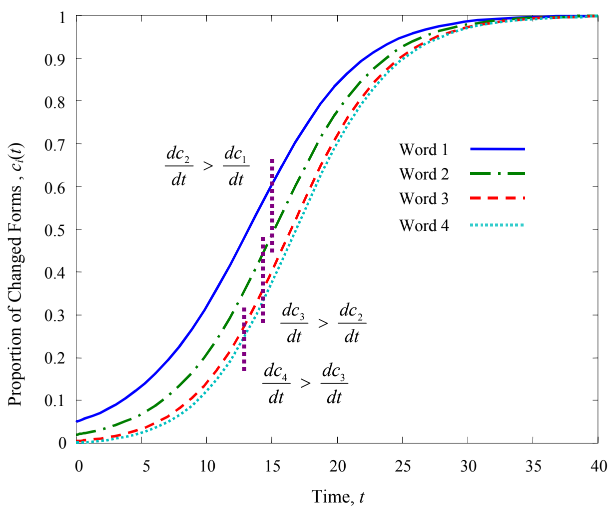

Figure 7. The snowball effect for four words — simplified model.
Marked on the figure is the threshold time beyond which the rate of increase of the frequency of changed forms of each word exceeds that of the preceding word.

predictions about the behavior of the frequencies of changed forms for a sound change in progress, the values of the frequencies of changed forms of each word and the coupling rate between each pair of words would have to be known. While the former may be estimated (with some difficulty) by polling speakers of the dialect undergoing the sound change, the latter are not directly measurable. We must therefore perform a test of the model, much like that performed by Shen (1997), by collecting data for various sound changes that have run their course and estimating the values of coupling and initial frequency that produce the best fit between predicted behavior and observed behavior. We intend to test our multiple-word model using the two data sets used by Shen (1997), allowing the two models to be compared directly, as well as locating other data, such as can be found in (Lee 2002), and collecting new data sets.

Both the dynamical system developed by Shen and the extension of it that we present here model lexical diffusion under highly regular conditions: a sound change within an isolated, static population of language users, each behaving identically. While such models may capture very well the expected evolution of a sound change in isolation, the reality is nowhere near so regular. As Ogura (1993) shows graphically, diffusion often proceeds in fits and starts; Figure 8 shows the progress of a syntactic change (periphrastic *do* in English) in various sentence types. Clearly the change was not regular; nor did it complete, the change being reversed in one of the sentence types analyzed. Realistic models of lexical diffusion should be able to capture such kinds of behavior. One way to extend the models discussed earlier to achieve this is to allow the coupling rates to vary with time. This should allow dynamic social and phonological pressures

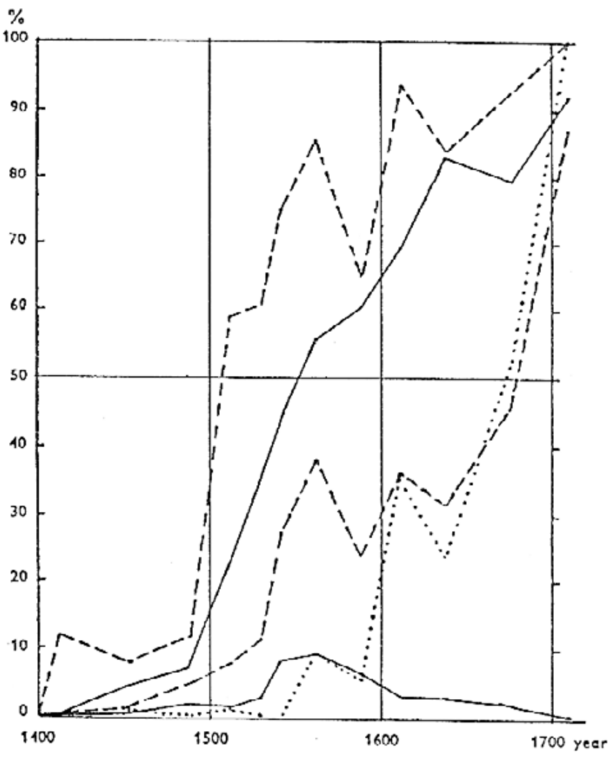

Figure 8. The progress of a syntactic change (periphrastic *do* in English), after Ogura (1993). Lines represent the progress of the change in sentences of distinct types.

to be modeled — diffusion can be "switched off" by setting the coupling rates to zero; competing changes can be modeled by introducing multiple changed forms; and so on. Of course, as we extend the expressive power of the model, so we render more complex the task of discovering the history of a lexical diffusion process.

2.4 Relationships to other systems

Abrams and Strogatz (2003) have recently developed a dynamical systems model to describe language death. The model deals with a population of speakers who must choose between either of two languages, an endangered language and a prestige language, which differ in terms of their proportion of speakers and their social status. They assume that a language becomes more attractive as its number of speakers and its social status increase. The proportion of endangered language speakers, x, is modeled by the differential equation

$$\frac{dx}{dt} = (1-x)P_{yx}(x,s) - x\,P_{yx}(1-x,1-s), \tag{16}$$

where s is a constant indicating the social status of the endangered language and $P_{yx}(x,s)$ is the attractiveness of that language to speakers of the prestige language. Abrams and Strogatz simplify the system by assuming that $P_{yx}(x,s)$ has the form

$$P_{yx}(x,s) = c\,x^{a}\,s\,, \tag{17}$$

where c is a constant and a is a parameter that adjusts the effect of number of speakers on the attractiveness of a language. They show that such a system provides an excellent fit to data for the number of speakers of several endangered languages, including Scottish Gaelic, Quechua and Welsh. Their results indicate that each of these languages will soon become extinct unless strategies to increase their social status are undertaken.

The similarity between the Abrams and Strogatz model for language death and Shen's model for lexical diffusion is striking — by setting $a = 1$ in equation (17) (and suitable re-labeling), the Abrams and Strogatz model reduces to Shen's model (cf. equation (5)). Although the model predicts that one language will always drive the other to extinction when the social status, s, is constant, other stable states become possible when s is allowed to vary. For example, control on s by active feedback can lead to a stable bilingual state, suggesting that linguistic diversity can be maintained by taking appropriate action to enhance the status of endangered languages (Abrams & Strogatz 2003). Extended to lexical diffusion, this approach may help us to model the effects of shifts in the relative social status of competing languages to the progress of a sound change. For example, there is some evidence that the sound change word initial /ŋ/ to nothing that is currently in progress in Hong Kong Cantonese is reversing, probably due to the ongoing increase in the social status of Mandarin ever since the return of sovereignty of Hong Kong to China in 1997. Application and extension of the Abrams and Strogatz model may bring us new insights into how sound changes and other lexical innovations diffuse across a population of speakers.

As we consider in the next section, another computational model, which has been used to study the evolution of universal grammar, might also be adapted to the task of modeling lexical diffusion.

3. Modeling language evolution — dynamical systems

The area of lexical diffusion discussed in the previous section deals with language evolution at an intermediate window size, with changes taking place over decades or centuries. Language evolution can be studied at many time scales, however, from the short interchanges between mother and child in early language acquisition to the many millennia over which language has evolved since its first emergence. We now turn to the question of language emergence.

Computational studies of language emergence provide a very valuable antithesis to the currently popular innatist position that there is literally an autonomous organ for language, or that language requires a special bioprogram, or that language is based on any instinct exclusive to it. If language emerged only very recently as recent studies in population genetics indicate, then the likelihood is small indeed that biological evolution could have put such an organ in place.

On the other hand, it is obvious that a very wide array of abilities must have been in place before our ancestors were ready for language, ranging over sensory, motoric, memorial and cognitive dimensions, as well as social skills in courtship, forming alliances, collaborating in group activities, and strategizing against enemies. Some years back, Wang (1978:116) characterized this point of view with words like 'mosaic' and 'interface'. It was in this spirit that Tzeng and Wang (1983) carried out a set of experiments to argue for a common neuro-cognitive mechanism for both language and movements.

A basic assumption that many computational studies make is that numerous interactions among members of a community, as well as among members across communities, over a long span of time can result in behaviors and structures which are quite complex. When Murray Gell-Mann (1994) wrote of the evolution of "highly complex forms," he could have easily included languages among his examples.

The field of modeling language evolution by computer essentially began with Hurford's (1989) discussion of the emergence of a consistent lexicon. Hurford considered the relative merits of three highly idealized learning strategies. Individuals adopting the *Imitator* strategy produce a particular utterance to indicate a certain object when they observe that nearby individuals typically produce that utterance to indicate the object. Similarly, they attend to a particular object when perceiving a certain utterance when they observe that nearby individuals typically attend to that object in response to the utterance. *Calculators*, however, use a particular utterance to indicate a certain object when they observe that others attend to the object when perceiving that utterance. Similarly, they attend to a particular object in response to an utterance when nearby individuals use that utterance to indicate the object. Hurford shows, however, that a better strategy is to follow the approach he refers to as *Saussurean*. Individuals adopting the Saussurean strategy copy the speech production patterns of nearby individuals, like Imitators.

But unlike Imitators, they make their perception consistent with their production. Thus, Saussureans base both their production and perception on the speech production of other individuals.

Surprisingly, perhaps, Hurford's paper did not stimulate an immediate interest to take up the challenge of modeling language emergence by computer. The field had to wait for further stimulation in the fields of game theory, cellular automata and artificial neural networks before contributions in the mid-1990's, such as by Clark & Roberts (1993), Batali (1994), Hutchins & Hazlehurst (1995), Steels (1995), and Noble & Cliff (1995), began to appear.

While many models for language evolution have adopted the agent-based simulation paradigm (which we discuss in Section 4), comparatively few models have been based on dynamical systems. An appropriately designed dynamical system is most useful for describing the *qualitative* behavior of a system rather than predicting the *exact* behavior for a particular instance of the system. The behavior of a system can be described in terms of its stable and unstable equilibria, oscillations, the values of certain parameters at which the system bifurcates, and so on. Much of the recent work on modeling language change from a dynamical systems perspective has come from Martin Nowak and his associates. They have focused particularly on modeling the evolution of universal grammar, which we now discuss.

3.1 Evolution of universal grammar

Nowak, Komarova and Niyogi (2001) have proposed a dynamical system model of the evolution of Universal Grammar (UG) among a population of heterogeneous language agents. Universal grammar is an abstract representation of one currently popular view of language acquisition. UG consists of a "mechanism to generate a search space for all mental candidate grammars" and "a learning procedure that specifies how to evaluate sample sentences" (Nowak et al. 2001). The UG model fits very well with the innatist viewpoint of language emergence: that our language faculties derive from a language-specific organ, the Language Acquisition Device (LAD). However, the model can equally well be applied without adopting an innatist position, provided we assume that children are able to develop a learning algorithm for acquiring language without recourse to exclusively innate capabilities of some LAD. Nowak and colleagues also assume that the UG consists of a finite set of grammars.

In the model, each agent uses a single grammar. In some cases, a sentence that can be parsed in one grammar may also be parsed in another grammar. Thus users of different grammars may often be able to communicate with some degree of success. For example, while the sentence "I might could do" is not grammatical in many English dialects (although many people would understand the intended meaning), it is grammatical in parts of Scotland and the USA (Trask 1996); nevertheless, many sentences that are grammatical in those particular dialects are also grammatical in many other English dialects. The result is that such dialects are mutually intelligible to such a degree that communication between them is essentially unimpaired.

For each pair of grammars, the probability that a user of grammar G_i produces a sentence that can be parsed by a user of grammar G_j is denoted by a_{ij}, with $0 \leq a_{ij} \leq 1$ and $a_{ii} = 1$. The probability that two agents who use grammars G_i and G_j respectively can communicate successfully is therefore given by

$$F\left(G_i, G_j\right) = \tfrac{1}{2}\left(a_{ij} + a_{ji}\right). \tag{18}$$

The *payoff* of grammar G_i is then defined as the average probability that a speaker of G_i produces a sentence that can be parsed by an arbitrary individual, i.e.

$$f_i = \sum_{j=1}^{n} x_j F\left(G_i, G_j\right), \tag{19}$$

where x_j is the frequency of individuals using grammar G_j. The payoff associated with a particular grammar is assumed to be linked to the reproductive success of individuals who use that grammar. Nevertheless, children do not always learn the grammar of their parents.

In the model, the distribution of the grammars that a child may learn depends on a matrix, Q_{ji}, whose values indicate the probability that a child with a parent who uses G_i learns to use G_j. Children do not select a grammar that is optimal in any sense; rather they acquire a grammar according to the probability distribution Q_{ji}. Nevertheless, selection pressure due to the linkage between payoff and reproductive success will tend to cause grammars that can parse sentences produced by many other grammars with high probability to gradually diffuse across the population over successive generations.

Nowak et al. propose the following language dynamical equation to explain the evolution of the system:

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} x_j f_j Q_{ji} - \phi x_i , \tag{20}$$

where $\phi = \sum_j x_j f_j$, termed the *grammatical coherence*, measures the average probability of mutual understanding among members of the population. Equation (20), much like equations (8) and (10) in the section on lexical diffusion, describes a kind of diffusion process, although, as Nowak and colleagues demonstrate, the diffusion does not necessarily complete. They display a graph, repeated in Figure 9, to indicate under what (simplified) conditions a population converges to a single dominant grammar — distinct grammars are assumed to be equidistant in the sense that $a_{ij} = a$ for some constant $0 \le a \le 1$ and $i \ne j$. They find that a single dominant grammar only diffuses across the population when the probability that children learn the grammar of their parents, Q_{ii}, is sufficiently high; otherwise a state in which each grammar has equal frequency emerges. Even when a single dominant grammar does emerge, it does not diffuse across the entire population unless children learn the grammar of their parents perfectly — other grammars continue to be used by some members of the population, although with relatively small frequency.

The UG model, (20), appears to have a formulation very similar to that of the lexical diffusion (LD) model, (8), which we discussed earlier. Since, as we have just observed, the UG

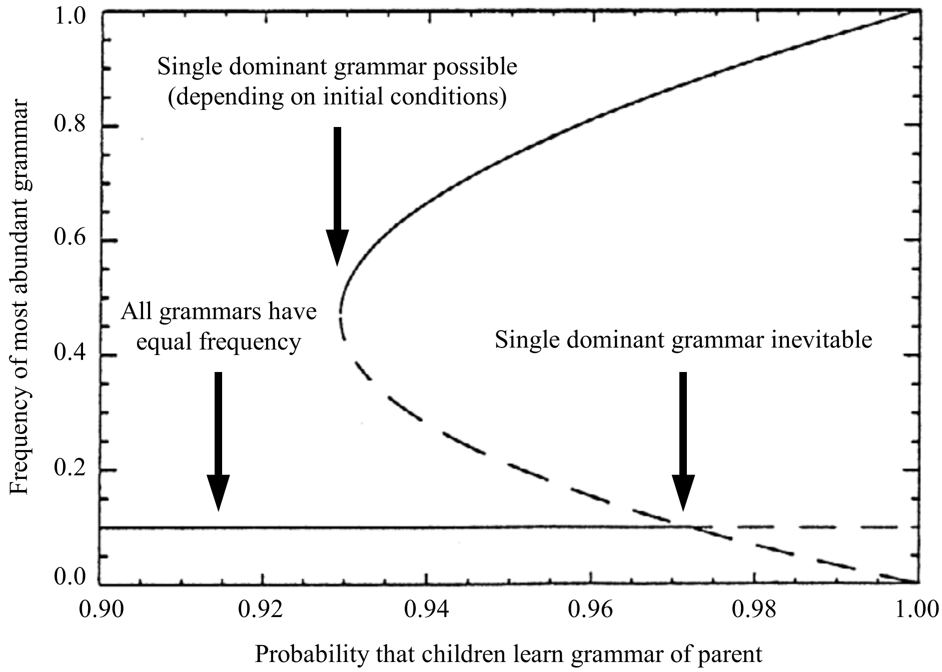

Figure 9. Conditions under which population which population converges to a single dominant grammar, adapted from Nowak et al. (2001).

model sometimes produces diffusional behavior, the question arises whether this model can be used to model some features of lexical diffusion. Equation (20) can be rearranged as

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} \left(Q_{ji} - x_i\right) f_j \, x_j \; . \tag{21}$$

The lexical diffusion equation (8) can be written as

$$\frac{dc_i}{dt} = \sum_{j=1}^{n} \left(1 - c_i\right) \alpha_{ij} \, c_j \; . \qquad (22).$$

The two sets of variables, x_i in (21) and c_i in (22), both represent the frequencies of a set of time-varying, measurable linguistic quantities. In the UG system the frequencies must sum to 1 — the various grammars must therefore "compete" for users. However, in the LD system there is no such competition — the frequency of changed forms of any particular word may reach 1 without forcing the frequencies for other words to fall. This substantially different behavior is introduced by the factor $(Q_{ji} - x_i)$; the UG frequencies increase whenever $Q_{ji} > x_i$ but fall whenever $Q_{ji} < x_i$.

In his initial formulation of lexical diffusion (1969), Wang had in mind that sound changes sometimes compete for words, causing apparent irregularities to appear in the changing phonological system. For example, consider the two hypothetical sound changes $\mathbf{R_1} : \mathrm{A} \rightarrow \mathrm{B}$ and $\mathbf{R_2} : \mathrm{A} \rightarrow \mathrm{C}$ that describe the change of a segment A either to B or to C. Suppose further that $\mathbf{R_1}$ commences first but is interrupted by $\mathbf{R_2}$ before it completes. Those words that have already acquired B due to $\mathbf{R_1}$ will not be affected by $\mathbf{R_2}$. Only those words that still maintain the unchanged segment A will be affected by $\mathbf{R_2}$. Some words, however, may be in the process of change, with some members of the population still using the unchanged form, A, while others use the changed form, B — in such cases the two sound changes will tend to compete for both words and speakers. The result then will be that some words will have replaced A with B, while others will have replaced A with C (some free variation between B and C (and perhaps A also) might persist among some speakers). A similar, apparent irregularity would emerge if the second sound change were instead $\mathbf{R_2}'$: B → C.

What then can Nowak's model of the evolution of universal grammar tell us about lexical diffusion and, more generally, about linguistic diffusion processes at large? Equation (20) tells us that the adoption of a single grammar by an entire population occurs only rarely, if ever. Instead of using the x_i in (20) to represent the frequencies of the numerous possible grammars, we could consider using the x_i to represent the frequencies of the various sounds that can arise due to a set of competing sound changes. If we take the parameters Q_{ji} to indicate the probability that sound i is replaced by sound j, and substitute for f_i a coefficient that measures the effect of coupling between both pairs of words and between sounds, we obtain a dynamical system describing the evolution of a set of competing sound changes. Developing such a model might allow us to gain more insight into the circumstances under which qualitatively different types of behavior emerge when a set of sound changes compete.

4. Modeling language evolution — multi-agent systems

The model proposed by Nowak et al. in studying language evolution is concerned with the evolution of universal grammar from the perspective that language as a whole evolves under the mechanism of natural selection. While this analytic approach is promising in providing a framework to study the dynamics of language evolution as a whole system, it gives us little hint on *how* a language system came into being. When speculating on the process of language emergence, there can be few who believe that a complex language system with elaborate lexicon, morphology and syntax could have sprung up all of a sudden from scratch. Language must have emerged and evolved gradually and incrementally to reach its modern form. And this process must have developed due to the communication interactions among individuals. It is through these communication interactions that language emerged and evolved to meet the increasing communication needs.

Our proposition is that, given the cognitive and physiological prerequisites being available, language emergence and evolution is basically a continuous conventionalization process, from the individual innovation of a new linguistic item, a word, a phrase or a syntactic construction, to the diffusion of the innovation through interactions among individual language users by imitation and learning during language acquisition. A compelling scenario for the emergence of language is that first a set of early words or holistic signals emerged, and that later different word orders or relationships between words came to be used to signify different aspects and moods, etc. in order to cope with the increasing need to express more complex meanings. Our view, therefore, is that one of the first steps in exploring the emergence of language in the macro scale should be to study how the first words emerged.

4.1 The emergence of vocabulary

Words are the smallest communicative functional units in language, a word (usually) being an arbitrary association between a meaning and a signal[5]. A modern individual typically has many thousands of words in his vocabulary through which he perceives his universe, and by means of which he communicates his needs and desires. At the outset, however, such symbols were much fewer — zoologists tell us that no animal in its natural state has more than several dozen symbols, be they vocal calls, facial expressions or body gestures (Wilson 1972, Hauser 1996). While animals' communication systems are mostly imprinted innately (though some of them are affected by learning, e.g. bird song, reported by Marler (1987)), the words used in a linguistic community are mostly established by conventionalization. Two great philosophers have pronounced similar ideas regarding the conventionalization of naming. Xunzi in China taught that "words have no intrinsic correctness" and "words have no intrinsic content" (translation by Wang (1989)), while at about the same time in Greece, Plato wrote that "any name which you give is the right one, and if you change that and give another, the new name is as correct as the old" (translation by Jowett (1953), quoted by Wang (1989)).

We have designed several models to simulate the process of conventionalization leading to emergence of a shared vocabulary from a phylogenetic point of view (Ke et al. 2002b). We have made a number of assumptions which are hypothetically plausible for early hominids. First, the agents are assumed to already possess the ability of naming, or, more generally, are able to use symbolic signs, which is considered to be a species-specific trait of *homo sapiens* (Deacon 1997). Second, there exists a set of meanings that are particularly salient in their daily life. Third, the agents are all able to produce the same set of utterances. The agents intentionally interact with each other to communicate these meanings by manipulating these utterances. The association between meanings and utterances can be represented in various ways, for example, by a look-up table, an association matrix, or a neural network. Models reported in Ke et al. (2002b) adopt two different forms, following earlier studies, i.e. look-up tables (Steels 1995) and probabilistic matrices (Hurford 1989, Oliphant 1997, Nowak et al. 1999).

The emergence of a shared vocabulary refers to a stage in which agents have the same set of associations between meanings and utterances, for both speaking and listening. The question then is how these associations are formed and how members of a population reach the same set of associations (whether starting from scratch or from random creation by each agent). The answers to these questions lie in the modes of interaction among agents during communication.

---

[5] In the following discussion of vocabulary or words, the terms of "signal", "utterance" and "form" are used interchangeably,

Imitation is one of the most likely interaction mechanisms. The strong ability of humans to imitate, even from early infancy, has been extensively documented in the studies reported by many investigators, e.g. Meltzoff (1996). While other social animals, particularly the primates, also imitate (Dugatkin 2000), it appears that the tendency is by far the strongest and most general in our species. We assume that imitation may serve as the most explanatory mechanism for the formation of a common vocabulary. Before establishing a consistent way of naming things, early humans very likely made use of their propensity for imitation; the younger ones imitating their elders, the followers imitating the leaders or, just by chance, their neighbors. In the simulation model, we assume there to be a number of agents, each of which initially has its own set of mappings between meanings and utterances. When two agents interact, one imitates the other according to some strategy, either by random, by following the majority in the population, or by avoiding homophones. We demonstrate that the agents in the population always converge to a single identical vocabulary. Mathematical modeling using Markov chains has been used to prove the convergence (Ke et al. 2002b).

While in the imitation model we adopt only one set of mappings for the associations between meanings and utterances, in a second simulation model we distinguish the active and passive vocabulary by using two sets of mappings: a speaking matrix and a listening matrix. This representation is considered to be more realistic, and is necessary when considering the fact that active and passive vocabularies are generally not identical. An example of the two matrices with three meanings and three utterances is given in Table 3. Each element of the matrices represents the probability that an agent has an association between a certain meaning and a certain utterance. The two matrices are stochastic matrices, having the constraint that each row of the speaking matrix and each column of the listening matrix sum to one, to meet the assumption that each meaning is expressible, and each utterance is interpretable.

We hypothesize another type of interaction, in which probabilistic changes are applied to the mappings, rather than imitation in the discrete manner used in the above model. At the beginning of the simulation, the speaking and listening matrices of each agent are both randomly initialized. When two agents interact, a successful interaction occurs when the listener interprets the received utterance as the meaning intended by the speaker, resulting in a reinforcement of

Table 3. An example of the speaking and listening matrices in the interaction model in Ke et al. (2002b)

p_{ij}	u_1	u_2	u_3
m_1	0.3	0.4	0.3
m_2	0.4	0.55	0.05
m_3	0.7	0.2	0.1

q_{ij}	u_1	u_2	u_3
m_1	0.1	0.3	0.6
m_2	0.5	0.3	0.3
m_3	0.4	0.4	0.1

the mapping used in the speaker's speaking matrix and the listener's listening matrix. On the contrary, if the listener interprets a meaning that differs from the one intended by the speaker, such a failed interaction will lead to weakening of the corresponding associations used by the two agents. The agents thus go through a process of iterative self-organization according to a sequence of such interactions.

After a number of interactions, the system converges: a common lexicon emerges in the population (inter-agent convergence), and the speaking and listening matrices of each agent reach a compatible state (intra-agent convergence). When the number of meanings equals the number of utterances, the speaking and listening matrices of each agent are identical. However, when the number of meanings is larger than the number of utterances, the speaking matrix of each agent is a subset of the listening matrices. Nevertheless, in both cases, the speaking and listening matrices of each agent become compatible.

The intra-agent convergence is an emergent property of the system as there is no explicit and obligatory mechanism forcing the speaking and listening matrices to be compatible. This leads us to speculate that it might not be necessary to presume a Saussurean strategy for the formation of vocabulary as proposed by Hurford (1989) — Hurford shows that a Saussurean strategy, in which the speaking and listening matrices are assumed to be identical, has the advantage of high communication effectiveness over other strategies and therefore might have been selected by biological evolution.

Figure 10 shows the trends of three measures of convergence: similarity (SI), population consistency (PC) and individual consistency (IC), from a typical run of simulation. In the run, the population consists of 10 agents, each starting with a vocabulary in which each meaning is randomly associated with each utterance with a different probability. A consistent vocabulary is formed and shared by all agents after a long period of fluctuation.

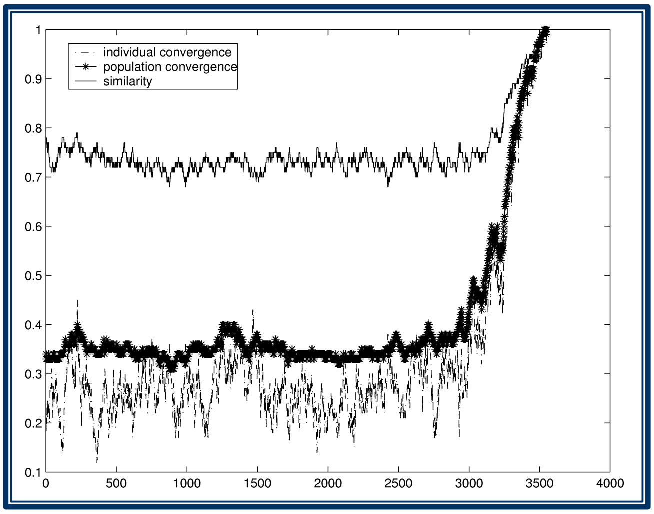

Figure 10. The convergence trends from an example simulation of the interaction model. Adapted from Ke et. al (2002)

It can be seen that the convergence is not gradual but rather is quite abrupt after about 3000 interactions, exhibiting a "phase transition" characteristic. For a long period of time, the interactions among agents only result in fluctuation, and there is little consistency in the population's vocabularies. However, at some instant, there is an abrupt rise of the consistency, and the population converges quickly after that period. The conditions of the model have not changed at all in the process toward convergence. The abrupt emergence of order in the population is the result of a sequence of interactions from which point agents evolve in the same direction, thus bringing about a converging momentum.

For this interaction model, we have also investigated the effects of various parameters. When the agents are prohibited from interacting with themselves, i.e. no "self-talk", we observe an interesting window in the optimal population size, between 5 and 15, in which the population converges the most quickly. The existence of an optimal population size is unexpected. This may be an artifact of a very small population, in which contradictory changes happen often and therefore the matrices oscillate heavily.

When "self-talk" is allowed, however, the convergence becomes much faster, and there is no such window effect: the smaller the population size, the easier it is for the population to achieve a consistent vocabulary. This is because self-talk allows an agent's speaking behavior to influence not only other agents' listening behaviors but also his own listening behavior through interaction between his own speaking and listening behaviors, thereby speeding up the convergence. However, the most important finding is that convergence time does not increase linearly with the increase of the population size, which suggests that there may be a threshold of population size for the convergence to be realistically possible within a bounded period of time. This finding may be linked to speculations that our ancestors tended to gather into populations of 50 to 100 members. It has been proposed that the social structure and size of groups of *homo sapiens* is one of the most important factors in the emergence of language (Dunbar 1993).

The above models are highly simplified in many aspects, for example, the numbers of meanings and utterances are pre-assigned, and each meaning is obliged to be associated with at least one utterance in the beginning, although in reality the meaning space gradually increases. Furthermore, the agents are considered to be immortal, a simplification that is equivalent to considering only a single generation. Much further work should be carried out for various enhancements, for example, to take into account an increasing semantic space, to simulate language learning in overlapping generations and in populations with different social structures, and to see how different clusters of agents share their specific subsets of the lexicon.

4.2. Homophony and ambiguity

The existence and abundance of ambiguity in languages has intrigued linguists for a long time. If we view language as a coding system to encode meanings with signals, it would seem that language is not optimal at all, because in an ideal code one signal should correspond to exactly one meaning. If there are one-to-many correspondences, ambiguities arise. Yet all languages are rife with such ambiguities at various levels, from polysemy to homophony to syntactic ambiguities. Indeed ambiguity has been the most formidable barrier to computational linguistics since its start — from automatic summary, to machine translation, to speech recognition — and remains so today, even as methods of disambiguation are becoming increasingly sophisticated and powerful[6].

In the following, we report some preliminary modeling work to investigate homophony, which is usually considered to be a major source of ambiguity in speech (Ke et al. 2002a). Homophones are pairs or sets of words which have the same phonological forms but different unrelated meanings. We are interested in understanding why and how homophones arise, and how they can survive. Nowak et al. (1999) use a mathematical model to demonstrate that within a given limited signal space, homophony is unavoidable if the least error limit is to be achieved. Steels and Kaplan (1998) have shown with a simulation model how homophony arises and persists in the lexicon from the language change point of view. However, our model, also adopting a simulation framework, addresses the problem from an emergence perspective.

The simulation model is designed within the "naming games" framework proposed by Steels (1995). Agents in the model are assumed to be able to produce a number of distinctive utterances and to make use of such utterances to communicate a set of meanings. At the beginning, the agents do not have any words, a word here referring to an association between a meaning and an utterance. Agents can create new words at random, as well as learn the words created by other agents.

At each time step, two agents are chosen to communicate, one as the speaker and the other as the listener. The speaker decides a meaning he/she wants to communicate, looks for or creates an utterance which is associated with the meaning, and transmits the utterance to the listener. The listener perceives the utterance and tries to interpret the meaning by searching his existing vocabulary. If he interprets the same meaning for the utterance, then this is considered to be a successful communication. Each word has a score; after each successful communication, the score of the word is increased. Otherwise, the score is decreased. When the score of the word becomes too small, the word is removed from the vocabulary. Upon failure, the listener learns the word from the speaker by adding an association between the perceived utterance and the intended meaning of the speaker. At the beginning, all agents have an empty vocabulary. However, after a long period of interaction, we observe that a set of associations between objects and utterances are shared by all agents.

With the above construction, we compare the convergence for two different ratios between the number of meanings (M) and the number of utterances (U). Figure 11 shows simulation results for the two different ratios. When M = U, we can see that agents are able to acquire the same vocabulary, and their communications are successful 90% of the time, 20% of the words having homophones. When we increase the number of meanings that are to be communicated

[6] Ambiguities sometimes serve various purposes in linguistic play — in puns, jokes, etc. — but these are surely developments which arose much later after ambiguities have taken root in languages.

by agents, for example, setting the meaning-utterance ratio to 3, the vocabularies of the agents no longer converge — every word has homophones — resulting in a low rate of communicative success (only about 30%).

This situation, M > U, simulates a more realistic situation in which our semantic needs far exceed the number of forms we can utilize. Cheng (1998) has shown that there exists a general limit to the size of the active vocabulary used by various writers. If we assume that the number of meanings that humans can and desire to manipulate is infinite, the limit on the vocabulary size suggests that there may exist a cognitive constraint on the number of forms which can be memorized as a whole. As a result of the limited number of forms, the condition of M > U is realistically true. To meet the semantic need, it is obvious that the existence of homophony is inevitable under this condition. However, in spite of the considerable ambiguity implied by homophones, our daily communication does not seem to be much hampered by it, contradictory to what the above model shows. Therefore, we need to seek explanations for the effective communication under the condition that M > U which the current model seems unable to demonstrate.

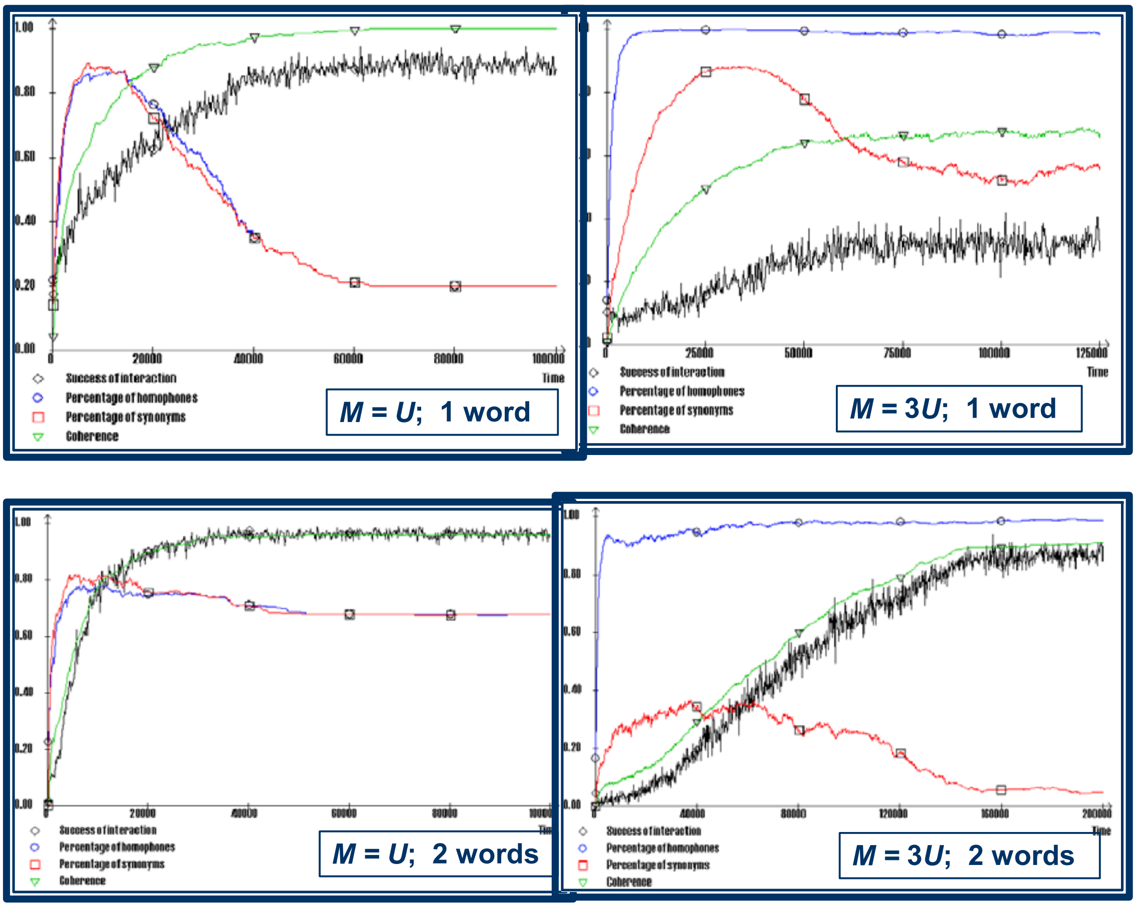

Figure 11. Simulation results the homophone evolution under four conditions.
(upper left: M = U, one-word communication; upper right: M = 3U, one-word communication; lower left: M = U, two-word communication; lower right: M = 3U, two-word communication)

In the above simulation, only one meaning is transmitted during each communication event. In a real situation, most of the time, we communicate with a phrase or a sentence. Words in the phrase or sentence are always semantically related. To simulate this situation, we have designed a two-word communication model. In a communication event, the speaker chooses two meanings (m1 and m2) which are close to each other in the semantic space, and produces two utterances (u1 and u2) to communicate with another agent. The listener receives the two utterances. If u1 has only one meaning, m1, and u2 has two or more meanings, say m3 and m4, then the listener will choose between m3 and m4 according to which is closer to m1 in the semantic space. If neither u1 nor u2 has a unique meaning, the same principle of disambiguation can be applied: the listener will choose that pair of meanings from the u1-meaings and the u2-meaings that are the closest in the semantic space.

In this formulation, while the semantic proximity helps to disambiguate homophonous utterances, random learning could cause trouble when the listener learns the wrong order of association. Nevertheless, we observe a gradual increase in the rate of communicative success through successive interactions. When M = U, we see that the communicative success reaches 100%, much better than the early case of one-word communication, even with a degree of homophony as high as 70%. When we increase the semantic demand, we can see a much clearer improvement of the system owing to the two-word communication. Homophony can be tolerated up to about 100%, with the rate of communicative success still rising to more than 80%. This simulation demonstrates clearly that, with the help of context, the lexicon can tolerate a high degree of homophony, even when the number of meanings greatly exceeds that of utterances.

The simulation model reported here illustrates the point that homophony can persist in the vocabulary while still maintaining a high communication effectiveness, given a realistic multiple-word communication condition. As mentioned earlier, this model is undertaken from the perspective of emergence, i.e. vocabulary starting from scratch. However, we know that homophones constantly emerge as the result of sound merger from language change. Also pairs of homophones exhibit various differentiation characteristics such as in frequency, in part of speech, etc., illustrating self-organization in a language system (Ke et al. 2002). How such self-organization is implemented in the process of agents' interactions will be a challenging topic for further modeling studies.

5. Discussion

It is clear that we should be encouraged by what the new area has achieved so far. The knowledge base for research on language evolution must rest on what linguistics has to offer, regarding how the several thousand languages available to us are organized, from the common core of this organization that is shared by all languages extending to the most idiosyncratic features observed for just a few languages, which marks the outer periphery of what a language can be like. This knowledge base grew tremendously in the 20th century, when linguists described a broad range of languages in many parts of the world which had not been studied scientifically before[7]. This linguistic knowledge has been joined by genetic knowledge since the 1980s in research on language evolution, and by computational studies since the 1990s.

As we look back on this decade or so of CSLE, it is clear that the achievements have been impressive and encouraging. At the same time, we see that there are many central topics on language evolution which await careful formulation and investigation. In this final section, we would like to offer some concluding remarks on this exciting new area of CSLE regarding the assumptions and limitations of current methodologies, as well as regarding the road that lies ahead.

Computational models can be used to demonstrate how certain linguistic structures emerge and/or change, such as the lexical diffusion model and the models of the emergence of vocabulary and homophony we reported above. It is an advantage of computational models that various assumptions must be made explicit and implementable, and thus can be examined, verified and compared. For example, in simulating communication interactions among agents, the models have to clearly specify various details as to how agents' meanings are represented, how meanings are transmitted by the speaker, and how the listener interprets the received signals. And in simulating language acquisition, the models must be explicit regarding the properties that the learners are assumed to be endowed with, such as the learning algorithm, if any, which determines how the learners construct their own language by memorizing and extracting regularities from the linguistic input.

A hypothesis supported by one model might not be supported by another model which is implemented based on different assumptions. For example, Kirby (2001) demonstrates that a compositional language can emerge from a set of random meaning-signal mappings by an iterative-learning model. However, in his model the mappings are represented by a version of Definite Clause Grammar and learners are assumed to have an induction algorithm which can look for common substrings and infer generalized rules generating them, which are highly biased toward language-like systems. He hypothesizes that a bottleneck effect, by which the learner is only exposed to a small subset of the possible language, is necessary for the emergence of compositional language. However, in a critique of this model by Tonkes and Wiles (2002), a neural network model is implemented for which no explicit rules or generalizations are required. They show that compositionality still emerges without requiring a learning bottleneck. From these models, we can see that computational models allow different hypotheses to be evaluated and compared objectively as long as the assumptions and representations have been explicitly stated.

---

[7] It is sad to note concurrently that the 20th century also marks the accelerated extinction of indigenous languages as these are replaced by a few international languages, empowered by economic and technological success. This development has a homogenizing effect which simultaneously expands the common core and shrinks the outer periphery of the space within which language locates.

Currently most models make rather strong assumptions or great simplifications of the real situations. For example, in our model of lexicon formation, we assume that meanings are transmitted explicitly and listeners have no problem at all in knowing the meaning intended by the speaker. Many other computational models simulating the interactions between individuals adopt a similar assumption, especially those in which agents are represented by neural networks and learn the meaning-signal mappings by some training process (e.g. Batali 1998). However, the transparency of meaning in communication may not be true in many real situations as ambiguous interpretations are almost always possible.

Moreover, in the case of first language acquisition, it is not well-known how children can identify the meanings intended by adults. There have in fact been some studies addressing this problem by making the meaning interpretation more realistic. For example, in Steels (2002) the meanings being communicated are embodied by object detection and feature extraction; Cangelosi et al. (2002) incorporate environmental information together with the signal during communication, and the listener interprets the meaning from the environmental information. With these embodiments implemented in the models, consistent communication still emerges without the prerequisites of meaning transference. It is through the process of identifying and then relaxing assumptions in the simulation models that more realistic frameworks are established.

While there have been exciting simulations on the emergence of the lexicon, on the formation of phonological systems, and on the emergence of compositionality in syntax, not much is known about how hierarchical syntax emerged. Hierarchical structures are a hallmark of complexity, as Herbert Simon noted decades ago (1962). When a chimpanzee takes off the top of a box to get at the banana inside, it presumably recognizes that the two parts of the box are discontinuous constituents of a single hierarchical unit which holds the banana. Cognitively it is comparable to separating constituents of language, such as taking apart ‘call up’ in ‘call him up’, or embedding large constructions within expressions like ‘what for’, such as in ‘what did you call him up for?’ Linguists have studied the dependency relations of constituents in great depth in a variety of languages — what can be moved, what can be deleted, what can cross over, etc. — and we can hope that computer simulations will soon be able to model such dependency relations within hierarchical structures.

Hierarchical structures are the bases of recursiveness, and recursion is the central mechanism that makes language infinite via repeated conjoining and embedding. While it is undeniable that there is no longest sentence, the fact remains that most utterances in everyday language are quite short, and statistical approximations to these utterances can be very useful in helping us understand the structure and function of such language.

Another question that has intrigued us a lot in recent years is that of ambiguity. From a CSLE vantage point, an interesting research topic would be to see at which points various types of ambiguities emerge as the most rudimentary languages with the simplest lexicons gradually grow toward the level of complexity of modern languages. Embedded in this topic are several questions concerning a typology of ambiguities in the languages of the world: are there universal ambiguities, how do we typologize them, and how do we predict them from the structures in which they reside? Since ambiguities are at once a robust phenomenon and probably unique to human communication, simulating their emergence can tell us much about the nature of language.

As our last point here, we would like to emphasize the tremendous heterogeneity of language. To get our computer simulations started, it is natural to have small and simple models, with a limited community of members who speak a homogeneous language. However, as the simulations continue, as the members and generations multiply, and as the number of interactions grows very large, we should expect the languages to become greatly diversified and the linguistic behaviors of the speakers increasingly heterogeneous.

The fact that two people are talking with each other by no means leads to the conclusion that they really understand each other, or that they share the same grammar and linguistic representations. As communities become larger and more complex, their speakers become more diverse as well. One school of linguistics once claimed that the central focus of its research was on an ideal speaker-listener situated in a homogeneous community, an attitude that has been criticized as 'monastic.' As the empirical foundations for linguistics have grown, however, there has come to be a fuller and fuller realization of just how much speakers differ from each other, even in the same family. It is such variability, of course, when amplified manifold across time and space, which produces dialects, and eventually distinct languages. It would be a worthy goal for CSLE to eventually be able to simulate such evolutionary processes with realism. Given that the area has been progressing at such an exciting pace, such a goal may not be too far away.

References

Abrams, Daniel M. & Steven H. Strogatz. 2003. "Modelling the dynamics of language death". *Nature* 424 (21 August 2003):900.

Batali, John. 1994. "Innate biases and critical periods: combining evolution and learning in the acquisition of syntax". *Artificial Life IV* ed. by R. Brooks and P. Maes. 160–171. Cambridge, MA: MIT Press.

Cangelosi, Angelo & Domenico Parisi, eds. 2002. *Simulating the Evolution of Language*. Springer: London.

Cangelosi, Angelo, Alberto Greco and Stevan Harnad. 2002. "Symbol grounding and the symbolic theft hypothesis". In Cangelosi & Parisi, 2002, 191–210.

Cavalli-Sforza, Luigi Luca & Marc Feldman, 1981. *Cultural Transmission and Evolution*. Princeton University Press: Princeton.

Cheng, Chin-Chuan. 1998. "Quantification for understanding language cognition". *Quantitative and Computational Studies on the Chinese Language*, ed. by Benjamin K. T'sou, Tom B. Y. Lai, Samuel W. K. Chan & William S-Y. Wang. 15–30. City University Press: Hong Kong.

Clark, R. and I. Roberts. 1993. “A computational model of language learnability and language change”. *Linguistic Inquiry*. 24:299–345.

Deacon, Terrence W. 1997. *The Symbolic Species: the Co-evolution of Language and the Brain*. New York: W.W. Norton.

Dunbar, R. 1993. “Coevolution of neocortical size, group size and language in humans”. *Behavioral and Brain Sciences*, 16:681–735.

Dugatkin, Lee Alan. 2000. *The Imitation Factor : Evolution Beyond the Gene*. New York: Free Press.

Freedman, David A. and William S-Y. Wang. 1996. “Language polygenesis: a probabilistic model”. *Anthropological Linguistics* 104.2:131–138.

Gell-Mann, Murray. 1994. *The Quark and the Jaguar*. W. H. Freeman: New York.

Greenberg, J. H. (ed.).1963. *Universals of Language*. Cambridge, Mass.: MIT Press.

Harnad, Stevan R., Horst D. Steklis, and Jane Lancaster. 1976. *Origins and evolution of language and speech*. New York: New York Academy of Sciences.

Hauser, Marc. 1996. *The Evolution of Communication*, Cambridge, Mass.: MIT Press.

Hockett, C.F. 1960. “The origin of speech”. *Scientific American*. Reprinted in Wang, ed. 1982.

Hurford, James R., Michael Studdert-Kennedy & Chris Knight, eds. 1998. *Approaches to the Evolution of Language: Social and Cognitive Bases*. Cambridge University Press: Cambridge.

Hutchins, Edwin & Brian Hazlehurst. 1995. “How to invent a lexicon: the development of shared symbols in interaction”. *Artificial Societies: the Computer Simulation of Social Life*, ed. by G. N. Gilbert & R. Conte. London: UCL Press.

Jakobson, Roman. 1941/1968. Child Language, Aphasia, and Phonological Universals. The Hague: Mouton.

Jowett, Benjamin, ed. and trans. 1953. *The Dialogues of Plato*, 4th ed. Oxford: Oxford University Press.

Kauffman, Stuart A. 1995. *At Home in the Universe: the search for laws of self-organization and complexity*. New York: Oxford University Press.

Ke, Jinyun & Christophe Coupé. 2002a. “The fall and rise of homophones: a window to language evolution”. Paper presented at the *Fourth International Conference on Language Evolution*. Harvard, March 27–30, 2002.

Ke, Jinyun, James W. Minett, Ching-Pong Au & William S.-Y. Wang. 2002b. “Self-organization and selection in the emergence of vocabulary”. *Complexity* 7:4:1–14.

Komarova, Natalia L., Partha Niyogi & Martin A. Nowak. 2001. “The evolutionary dynamics of grammar acquisition”. *Journal of Theoretical Biology* 209:43–59.

Lee, Sang-Oak, Hyung-Soo Kim & Ik-sang Eom. 2002. “The lexical diffusion of sound change in Korean and Sino-Korean”. *Journal of Chinese Linguistics Monograph 19*.

Marler, P. 1987. “A comparative approach to vocal learning: song development in white-crowned sparrows”. *Journal of Comparative and Physiological Psychology* 71:1–25.

Meltzoff, A.N. 1996. “The human infant as imitative generalist: a 20-year progress report on infant imitation with implications for comparative psychology”. *Social Learning in Animals: the Roots of Culture*. ed. by C.M. Heyes and B.G. Galef, 347370. San Diego: Academic Press.

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

Noble, Jason & D. Cliff. 1995. “On simulating the evolution of communication”. *Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, ed. by Pattie Maes, Maja J. Mataric, Jean-Arcady Meyer, Jordan Pollack, and Stewart W. Wilson. Cambridge, MA: MIT Press.

Nowak, Martin A., David C. Krakauer & A. Dress. 1999. “An error limit for the evolution of language”. *Proceedings of the Royal Society of London* B:266:2131–2136.

Nowak, Martin A., Joshua B. Plotkin, & David C. Krakauer. 1999. “The evolutionary language game”. *Journal of Theoretical Biology* 200:147–162.

Nowak, Martin A., Natalia L. Komarova & Partha Niyogi. 2001. “Evolution of universal grammar”. *Science* 291:114–118.

Ogura, Mieko. 1993. “The development of periphrastic *do* in English”. *Diachronica* 10:1:51–85.

Ogura, Mieko & William S.-Y. Wang. 1996. “Snowball effect in lexical diffusion: the development of *-s* in the third person singular present indicative in English”. *English Historical Linguistics 1994* ed. by Derek Britton, 119–141. Amsterdam: John Benjamins

Schoenemann, P.Thomas. 1999. “Syntax as an emergent characteristic of the evolution of semantic complexity”. *Minds and Machines* 9:309–346.

Schoenemann, P.Thomas and William S-Y.Wang. 1996. “Evolutionary principles and the emergence of syntax”. *Behavioral and Brain Sciences* 19:646–7.

Shen, Zhongwei. 1990. “Lexical diffusion: a population perspective and a mathematical model”. *Journal of Chinese Linguistics* 18:159–201.

Shen, Zhongwei. 1997. “Exploring the Dynamic Aspect of Sound Change”. *Journal of Chinese Linguistics*, Monograph 11.

Sherman, Don. 1975. “Noun-stress alternation: an example of lexical diffusion of sound change in English”. *Linguistics* 159:43–71.

Simon, Herbert A. 1962. “The architecture of complexity”. *Proceedings of the American Philosophical Society* 106:467-482.

Steels, Luc. 1995. “Emergent Adaptive Lexicons”. *Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, ed. by Pattie Maes, Maja J. Mataric, Jean-Arcady Meyer, Jordan Pollack, and Stewart W. Wilson. Cambridge, MA: MIT Press.

Steels, Luc. 2002. “Grounding symbols through evolutionary language games”. In Cangelosi & Parisi, 2002, 211–226.

Steels, Luc and F. Kaplan. 1998. Spontaneous lexicon change. *Proceedings of COLING-ACL'1998*, Montreal, 1243-1249.

Trask, Robert L. 1996. *Historical Linguistics*. London; New York: Arnold.

Tzeng, Ovid & William S.-Y. Wang. 1983. Search for a common neuro-cognitive mechanism for language and movements. *American Journal of Physiology* 246:R904–R911.

Wang, William S.-Y. 1969. Competing change as a cause of residue. *Language* 45:9–25.

Wang, William S.-Y. 1978. Explorations in Language Evolution. Diamond Jubilee Lecture, Osmania University, Hyderabad, India, Dec. 1978. Reprinted in William S.-Y. Wang. 1991. *Explorations in Language*. 105–131. Taipei: Pyramid Press.

Wang, William S-Y., ed. 1982. Human Communication: Language and its Psychobiological Bases : readings from Scientific American. San Francisco: W. H. Freeman.

Wang, William S-Y. 1989. Language in China: a chapter in the history of linguistics. *Journal of Chinese Linguistics* 17:2:183–222.

Wang, William S-Y. and Jinyun Ke. 2003. “Language heterogeneity and self-organizing consciousness,” a commentary on “The self-organizing consciousness” by P. Perruchet and A. Vinter. *Behavioral and Brain Sciences* 25.3:358-359.

Wilson, Edward O. 1972. “Animal communication”. *Scientific American* 227 (September): 52-60.

Wray, Alison, ed. 2002. *The Transition to Language*. Oxford: Oxford University Press.

Language Engineering Laboratory,
Department of Electronic Engineering,
City University of Hong Kong,
Kowloon, Hong Kong, P.R.C.
Correspondence to: Wang
Tel: (+852) 2788 7196
Fax: (+852) 2788 7791
wsyw@ee.cityu.edu.hk (Wang),
jyke@ee.cityu.edu.hk (Ke),
minett@ee.cityu.edu.hk (Minett).