J. theor. Biol. (2000) **203,** 000–000
doi:10.1006/jtbi.2000.1085, available online at http://www.idealibrary.com on IDEAL®

The Basic Reproductive Ratio of a Word, the Maximum Size of a Lexicon

MARTIN A. NOWAK

Institute for Advanced Study, Princeton, NJ 08540, *U.S.A.*

Language is about words and rules. While there is some discussion to what extent rules are learned or innate, it is clear that words have to be learned. Here I construct a mathematical framework for the population dynamics of language evolution with particular emphasis on how words are propagated over generations. I define the basic reproductive ratio of word, R, and show that $R > 1$ is required for words to be maintained in the lexicon of a language. Assuming that the frequency distribution of words follow Zipf's law, an upper limit is obtained for the number of words in a language that relies exclusively on oral transmission.

1. Introduction

Many discussions about language evolution are of a verbal nature, but evolution is a discipline well grounded in mathematical theory. Concepts like *fitness*, *selection* or *mutation* are best described by mathematical equations and in many cases only become clear in the context of specific mathematical models. Thus, it is essential to construct a mathematical theory for language evolution. The aim of the paper is to design such a theory, which describes language propagation and change in a population over many generations. In other words, we are aiming to describe the *population dynamics of language evolution.*

Contributions to the evolution of language have come from many different areas including studies of animal communication and social behaviors (Marler, 1970; Smith, 1977; Cheney & Seyfarth, 1990; Hauser, 1996), diversity and historical change of existing human languages (Bickerton, 1990; Cavalli-Sforza & Cavalli-Sforza, 1995), language development in children (Hurford, 1991; Bates, 1992), the genetic and anatomic correlates of language competence (Lieberman, 1991; Deacon, 1997), and more general models of cultural evolution (Cavalli-Sforza & Feldman, 1981; Boyd & Richerson, 1985; Niyogi & Berwick, 1997). Chomsky (1965, 1975, 1980) argues that all human languages have the same underlying *universal grammar*, which is the product of a *language organ*. Since only humans, but no other animals have universal grammar, Chomsky suggests that the language organ might have evolved for other purposes and was later accidentally taken over for language. Pinker (1995) notes that language is an enormously complicated trait and as such could only have arisen gradually and by natural selection, rather than being the by-product of some other process (see also Pinker & Bloom, 1990). Maynard Smith & Szathmary (1995) describe human language as a major transition in evolution. For a recent review on approaches to language evolution see Hurford *et al.* (1998).

Our own previous work in this area was devoted toward constructing a mathematical theory which describes how basic features of human language can evolve gradually and by natural selection from simple precursors. Specifically, we have explored how associations between signals and objects can evolve, how word formation can

0022–5193/00/000000 + 00 $35.00/0

overcome a linguistic error limit and how simple grammatical rules emerge as mechanisms to reduce mistakes in communication (Nowak & Krakauer, 1999; Nowak *et al*., 1990a, b). Other work explored the spatial dynamics of the evolutionary language game (Grassly *et al*., 2000).

The size of large English dictionaries has grown roughly exponentially over the last 400 years. Robert Cawdrey's "Table Alphabeticall" published in 1604 listed 2500 words, Samuel Johnson in 1755 conceived definitions for 40 000 words, James Murray's Oxford English Dictionary in 1928 had 400 000 entries. But many of these 400 000 entries are compounds whose meanings can be derived. In an attempt to estimate the number of words known by a person, a recent study by Nagy and Anderson begins with about 228 000 words, but only 89 000 of these are either simple roots, stems or compounds whose meaning cannot be derived. They showed that the average American high school graduate knows about 45 000 of these words. This is an underestimate because proper names, foreign words, acronyms and numbers were excluded. The actual number maybe closer to 60 000. If word learning begins at around 1 year of age, then a 17-year-old high school graduate must have learned about 10 new words per day for 16 years. An average 6-year-old knows about 13 000 words. Shakespeare used 15 000 words.

A formal definition of the concept "word" is not easy, and hence counting the number of words in a language or the number of words known to a person is not straightforward. When pressed for a definition, linguists may say that words are the units of language which are the products of morphological rules and which are unsplittable by syntactic rules (see Pinker, 1995; Miller, 1991). In a word, a word is a syntactic atom. A very different definition describes a word as a memorized chunk: a string of linguistic stuff that is arbitrarily formulated with a particular meaning (Pinker, 1995). A word is a "listeme", an item of the long list called "mental dictionary". It is this second concept that we have in mind when writing this paper.

In particular, we will analyse how words are being maintained in a population. Each individual has an internal lexicon. Children acquire this internal lexicon by learning words from their parents or other individuals. Not every individual knows all the words of a language. There is always a chance that children will miss out on some rare words. The average waiting time for a specific rare word to be uttered by the parent could exceed the duration of the childhood. An interesting question is, what is required for a (rare) word to be maintained in the lexicon of a language that relies on oral transmission? Note that writing is, of course, a very recent invention, and, hence, for most of the time language evolution occurred without the aid of written records.

We will explore various models to answer this question. The models will differ in the basic assumptions such as whether children learn the language from one or several individuals and whether language contributes to biological fitness or the chance to act as a language teacher. For all such models we find that the *basic reproductive ratio of a word* is given by

$$R = BQ,$$

where B is the average number of language teachers per child and Q is the probability that a child will learn the word from any one teacher. (Because of symmetry, B is also the average number of children that learn the language from any one teacher.) For a new word to invade a population from a very low initial abundance, we require $R > 1$. For a word to be stably maintained in the lexicon of a population, $R > 1$ is for all models a sufficient condition and for some models also a necessary condition. An immediate consequence is that learning the language from only one individual, $B = 1$, can never lead to a basic reproductive ratio above 1 (since $Q \leq 1$). In this case, words can never invade when rare, but have to overcome a minimum initial level of abundance (an *invasion barrier*).

We will then relate the probability, Q, of acquiring a word from a language teacher to the frequency of occurrence of this word in spoken language. This will allow us to formulate a condition for the minimum frequency of occurrence that is required for a word to be maintained in a language.

Assuming that word frequency distributions follow Zipf's law (Estoup, 1916; Zipf, 1935; Mandelbrot, 1958), which states that words have a frequency inversely proportional to their position in a frequency ranking, we derive the maximum number of words, n_{max}, that can be maintained in a lexicon. We find that n_{max} is implicitly given by the equation

$$n_{max} \ln n_{max} = BZq,$$

where B is the number of teachers, Z is the total number of words told to the child by one teacher and q is the probability to memorize a word after a single encounter.

The core of the paper is in three parts. Section 2 devises models for children learning the language of their parents. Section 3 explores language learning from several individuals. Section 4 analyses the maximum size of a lexicon. In Section 2, we will encounter a quasispecies equation (although with frequency dependent fitness values). In Section 3, we will define the basic reproductive ratio of a word and find equations that are reminiscent of epidemiological theory and models of cultural evolution.

2. Learning the Language of your Parents

In this section, we will outline the basic model for the situation where children learn the language of their parents. For simplicity, we will start by assuming the children acquire the language of only one parent, later we will extend the model to two parents. There is a qualitative difference between these two cases. Learning the language from only one individual is like asexual reproduction. Learning from two individuals is like sexual reproduction. Sexual reproduction allows recombination. If a child learns the language from only one individual then the child can only build up a lexicon which is a subset of the lexicon of the teacher (unless the child invents new words). If a child learns from two or more individuals, then the lexicon of the child need not be a subset of the individual lexica of either teachers: new combinations of words can emerge.

There is also an interesting link to Muller's ratchet in genetics. Muller's ratchet states that in a small, asexual population genetic information will disappear over time because genes will accumulate deleterious mutations (Muller, 1964). In a sexual population, however, recombination can work against the accumulation of deleterious mutations. A similar phenomenon applies to language: in a small population, learning words from only one teacher will lead to a decline in the total vocabulary over time; learning from several teachers can in principle maintain a stable number of words.

In the paper, we will sometimes refer to the *frequency of occurrence* of a word and sometimes to the *relative abundance* of a word in the population. By frequency of occurence we mean how often the word is used by an individual who knows the word. By relative abundance of a word we denote the fraction of people who know the word.

Throughout the paper we assume that an individual either knows a word or does not know it. There cannot be varying degrees of knowledge. This is a simplification. One can imagine that a word can be more or less strongly incorporated into the mental lexicon of an individual and this could also affect the frequency at which this individual uses the word. Extending our models into these directions seems a promising task for future research.

2.1. INTERNAL LEXICA AND PAYOFF

Suppose there are n words. Individuals are characterized by binary strings, S, which indicate all the words known to a given individual. Thus, the binary string, S, describes the *internal* or *mental lexicon* of an individual. There are $v = 2^n$ many internal lexica, which we label $j = 0, \ldots, v-1$ (corresponding to the integer representation of the binary string). Lexicon j is defined by the binary string

$$S_j = (s_j(1), s_j(2), \ldots, s_j(n)). \tag{1}$$

Here $s_j(i) \in \{0, 1\}$. If lexicon j contains word i then $s_j(i) = 1$, otherwise $s_j(i) = 0$. The total number of words in lexicon j is given by

$$n_j = \sum_{i=1}^{n} s_j(i). \tag{2}$$

The payoff for an individual with lexicon j conversing with an individual with lexicon k is given by the number of words they have in common,

$$F(S_j, S_k) = \sum_{i=1}^{n} s_j(i) s_k(i). \tag{3}$$

The payoff function is based on the simplifying assumption that each word contributes the same amount (of 1 point) to the overall payoff. Other assumptions are possible. Some words will be used much more frequently than others. Hence, the payoff for having a common word could be proportional to its frequency of occurrence, ϕ_i:

$$F(S_j, S_k) = \sum_{i=1}^{n} s_j(i) s_k(i) \phi_i. \tag{4}$$

Equation (3) and (4) both assume that the interaction between individuals j and k consists of j talking to k and k talking to j: that is the interaction is symmetric.

Following the standard assumption of evolutionary game theory, we interpret the payoff received in the game as fitness (Maynard Smith, 1982; Hofbauer & Sigmund, 1998). Thus, individuals that communicate well produce more offspring.

2.2. A QUASISPECIES EQUATION FOR INTERNAL LEXICA

Children inherit (or learn) the language of their parents, but they can miss out on certain words. Denote by Q the probability that a specific word is transmitted from parent to child. Let us neglect the possibility that children invent new words (which is a rare event). Thus, we can define Q_{jk}, the probability that a parent with lexicon S_j gives rise to an offspring with lexicon S_k, as

$$Q_{jk} = Q^{n_k}(1 - Q)^{n_j - n_k}. \tag{5}$$

This equation holds for all lexica S_j and S_k where $s_k(i) = 1$ implies that $s_j(i) = 1$. Otherwise, $Q_{jk} = 0$. This just means that it is possible to lose words but not to gain words.

Denote by x_j the abundance of individuals with lexicon S_j. We have $\sum_{j=0}^{v-1} x_j = 1$. The abundance of each lexicon S_j changes over time according to

$$\dot{x}_j = \sum_{k=0}^{v-1} x_k f_k Q_{kj} - \Phi_{kj}, \quad j = 0, \dots, v - 1. \tag{6}$$

Here, f_k denotes the fitness of lexicon S_k and is given by

$$f_k = \sum_{j=0}^{v-1} x_j F(S_k, S_j). \tag{7}$$

The average fitness of the population is given by

$$\Phi = \sum_{j=0}^{v-1} x_j f_j. \tag{8}$$

Note that eqn (6) implies that each child acquires the language from one parent. We will extend this framework later. Equation (6) is reminiscent of the quasispecies equation devised by Eigen & Schuster (1978) for describing molecular evolution and theories for the origin of life. The main difference, however, is that our model has fitness values that are frequency dependent, while the standard quasispecies theory has constant fitness values. In the context of quasispecies theory, x_i denotes the frequency of nucleic acid i which is given by a sequence of bases of length n. The matrix Q contains the mutation rates; Q_{kj} is the probability that replication of sequence k results in sequence j. In contrast to our language model, quasispecies theory usually has symmetric mutation matrices. Standard quasispecies dynamics can be seen as an optimization procedure on a constant fitness landscape, our language model describes mutation and adaptation on a changing fitness landscape.

We are interested in calculating any potential limit on the total number of words that can be stably transmitted from one generation to the next in the system described by eqn (5)–(8). The next section shows a simple way how to do this.

2.3. THE EVOLUTIONARY STABILITY OF WORDS

Suppose all individuals of a population have $n - 1$ words in common and differ only in whether they have or have not the n-th word. Denote by x_1 the frequency of individuals who have the n-th word and by x_0 the frequency of individuals who do not have the n-th word in their mental lexicon. We can write

$$\begin{aligned} \dot{x}_1 &= x_1 f_1 Q - x_1 \Phi, \\ \dot{x}_0 &= x_1 f_1 (1 - Q) + x_0 f_0 - x_0 \Phi. \end{aligned} \tag{9}$$

Assuming that each word contributes the same amount to the payoff, the fitness of x_1 individuals is given by $f_1 = f + nx_1 + (n-1)x_0 = f_0 + x_1$. The fitness of x_0 individuals is given by $f_0 = f + n - 1$. Here, f is the background fitness independent of language. Since $x_1 + x_0 = 1$ the system is described by a differential equation in one variable:

$$\dot{x}_1 = x_1[-x_1^2 + x_1Q - (n-1+f)(1-Q)]. \tag{10}$$

The potential equilibria are given by

$$x_1 = \frac{Q}{2}[1 \pm \sqrt{1 - 4(1-Q)Q^-2(f+n-1)}]. \tag{11}$$

The equilibria exist if

$$n < \frac{Q^2}{4(1-Q)} - f + 1. \tag{12}$$

We see that the lexicon size, n, is limited. The system cannot maintain an arbitrarily large number of words. If condition (12) is fulfilled then eqn (10) has two equilibria. The smaller equilibrium is unstable and represents an invasion barrier, the larger equilibrium is stable and represents the equilibrium frequency of individuals knowing the n-th word. The smaller n is compared to $Q^2/[4(1-Q)] - f$, the smaller the invasion barrier and the larger the stable equilibrium frequency of x_1. The invasion barrier implies that a new word cannot invade a population (cannot spread in abundance) if only a very small fraction of the population uses the word. It can only invade if the fraction of people using the word has overcome the invasion barrier.

Condition (12) essentially implies that for maintaining a reasonable number of words in the population, the probability Q has to be very close to 1. The reason for having an upper limit for the number of words in this model is that as the number of words increases, each word contributes smaller proportions to the fitness of an individual. As n becomes very large, the difference between f_0 and f_1 becomes negligible. In Section 4, we will also derive a limit for the maximum number of words in the combined mental lexicon of a population, but this limit will be independent of fitness considerations.

Figure 1 shows a computer simulation which is an application of eqn (6) to finite populations. The total population size is 100. There are $n = 10$

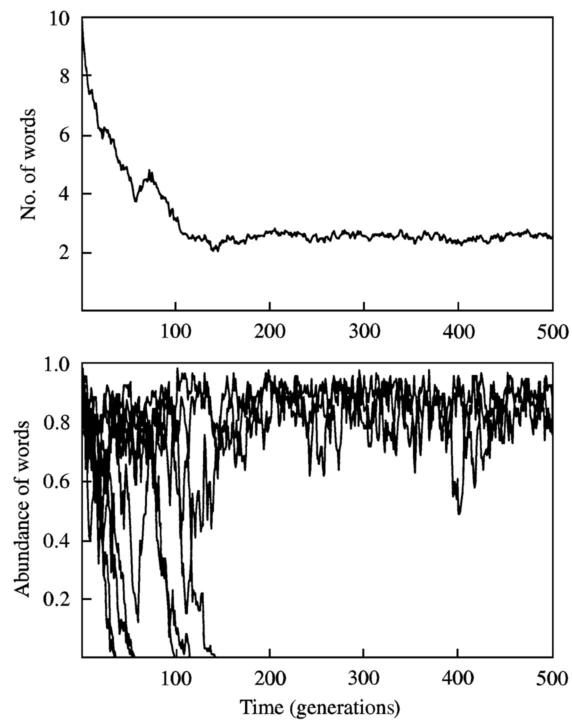

FIG 1. Only a limited number of words can be stably propagated from one generation to the next. In this simulation, children learn the language from one parent. Initially, all individuals have $n = 10$ words. The probability that a word is transfered from parent to offspring is $Q = 0.95$. The population consists of 100 individuals, described by binary strings on length 10. The fitness of each individual is evaluated according to eqn (3). Individuals are chosen for reproduction proportional to their fitness. Each new offspring replaces another, randomly chosen, individual. Hence, the population size remains constant. The language independent, background fitness if $f = 1$. Thus, eqn (12) suggests that the stable number of words cannot exceed 4. We find that after 500 generations, only 3 words are left. Some individuals have fewer words. On average there are 2.53 words per individual (averaged over all individuals between $t = 400$ and 500). The small population size cannot maintain the maximum possible number of words, 4. The last three words are stable for a long period. (Note, however, that all words will be lost eventually. This is simply a consequence of the finite population size and the fact that words once lost do not reappear in this simulation.) (a) Average number of words per individual over time. (b) Abundances of individual words in the population. After about 150 generation, seven words have become extinct, the remaining three fluctuate around abundance levels of about 80%.

words. Each individual is described by a bit string of length 10. The fitness of each individual is determined by summing up the payoffs obtained in all interactions with all other individuals. Individuals generate children proportional to their fitness. Each child replaces a randomly chosen individual; thus the population size is constant. Each child learns the words of its parent. The probability that a given word is passed on is $Q = 0.95$. The population shows 500 generations. At the end, three words are stably propagated in the population. Note that eqn (12) predicts four words as an upper limit. Presumably, the small population size eliminated the 4th word in a random fluctuation. Three words are more stable that four words, because then each word contributes relatively more to the overall fitness of an individual.

2.4. ACQUIRING THE LANGUAGE FROM TWO PARENTS

Sofar each individual had a single parent for the purpose of language learning. We can move closer to reality by admitting the possibility that individuals can acquire their language from two parents. In analogy to eqn (9) we write

$$\dot{x}_1 = f_1^2 x_1^2 [1 - (1 - Q)^2] + 2 f_0 f_1 x_0 x_1 Q - \Phi x_1. \tag{13}$$

Since $x_0 = 1 - x_1$ it is not necessary to write the equation for $\dot{x}_0$. We have $f_0 = f + n - 1$, $f_1 = f_0 + x_1$ and $\Phi = (f_0 + x_1^2)^2$. The main new feature of eqn (13) is that words can spread from very low abundances in the population. This was not possible for the models discussed in Sections 2.2 and 2.3, where there is always an invasion barrier (a minimum level of abundance) which has to be overcome. In eqn (13), the x_1 individuals increase from very low levels provided

$$Q > 1/2. \tag{14}$$

Thus, acquiring the language from two rather than from one individual adds a qualitatively new feature: words can invade populations starting from arbitrarily low frequency. This brings us straight to the next section.

3. Learning the Language of Others

Let us consider a system where language is not acquired only from the parents, but from several different individuals (perhaps including the parents). As in Section 2, we assume that there are n words and lexica are characterized by bit strings of length n defining whether or not they have a certain word. Consider the following system of equations:

$$\dot{x}_k = \delta_k - \Phi x_k + b \sum_{i=0}^{v-1} \sum_{j=0}^{v-1} (x_i x_j Q_{ijk} - x_k x_i Q_{kij}),$$
$$k = 0, \ldots, v - 1. \tag{15}$$

Here $\delta_0 = \Phi$ and $\delta_k = 0$ for $k > 0$. Again the total population size is scaled to one: $\sum_k x_k = 1$. The average fitness of the population is given by $\Phi = \sum_k f_k x_k$, where f_k is the fitness of individual k and is calculated as in Section 2. Each newborn individual is in the x_0 class, that is it does not know any word. Individuals learn by interacting with other individuals. The rate of this interaction is b, and Q_{ijk} denotes the probability that an individual with lexicon S_i learning from an individual with lexicon S_j will end up with lexicon S_k.

Q_{ijk} is constructed as follows: (i) If lexicon S_i has a certain word then lexicon S_k will have it, too. (It is not possible to lose words.) (ii) If S_i does not have a certain word, but S_j has it, then S_k will have it with a probability Q. This process describes the learning of new words. (iii) If neither S_i nor S_j have a certain word, then S_k will also not have it.

As before, we are interested in the conditions for maintaining rare words in the population. These conditions can be calculated by analysing a simple model that describes the dynamics of each word separately.

3.1. THE BASIC REPRODUCTIVE RATIO OF WORDS

In this section, we derive a basic reproductive ratio for words. Writing eqn (15) for only one word, we get

$$\dot{x}_0 = \Phi(1 - x_0) - bQx_0x1,$$
$$\dot{x}_1 = -\Phi x_1 + bQx_0x_1. \tag{16}$$

Clearly, this is just a one-dimensional system, since $x_0 + x_1 = 1$. We have $\Phi = f_0x_0 + f_1x_1$, $f_0 = f + n - 1$ and $f_1 = f_0 + x_1$. We find that x_1 can increase from low values and converge to a stable equilibrium provided

$$(b/f_0)Q > 1. \tag{17}$$

The ratio b/f_0 denotes the number of language teachers an individual has through its lifetime. Let us write $B = b/f_0$. Then the condition becomes

$$BQ > 1. \tag{18}$$

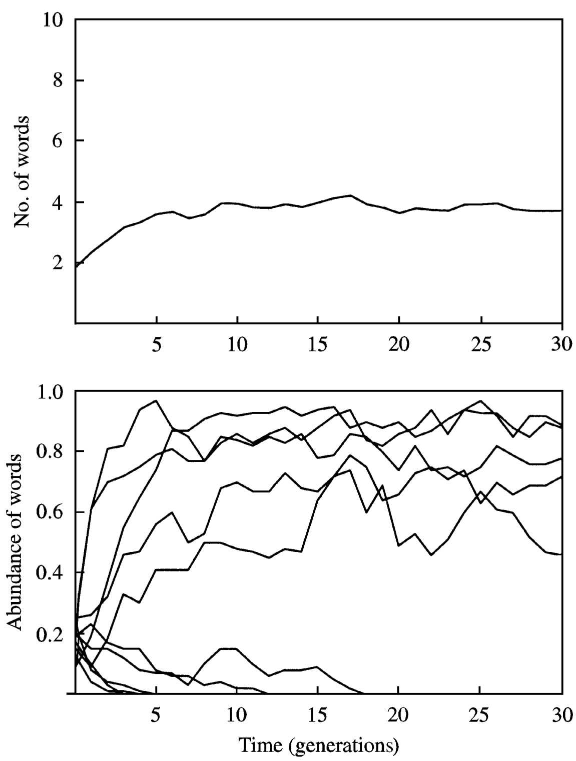

FIG 2. If children learn words from more than one individual, then the basic reproductive ratio, R, of words can exceed one. This means that words can increase in abundance from low starting frequencies; there is no invasion barrier. In this simulation, there are $n = 10$ words and 100 individuals. Each child learns words from two individuals. In the notation of Section 3.1, this means $B = 2$. The probability of learning word i from any one teacher ranges from $Q_1 = 0.25$ to $Q_{10} = 0.75$. Words 1–5 have R values below 1, while words 6–10 have R values above 1. Initially, words have randomly chosen starting abundances in the population of roughly 0.2. This means a given word is known by 1 in every 5 individuals. (a) The average number of words per individual increases over time from 2 to about 4. (b) Words with $R_i > 1$ remain, while words with $R_i < 1$ become extinct.

The product, BQ, is the basic reproductive ratio, R, of a word; it is the product of the number of language teachers times the probability to acquire the word from one language teacher. There is, of course, a symmetry between teachers and students and B can be seen as the total number of students an individual has. Hence, intuitively and in complete analogy to epidemiology, the basic reproductive ratio of a word is the number of individuals that acquire the word from any one individual in the limit where almost no one has the word. Note also that $BQ > 1$ is in agreement with $Q > \frac{1}{2}$ which is the crucial condition in Section 2.4, where each individual learns from two parents.

Figure 2 shows results of a computer simulation for a finite population. There are 100 individuals. Each new individual replaces a randomly chosen individual and learns the language from two other individuals that are chosen according to their payoff. Hence, $B = 2$. As before, the payoffs are evaluated by each individual talking to every other individual. There are $n = 10$ words; each individual is characterized by a bit string of length 10. The probability that word i is transmitted from a teacher, who has word i, to a student, who does not have word i, is Q_i. After 50 generations, the five words with $R > 1$ remain in the population whereas the five words with $R < 1$ have become extinct.

3.2. THREE OTHER SYSTEMS THAT LEAD TO THE SAME BASIC REPRODUCTIVE RATIO

Equations (15) and (16) assume that reproduction is proportional to the payoff achieved in the language game, but the choosing of teachers is not. We could assume that individuals who talk well not only have more children but are also more likely to be chosen as teachers for language learning. In analogy to eqn (16) we can write

$$\begin{aligned} \dot{x}_0 &= \Phi(1 - x_0) - bQx_0x1f_1/\Phi, \\ \dot{x}_1 &= -\Phi x_1 + bQx_0x_1f_1/|\Phi. \end{aligned} \tag{19}$$

Remember that $\Phi = f_0x_0 + f_1x_1$. Thus, x_1 individuals are chosen as teachers proportional to their frequency, x_1, times their relative payoff, f_1/Φ. As before, we find that the basic reproductive ratio of x_1 individuals is given by $R = BQ$, where $B = b/f_0$ is the number of language teachers per individual.

Alternatively, we could assume that the payoff of the language game only affects the probability to be chosen as language teacher and has no effect on biological reproduction:

$$\begin{aligned} \dot{x}_0 &= 1 - x_0 - bQx_0x1f_1/\Phi, \\ \dot{x}_1 &= - x_1 + bQx_0x_1f_1/\Phi. \end{aligned} \tag{20}$$

Again the basic reproductive ratio is given by $R = BQ$, but this time $B = b$ denotes the number of teachers per individual (because here the life expectation of an individual is 1 rather than $1/f_0$.)

Finally, we could assume that the payoff of the language game neither confers biological fitness nor has an effect on being chosen as language teachers. In this case we obtain

$$\begin{aligned} \dot{x}_0 &= 1 - x_0 - bQx_0x1, \\ \dot{x}_1 &= - x_1 + bQx_0x_1. \end{aligned} \tag{21}$$

This system is equivalent to the basic model of infection dynamics (see Anderson & May, 1991) which can also be used for describing the spread of cultural innovations (Cavalli Sforza & Feldman, 1981). For eqn (21), the basic reproductive ratio of a word is $R = BQ$ with $B = b$. The equilibrium is given by $x_0 = 1/R$ and $x_1 = 1 - 1/R$.

All models lead to the same basic reproductive ratio of a word, but will differ in the equilibrium abundances of the word. The last two models do not relate the payoff of the language game to biological reproduction. The last model even makes no reference to fitness at all; individuals are *language eaters* that acquire as many words as they can. While the simplicity of such a model is useful for mathematical analysis it is important to note that somewhere there must be the (hidden) assumption that language confers biological fitness. Otherwise, individuals who do not care about language would not be selected against. We can, however, deal with this issue in a two-step approach. We can first study a model that shows that everyone has to care about language (for such a model, language must relate to biological fitness), and afterwards we simply study a model, like eqn (21), where everyone already has a language acquisition device. The next section presents a simple model for the evolution of a language acquisition device.

3.3. SELECTION OF LANGUAGE ACQUISITION

In this section, we compare two types of individuals that differ in their readiness to learn language. Specifically, they differ in the rate at which they acquire language teachers. We will observe that the "faster language eaters" win.

Denote by x_i and y_i individuals that have, respectively, $A = a/\Phi$ and $B = b/\Phi$ language teachers. Let $a > b$. Consider the following system:

$$\begin{aligned} \dot{x}_0 &= f_0x_0 + f_1x_1 - aQx_0(x1 + y_1) - \Phi x_0, \\ \dot{x}_1 &= aQx_0(x1 + y_1) - \Phi x_1, \\ \dot{y}_0 &= f_0y_0 + f_1y_1 - bQy_0(x1 + y_1) - \Phi y_0, \\ \dot{y}_1 &= bQy_0(x1 + y_1) - \Phi y_1. \end{aligned} \tag{22}$$

Here x_0 and y_0 denote individuals that do not have the n-th word, while x_1 and y_1 do have the n-th word. We have $\Phi = f_0(x_0 + y_0) + f_1(x_1 + y_1)$, $f_0 = f + n - 1$ and $f_1 = f_0 + x_1 + y_1$.

If $(a/f_0)Q > 1$ then the system converges to the boundary where $x_1 > 0$, $x_0 = 1 - x_1$ and $y_0 = y_1 = 0$. Here the "slower language eaters" become extinct.

If $(a/f_0)Q < 1$ then the system converges to the boundary where $x_0 + y_0 = 1$ and $x_1 = y_1 = 0$. The n-th word becomes extinct. We have to repeat the same analysis for the $(n - 1)$-st word, denoting by x_1 and y_1 individuals who do have the $(n - 1)$-st word and by x_0 and y_0 individuals who do not. In this case $f_0 = f + n - 2$ and $f_1 = f_0 + x_1 + y_1$. If we have now $(a/f_0)Q > 1$ then the slow language eaters become extinct. If instead $(a/f_0)Q < 1$ then the $(n - 1)$-st word becomes extinct and we turn to the $(n - 2)$-nd word and so on. Ultimately, there are only two possibilities: either all words become extinct or the slow language eaters become extinct. If language persists there is selection for fast language acquisition.

4. The maximum size of a lexicon

In the previous section, we showed that the basic reproductive ratio being greater than one was required for persistence of a word in the lexicon of a population. The basic reproductive ratio of a word is given by

$$R = BQ, \tag{23}$$

where B is the number of language teachers (or students) per individual and Q the probability that the student will acquire the word provided the teacher knows it. The probability Q will certainly depend on the frequency of occurrence, ϕ, of the word in the teacher's language and on the total number of words, Z, the teacher says to the student during the student's who language learning period. Let us suppose that q is the probability that the student memorizes the word after a single encounter. We can write

$$Q = 1 - (1 - \phi q)^Z. \tag{24}$$

This relation assumes that a word once memorized cannot be forgotten. Furthermore, attempts to memorize the word are independent of each other. From $BQ > 1$ we obtain

$$\phi > \frac{1}{q}\left[1 - \left(1 - \frac{1}{B}\right)^{1/Z}\right] \approx \frac{1}{BZq}. \tag{25}$$

The approximation holds for B well above 1. Hence, we find that for a word to have a basic reproductive ratio above 1 its frequency of occurrence must exceed $1/(BZq)$. This result is quite intuitive: BZ is the total number of words the student hears during his/her language acquisition period from all teachers. BZq is the number of all words that are heard and memorized by the student (frequent words will be "memorized" more than once). For a word to be memorized it must occur at least once in this set of BZq words. Hence, its frequency of occurrence must be at least $1/(BZq)$.

4.1. USING ZIPF'S LAW

Let us now consider a language with n words with frequencies $\phi_1 > \phi_2 > \cdots > \phi_n$. For all n words to be maintained in the language, we require $\phi_n > 1/(BZq)$. The least frequent word must satisfy inequality (25).

An empirical observation is that human languages have word frequency distributions which roughly follow Zipf's law (Zipf, 1935):

$$\phi_i = C_n/i. \tag{26}$$

The frequency of a word is indirectly proportional to its position in a frequency rank ordering. Plotting $\log(\phi_i)$ vs. $\log i$ gives a straight line with a slope very close to -1. While this observation is surprising at first sight, it still remains to be determined whether there is a deeper reason behind it. Miller & Chomsky (1963) point out that Zipf's law is almost like a null hypothesis; a source that randomly emits letters and spaces will also generate word frequencies that follow Zipf's law. In any case, we will use Zipf's law for the distribution of word frequencies.

Since $\sum_{i=1}^{n} \phi_i = 1$ we have $C_n = 1/\sum_{i=1}^{n} (1/i)$. For reasonable large n we obtain approximately $C_n \approx 1/(\ln n + \gamma)$, where $\gamma = 0.5772...$ is Euler's constant. Hence, the frequency of the least frequent word is given by

$$\phi_n = \frac{1}{n(\ln n + \gamma)}. \tag{27}$$

Combining eqn (25) with eqn (27), we obtain that the maximum number of words that can be maintained in a language, n_{max}, is implicitly given by

$$n_{max}(\ln n_{max} + \gamma) = BZq. \tag{28}$$

Table 1 shows n_{max} for various choices of BZq. For example, for maintaining about 10^5 words in the lexicon of a spoken language, we require BZq to be about 10^6. If for example $q = 0.1$ (that is the average probability of memorizing a word after one occurrence is 0.1) then $BZ = 10^7$. Thus, a child has to hear a total of 10^7 words. In a childhood lasting 10^8 s this means 1 word every 10 s.

If we do not use Zipf's law but instead assume that all words occur at the same frequency, $1/n$, then n_{max} is simply given by BZq, which again seems intuitively obvious.

TABLE 1

How many words does a child have to hear for a population to maintain a certain number of words in its lexicon?

n_{max}	BZq	BZ ($q = 0.1$)
10	29	290
100	520	5200
1000	7500	75 000
10 000	98 000	980 000
100 000	1 200 000	12 000 000

According to eqn (24), the frequency of the least frequent word, ϕ_n, has to exceed $1/(BZq)$. Assuming that words follow a frequency distribution as in Zipf's law, $\phi_i = C_n/i$ we find $C_n/n > 1/(BZq)$ where $C_n = \sum_{i=1}^{n}(1/i)$. The maximum number of words that can be maintained, n_{max}, is the largest integer which fulfills this inequality. B is the number of teacher of each child, Z is the total number of words from each teachers, and q is the probability that a child remembers a word after one occurrence. In some, sense BZq is the total number of words that are memorized by the child, while BZ is the total number of words a child hears from all its teachers together. For maintaining 10^5 words in a language, we need BZq to be about 10^6. If $q = 0.1$ this implies $BZ = 10^7$. There are about 10^8s in a childhood.

4.2. FREQUENCY AND ABUNDANCE

For the simple model given by eqn (29), the equilibrium abundance of a word in the population is given by $1 - 1/(BQ)$. Together with ( ), we can study the relation between the frequency of occurrence of a word and its abundance in the population. Recall that "frequency of occurrence" describes how often the word is used by a speaker who knows the word, while "abundance" refers to the fraction of individuals who know (and use) the word. Let us consider the abundance of a word *relative* to the abundance of a word with $Q = 1$. (This is a renormalization to ignore the fact that some individuals do not know the word because they have not had any language training so far.) Thus, we have for the relative abundance of word i.

$$x_i = \frac{1 - 1/(BQ_i)}{1 - 1/B}, \tag{29}$$

where $Q_i = 1 - (1 - q\phi_i)^Z$. Figure 3 shows a plot of x_i vs. i. There is a shoulder followed by a linear decline. The most frequent words are known to everyone. Then the fraction of people who know the word falls as a linear function of the rank of the word.

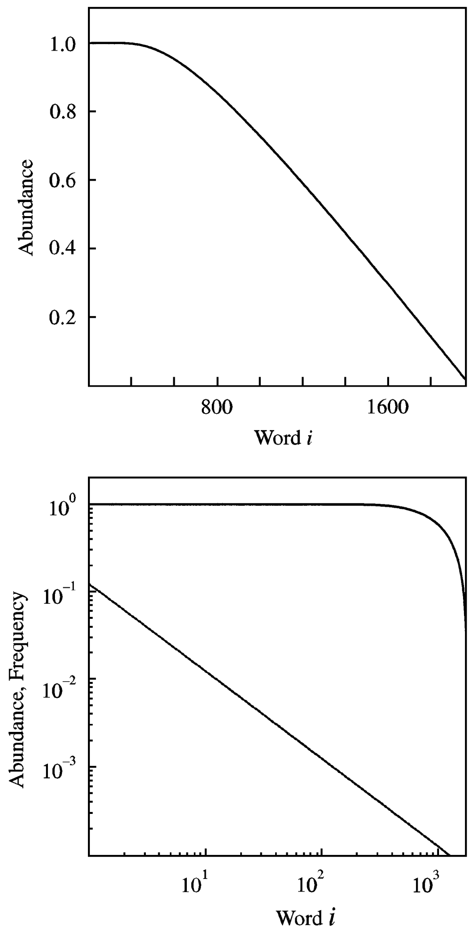

FIG 3. Abundance and frequency vs. word rank. The abundance of a word in the population is the fraction of people who know the word. The frequency of a word is its probabilty of occurrence in the speech of someone who knows the word. Words are labeled $i = 1, \dots, n$ according to decreasing frequency of occurrence, $\phi_1 > \cdots > \phi_n$. We assume Zipf's law: $\phi_i = C_n/i$. The rescaled abundance is given by $x_i = [1 - 1/(BQ_i)]/[1 - 1/B]$ where $Q_i = 1 - (1 - q\phi_i)^Z$. In this example, we have $B = 2$, $q = 0.1$ and $Z = 10^5$. Observe that the most frequent words are known to everyone. After this initial shoulder there is a linear decline. In a log–log plot, the frequency is, of course, a straight line with slope -1, while the abundance is absolutely flat initially and then falls off sharply.

5. Conclusion

In this paper, we have devised models for the population dynamics of language evolution. For various models we have calculated the basic reproductive ratio of a word, R, and shown that

$R > 1$ is the crucial condition for words to be maintained in the lexicon of a population. Assuming a simple model for memorizing words, we calculated the minimum frequency of occurrence of a word, ϕ, required for its survival. Specifically, we find that ϕ has to exceed $1/(BZq)$, where B is the number of language teachers per individual, Z the total number of words from any one teacher and q the probability of memorizing a word after a single encounter. Using Zipf's law for the distribution of word frequencies we find that the maximum size of a lexicon is approximately given by

$$n_{max} \ln n_{max} = BZq.$$

There are many possibilities to extend this line of research. Most importantly, more sophisticated models for memorizing words should be used. In a mental lexicon, words are not independent units but will affect each other. Words are stored in certain classes. Memorizing a word is not an all or nothing process. Different types of information have to be stored for each word: its meaning, its possible positions in a sentence, its relation to other words. All these factors will influence how the probability Q is related to the frequency of occurrence of a word. Furthermore, internal lexica will then not be described by bitstrings but by more complicated mathematical structures. In this regard, the present paper is only a first step.

Support from the Leon Levy and Shelby White Initiatives Fund, the Florence Gould Foundation, the J. Seward Johnson, Sr. Charitable Trusts, the Ambrose Monell Foundation and the Alfred P. Sloan Foundation is gratefully acknowledged.

REFERENCES

ANDERSON, R. M. & MAY, R. M. (1991). *Infectious Diseases of Humans*. Oxford: Oxford University Press.

BATES, E. (1992). Language development. *Curr. Opinion Neurobiol.* **2,** 180–185.

BICKERTON, D. (1990). *Species and Language*. Chicago: Chicago University Press.

BOYD, R. & RICHERSON, P. J. (1985). *Culture and the Evolutionary Process*. Chicago: The University of Chicago Press.

CAVALLI-SFORZA, L. L. & FELDMAN, M. W. (1981). *Cultural Transmission and Evolution*. Princeton, NJ: Princeton University Press.

CHENEY, D. L. & SEYFARTH, R. M. (1990). *How Monkeys See the World: Inside the Mind of Another Species*. Chicago: Chicago University Press.

CHOMSKY, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

CHOMSKY, N. (1975). *Reflections on Language*. New York: Pantheon Press.

CHOMSKY, N. (1980). *Rules and Representations*. New York: Columbia University Press.

DEACON, T. (1997). *The Symbolic Species*. Allen Lane: The Penguin Press.

ESTOUP, J. B. (1916). *Gammes Stenographique*, 4th Edn. Paris.

HAUSER, M. (1996). *The Evolution of Communication*. Cambridge, MA: The MIT Press.

HOFBAUER, J. & SIGMUND, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press.

HURFORD, J. R. (1991). *Cognition* **40,** 159–201.

HURFORD, J. R., STUDDERT-KENNEDY, M. & KNIGHT, C. (1998). *Approaches to the Evolution of Language*. Cambridge: Cambridge University Press.

LIEBERMAN, P. (1991). *Uniquely Human*. Cambridge, MA: Harvard University Press.

MANDELBROT, B. (1958). Les lois statistique macroscopiques du comportment. *Psychol. Francaise* **3,** 237–249.

MARLER, P. (1970). Birdsong and speech development: could there be parallels? *Am. Sci.* **58,** 669–673.

MAYNARD SMITH, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

MAYNARD SMITH, J. & SZATHMARY, E. (1995). *The Major Transitions in Evolution*. New York: Freeman.

MILLER, G. A. & CHOMSKY, N. (1963). Finitary models of language users. In: *Handbook of Mathematical Psychology*, Vol. 2. (Luce, R. D., Bush, R. & Galanter, E., eds), New York: Wiley.

MILLER, G. A. (1991). *The Science of Words*. New York: Scientific American Library.

MULLER, H. J. (1964). The relation of recombination to mutational advance. *Mutat. Res.* **1,** 2–9.

NIYOGI, P. & BERWICK, R. C. (1997). A dynamical systems model for language change. *Complex Syst.* **11,** 161–204.

NOWAK, M. A. & KRAKAUER, D. C. (1999). The evolution of language. *Proc. Nat. Acad. Sci. U.S.A.* (in press).

NOWAK, M. A., PLOTKIN, J. & KRAKAUER, D. C. (1999a). The evolutionary language game. *J. theor. Biol.* (in press).

NOWAK, M. A., KRAKAUER, D. C. & DRESS, A. (1999b). An error-limit for the evolution of language. *Proc. Roy. Soc. London B* (in press).

PINKER, S. (1995). *The Language Instinct*. Harmondsworth: Penguin.

PINKER, S. & BLOOM, P. (1990). Natural language and natural selection. *Behav. Brain Sci.* **13,** 707–784.

SMITH, W. J. (1977). *The Behaviour of Communicating*. Cambridge, MA: Harvard University Press.

ZIPF, G. K. (1935). *The Psychobiology of Language*. Boston: Houghton-Mifflin.

Author Queries

1. Cavalli-Sforza & Cavalli-Sforza 1995 not listed in the Ref. list
2. Grassly et al. 2000 not listed in Ref. list
3. Mandelbrot 1958 or 1959?
4. Eigen & Schuster 1978 not listed
5. Estoup 1916, please give publishers name
6. Nowak & Krakauer 1999; Nowak et al. 1999a,b, please update