THE ROYAL SOCIETY

An error limit for the evolution of language

Martin A. Nowak[1*], **David C. Krakauer**[1] **and Andreas Dress**[2]

[1]*Institute for Advanced Study, Princeton, NJ 08540, USA*
[2]*FSPM Strukturbildungsprozesse, Universitaet Bielefeld, D-33615 Bielefeld, Germany*

On the evolutionary trajectory that led to human language there must have been a transition from a fairly limited to an essentially unlimited communication system. The structure of modern human languages reveals at least two steps that are required for such a transition: in all languages (i) a small number of phonemes are used to generate a large number of words; and (ii) a large number of words are used to a produce an unlimited number of sentences. The first (and simpler) step is the topic of the current paper. We study the evolution of communication in the presence of errors and show that this limits the number of objects (or concepts) that can be described by a simple communication system. The evolutionary optimum is achieved by using only a small number of signals to describe a few valuable concepts. Adding more signals does not increase the fitness of a language. This represents an error limit for the evolution of communication. We show that this error limit can be overcome by combining signals (phonemes) into words. The transition from an analogue to a digital system was a necessary step toward the evolution of human language.

Keywords: language; words; phonemes; mathematical biology; adaptive dynamics

1. INTRODUCTION

Language is the defining characteristic of humans. It is the trait that sets us most clearly apart from all other animals. While many animal species have evolved sophisticated communication systems, they normally consist of a limited number of context based signals (Frisch 1967; Marler 1970; Cheney & Seyfarth 1990; Hauser 1996). Human language is essentially unlimited. This transition from a limited to an unlimited representation system is what interests us here.

According to Chomsky (1975, 1980), all of the roughly 6000 human languages that exist today have the same underlying universal grammar which is the product of a special circuitry in the brain, a language organ. Because the language organ is innate, so is universal grammar. Chomsky argues that all humans but no other animals have universal grammar. Because of the lack of simple precursors of human language, Chomsky argues that the language organ must have evolved for some other purpose (perhaps for combinatorial manipulation of mental representations of objects and processes) and was later taken over for language generation and comprehension.

In Bickerton's (1990) classification, animal signals use primary representations that refer to whole situations (e.g. food or predator) while human language uses secondary representations that consist of parts with their own meaning (e.g. nouns referring to objects or verbs referring to actions). Proto-languages have secondary representations, but lack some of the most fundamental properties of language such as a mapping between word order and meaning or the use of grammatical morphemes for structuring the sentence. Proto-languages lie somewhere between animal communication and full-blown human language. Users of proto-language include signing chimpanzees, children under about two years of age and first-generation speakers of pidgin. Both Chomsky and Bickerton have difficulties in imagining how human language could have evolved gradually by natural selection.

Pinker (1995; see also Pinker & Bloom 1990) argues that language is a complex trait, and that natural selection via gradual changes is the only mechanism that can account for the emergence of such a complex trait. The fact that our closest living relatives (primates) do not have language does not argue against its evolution by natural selection, because these species are not our direct ancestors. Instead, simple language or proto-language systems evolved gradually in our direct ancestors (most likely in *Australopithecus* or early *Homo* species who lived several million years ago). It seems unlikely that language is simply the by-product of a big brain. Instead language was favoured by natural selection because of its adaptive value. In agreement with Chomsky's theory, Pinker argues that all humans are born with a language instinct which is responsible for universal grammar.

In anatomical terms, perhaps the two most important events that are required for human speech are the lowering of the larynx and the lateralization of the brain. The first is required to produce the current variety of phonemes, the second is required for the fine control of the speech organs (Miller 1981; Lieberman 1991; Deacon 1997).

Our programme here and in related papers (Nowak & Krakauer 1999; Nowak *et al.* 1999) is to outline an evolutionary scenario that would allow the fascinating features

*Author for correspondence.

Received 17 June 1999 *Accepted* 16 July 1999

of human language to evolve gradually by natural selection. We are convinced that human language originated from animal communication systems, and we are therefore required to provide an explanation for how human language might have evolved from simpler communication systems. Specifically, in this paper, we study the evolution of communication in the presence of noise: individuals may mistake one signal for another (Smith 1971). We show that this limits the amount of information which can be transfered between individuals, and thus represents an error limit for the evolution of communication. The fitness of a language cannot be increased arbitrarily by just adding more signals. However, the fitness can be increased by combining (a small number of) signals into words. Linguists call the units that make up words phonemes. Modern human languages have a limited number of phonemes: all of the 317 languages in the University of California Los Angeles Segment Inventory Database (UPSID) have between 11 and 141 phonemes, but 70% of these languages have between 20 and 37 phonemes.

2. THE BASIC MODEL

Let us start by formalizing communication in terms of an evolutionary game. Consider a group of individuals (animals or early humans) which can communicate about a given number of objects. 'Objects' is used here in a broad sense to include objects in the environment, other animals, other people, concepts, or actions. We assume that successful communication is of benefit to both speaker and listener: if the information 'object i' is correctly transmitted from speaker to listener, then both get a pay-off a_i which defines the intrinsic 'value' of object i.

A more general framework should allow different pay-offs for speaker and listener, and also the possibility that communication about some objects may only be of benefit to one of them. We will explore this generalization in a subsequent paper, and concentrate here on the special case of 'symmetric communication' where equal pay-offs are given to both speaker and listener. Following the central assumptions of evolutionary game theory, pay-off is equated to fitness (Maynard Smith 1982). Thus, an individual that communicates well leaves more offspring.

Assume that language L has n specific signals to communicate about n objects. If two individuals who speak language L meet, their pay-off is

$$F = \sum_{i=1}^{n} a_i. \tag{1}$$

This assumes that communication about all objects i occurs with equal frequency, or else that any differences in frequency are included in the a_i values. If all objects have the same a_i value, then the overall pay-off is simply given by the number of objects, $F = n$. These equations describe error-free communication.

Let us now include the possibility of misunderstanding signals. Denote by u_{ij} the probability of mistaking signal i for signal j. The corresponding error matrix U is a stochastic $n \times n$ matrix. Its rows sum to unity. The diagonal values, u_{ii} define the probabilities of correct communication. Given this error matrix, the pay-off for language L becomes

$$F = \sum_{i=1}^{n} a_i u_{ii}. \tag{2}$$

Here, the implicit and natural assumption is that only correct communication leads to a reward (even though there may always be some spectacular exceptions).

The error matrix can be defined in terms of similarity between signals. Denote by s_{ij} the similarity between signals i and j. 'Similarity' should be a number between zero and unity, with unity denoting 'identity'. Thus we have $s_{ii} = 1$. The probability of mistaking signal i for j is now given by $u_{ij} = s_{ij}/\Sigma_{k=1}^{n} s_{ik}$; hence, the probability of mistaking signal i for j is defined by how similar signal i is to signal j compared to how similar signal i is to all other signals in the language. The probability of correct communication is given by $u_{ii} = 1/\Sigma_{k=1}^{n} s_{ik}$. Thus, the fitness function in terms of similarity becomes

$$F = \sum_{i=1}^{n} \left(a_i \Big/ \sum_{j=1}^{n} s_{ij} \right). \tag{3}$$

Let us now imagine that signals (or more specifically 'sounds' if we consider a spoken language) can be embedded in some metric space X and that d_{ij} denotes the distance between sounds i and j. The similarity should then be a monotonically decreasing function of their distance, $s_{ij} = f(d_{ij})$.

3. ALL OBJECTS HAVE THE SAME VALUE

Let us first consider the situation where correct communication about each object contributes the same amount to the fitness function $a_i = 1$ for all i. Thus we have

$$F = \sum_{i=1}^{n} \left[1 \Big/ \sum_{j=1}^{n} f(d_{ij}) \right]. \tag{4}$$

The following questions arise: (i) What is the optimum distribution of n sounds (signals) that maximize the fitness of the language? (ii) How does increasing the number of sounds n affect the fitness of the language? First, we describe three specific examples; then we present a general result.

(a) *The line*

Suppose that sounds can somehow be represented by a one-dimensional (1D) spectrum, for example the interval [0, 1]. Sound i is given by the number $x_i \in [0, 1]$. The distance between sounds i and j is given by $d_{ij} = |x_i - x_j|$. The perceived similarity between two sounds is a monotonically decreasing function of their distance. As a natural choice, we consider $s_{ij} = \exp(-\alpha d_{ij})$, with $\alpha > 0$. The parameter α can be interpreted as a measure of the resolution of perception: when α is high perception is more accurate.

For a given n, we want to find the optimum configuration $x_1, \ldots, x_n$ that maximizes

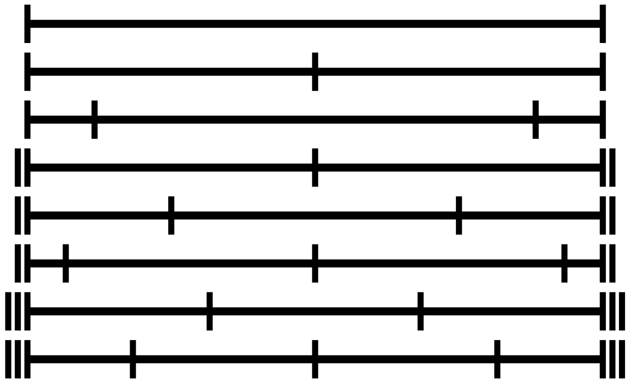

Figure 1. The optimum configuration for $n = 2$–9 signals on the line [0, 1]. The fitness function is given by equation (5). We choose $\alpha = 1$. Note that all configurations are symmetrical. For $n \geqslant 5$ we find multiplicity at the boundaries, 0 and 1.

$$F = \sum_{i=1}^{n} \left[1 \bigg/ \sum_{j=1}^{n} \exp\left(-\alpha |x_i - x_j| \right) \right]. \tag{5}$$

Clearly, for $n = 1$, any choice of $x_1 \in [0, 1]$ will do. For $n = 2$, the optimum configuration is to place the sounds at opposite ends of the spectrum: $x_1 = 0$ and $x_2 = 1$. For $n = 3$, we find $x_1 = 0$, $x_2 = 0.5$, $x_3 = 1$. For $n \geqslant 4$, the optimum configuration depends on α. Numerical simulations for $\alpha = 1$ and $n = 4$ suggest that $x_1 = 0$, $x_2 = 0.1305\ldots$, $x_3 = 0.8695\ldots$, $x_4 = 1$ is optimum. For $\alpha = 1$ and $n = 5$, the optimum configuration is, somewhat surprisingly, $x_1 = x_2 = 0$, $x_3 = 0.5$, $x_4 = x_5 = 1$. Hence, the maximum fitness is achieved by using only three different signals to communicate about five objects. For $\alpha = 1$ and $n = 8$, we find that three signals are at zero and three signals are at unity, while only two signals lie in the interior; thus the highest pay-off is achieved if four different signals are used to communicate about eight objects. Figure 1 shows the optimum configurations for $n = 2, \ldots, 9$ for $\alpha = 1$.

For arbitrary values of α, we find that for sufficiently large numbers n the optimum configuration always consists of collapsing some k sounds at the opposite ends of the spectrum (at zero and unity) while the remaining $n - 2k$ sounds are in the interior of the interval and distinct. Hence, multiplicity (i.e. identical sounds are used for distinct objects) occurs at the boundary of the interval (at $x = 0$ or 1) but not in the interior. The number k depends on α and n. All optimum configurations are symmetrical. The interior sounds are in a fairly regular arrangement; the distance between neighbouring sounds is roughly constant.

While the exact optimum configuration for specific values of α and n is difficult to calculate, we can prove that as n grows to infinity, the maximum value of F converges to the surprisingly simple expression

$$F_{\max} = 1 + \frac{\alpha}{2}. \tag{6}$$

This implies that for any given value of α the maximum fitness of the language cannot exceed a certain limit. Adding more and more signals (and objects that are described by these signals) does not increase the fitness of the language beyond this limit. Each new signal–object pair leads to a marginal fitness increase. This represents an error limit for language evolution. There will certainly be selection to increase the resolution α as much as possible, but we expect this process to reach some physical boundary.

(b) *The circle*

Another interesting case is generated by considering the 1D circle which corresponds to the not implausible situation where the spectrum of signals or sounds does not have a beginning or end. Sounds are defined by numbers in $[0, 1)$; the distance between x_i and x_j is given by $d_{ij} = \min\{|x_i - x_j|, 1 - |x_i - x_j|\}$. Again, let $s_{ij} = \exp(-\alpha d_{ij})$. We find that the optimum configurations are symmetrical (given by the regular polygons). Multiplicity cannot occur. For large n the maximum fitness converges to

$$F_{\max} = \frac{\alpha/2}{1 - \exp(-\alpha/2)}. \tag{7}$$

As before, this equation represents an error limit. Increasing the repertoire of the language does not allow the total fitness to exceed this limit.

(c) *Constant similarity*

The simplest possibility is to assume that there is a constant similarity s between any two distinct signals. Thus, we have $s_{ii} = 1$ and $s_{ij} = s$ where $s < 1$. (The geometric interpretation is that the signals are represented by the vertices of an n-dimensional simplex.) We obtain the fitness function

$$F = \frac{n}{1 + (n-1)s}. \tag{8}$$

This function is monotonically increasing and converges for large n to $F_{\max} = 1/s$. As before, increasing the size of the repertoire n does not allow the language to exceed this limit.

(d) *A general result*

For the general situation where the sounds (or signals) are embedded in some arbitrary bounded subset of R^k or, more generally, some arbitrary pre-compact metric space X, and where $f(d_{ij})$ is a declining function which is positive on some interval $[0, \epsilon)$, we can show that the fitness

$$F = \sum_{i=1}^{n} \left[1 \bigg/ \sum_{j=1}^{n} f(d_{ij}) \right], \tag{9}$$

is still limited by some constant c, which depends exclusively on X and f, but not on n. In other words, as the repertoire n increases, F cannot exceed a certain value. This is the most general formulation of the error limit (A. Dress and M. A. Nowak, unpublished results).

4. DIFFERENT VALUES

Let us now turn to the situation where objects can have different values a_i. Without loss of generality we label the objects according to their value: $a_1 \geqslant a_2 \geqslant \ldots$. Suppose a given language L describes n objects. Because we are interested in the maximum fitness, we always assume that L describes the n most valuable objects. The fitness function is given by

$$F = \sum_{i=1}^{n}\left(a_i \Big/ \sum_{j=1}^{n} s_{ij}\right). \tag{10}$$

It is clear that F need not be monotonically increasing with n.

We shall illustrate this point using the specific example of constant similarity. In this case, the fitness of a language which describes the n most valuable objects becomes

$$F(n) = \frac{1}{1+(n-1)s}\sum_{i=1}^{n} a_i. \tag{11}$$

Let us define the two limiting values $a = \lim_{i\to\infty} a_i$ and $A = \sum_{i=1}^{\infty}(a_i - a)$. The decisive inequality is

$$s > a/(a+A). \tag{12}$$

If this condition is fulfilled, then there exists a number n_0 such that $F(n) > F(n+1)$ holds for all $n > n_0$. In other words, the fitness $F(n)$ will obtain a maximum value for some intermediate value of n. Conversely, if $s \leqslant a/(a+A)$, then $F(n) \leqslant F(n+1)$ for all n, with $F(n)$ becoming constant for large n if and only if $s = a/(a+A)$ and the a_n become constant for large n. In both cases, however, the fitness is bounded and certainly cannot exceed the value a_1/s.

If the sequence a_i converges to zero, i.e. $a = 0$, or if A is infinite, then every positive s satisfies condition (12), and we always have the situation that maximum fitness is obtained by restricting one's language to describing an intermediate number of objects. However, if a is greater than zero and if A is finite, then there exists a critical value of s (given by $s_c = a/(a+A)$) such that if s is smaller than s_c, then no intermediate value of n will achieve maximum fitness, which is given by a/s.

5. EVOLUTION

So far we have calculated the pay-off that is obtained if the two communicating individuals use the same language L. The framework can be extended to describe the situation where the two individuals use different languages, L_1 and L_2. Suppose L_1 is given by $\{x_1, \ldots, x_n\}$ and L_2 by $\{y_1, \ldots, y_n\}$. Thus language L_1 uses signal x_i for object i, while language L_2 uses signal y_i for object i. Denote by $s(x_i, y_j)$ the similarity between x_i and y_j. The pay-off for L_1 versus L_2 is

$$F(L_1, L_2) = \frac{1}{2}\sum_{i=1}^{n} a_i s(x_i, y_i)\left[\frac{1}{\sum_{j=1}^{n} s(x_i, y_j)} + \frac{1}{\sum_{j=1}^{n} s(x_j, y_i)}\right]. \tag{13}$$

The two denominators appear because if L_1 speaks and L_2 listens then L_2 has to distinguish signal x_i from all y_j, whereas if L_2 speaks and L_1 listens then L_1 has to distinguish signal y_i from all x_j. This assumes that errors occur while receiving messages and that each individual expects to hear its own sounds and interprets all perceived sounds within the context of its own repertoire.

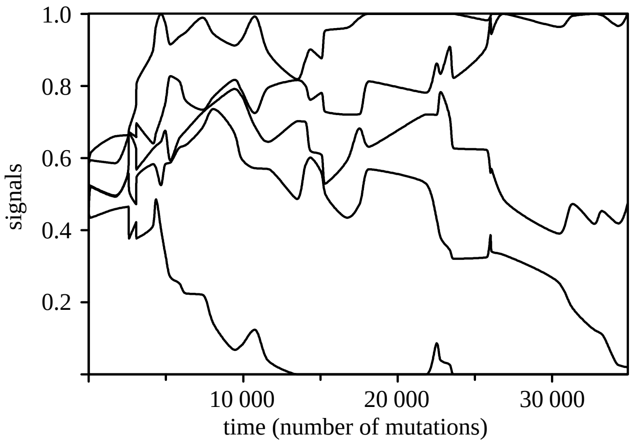

Figure 2. Stochastic adaptive dynamics of language evolution can lead to maximum fitness. There are $n = 5$ signals represented by numbers $x_1, \ldots, x_5$ from the interval [0, 1]. As similarity function we chose $s_{ij} = \exp(-\alpha|x_i - x_j|)$ with $\alpha = 1$. Initially all signals are identical, $x_1 = \ldots = x_5 = 0.5$. At each time-step a new mutant, L', is generated and tries to invade the existing language L. We always find that $F(L', L) < F(L, L)$. Thus a mutant can never invade in the setting of deterministic adaptive dynamics. There is, however, the possibility that $F(L', L') > F(L, L')$. In this case the new mutant would invade if it overcame the invasion barrier, defined by the unstable equilibrium between L and L': $y = [F(L', L') - F(L, L')]/[F(L, L) + 2F(L, L') + F(L', L')]$. Here y denotes the frequency of L' at the unstable equilibrium. We assume that the probability that L' invades and replaces L is given by $\exp(-\beta y)$. For the simulation we chose $\beta = 10$. The figure shows how the language converges after 40 000 invasion attempts (of which 44 were successful) towards the optimum configuration, $x_1 = x_2 = 0$, $x_3 = 0.5$, $x_4 = x_5 = 1$. The pay-off function is given by equation (13).

As a specific example let us assume that all sounds are taken from a 1D spectrum, the interval [0, 1], that $s(x_i, y_j) = \exp(-\alpha|x_i - y_j|)$ and that $a_i = 1$ for all i. A surprising numerical observation is that every randomly chosen language L seems to be a strict Nash solution, i.e. to fulfill $F(L, L) > F(L', L)$ for all $L' \neq L$. Thus, it is difficult to improve a language which has been adopted by everyone. Standard adaptive dynamics (Nowak & Sigmund 1990; Metz *et al.* 1996; Hofbauer & Sigmund 1998) does not work in this context. Stochastic adaptive dynamics (Wahl & Nowak 1999), however, provides an evolutionary path that can lead to the optimum (figure 2).

6. WORD FORMATION

In this section, we show how the error limit can be overcome by combining sounds into words. We will provide a very simple and intuitive argument.

Words are strings of sounds. Linguists call these sounds 'phonemes'. Suppose there are n phonemes. Let us at first only consider words of length two phonemes. There are n^2 such words. We assume that the similarity between two words is the product of the similarities between the phonemes in corresponding positions. Thus if word W_{ij} consists of phonemes i and j, then the similarity between the words W_{ij} and W_{kl} is

$$S(W_{ij}, W_{kl}) = s_{ik}s_{jl}. \tag{14}$$

Hence the fitness of a language that contains n^2 words to describe the same number of objects is

$$F = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(1 \Big/ \sum_{k=1}^{n}\sum_{l=1}^{n} s_{ik}s_{jl}\right). \tag{15}$$

This can be rewritten as

$$F = \left[\sum_{i=1}^{n}\left(1 \Big/ \sum_{j=1}^{n} s_{ij}\right)\right]^2. \tag{16}$$

Similarly, for word length L we obtain

$$F = \left[\sum_{i=1}^{n}\left(1 \Big/ \sum_{j=1}^{n} s_{ij}\right)\right]^L. \tag{17}$$

Hence, if $F_{\max}(L)$ denotes the maximum fitness that can be achieved for a given word length L, we have

$$F_{\max}(L) = F_{\max}(1)^L. \tag{18}$$

This equation describes the maximum fitness for a language that contains words of constant length L. We can also calculate the maximum fitness of a language containing words up to length L. Under the assumption that words of different lengths cannot be mistaken for each other, we get

$$F_{\max}(L) = \sum_{k=1}^{L} F_{\max}(1)^k = \frac{F_{\max}(1)^{k+1} - F_{\max}(1)}{F_{\max}(1) - 1}. \tag{19}$$

Hence combining sounds into words is an efficient way to overcome the constraint of the error limit. The total pay-off can grow exponentially with the length of words. Of course, we should not expect that natural selection will lead to ever increasing word lengths. Ultimately word length is limited by the number of objects or concepts that need to be described—see § 6(b). In addition there will be a reward for rapid communication which will reduce the average word length and, in particular, make frequently used words as short as possible.

(a) *A specific example: word formation with constant similarity between phonemes*

If any two distinct phonemes have the same similarity s (see § 3(c)), we obtain for a fixed word length l the fitness function

$$F = n^l/[1 + (n-1)s]^l = \left(s + \frac{1+s}{n}\right)^{-l} \approx (1/s)^l. \tag{20}$$

This formula remains valid even if we—realistically—assume that not all possible words are formed and that selection favours the formation of word sets with large Hamming distance. (The Hamming distance between two words is defined as the number of different phonemes in corresponding positions.) In this case, one could argue heuristically that if N words with average Hamming distance or 'effective word length' k are formed, the fitness that can be obtained is approximately (as in § 3(c)) given by

$$F = N/[1 + (N-1)s^k] \approx (1/s)^k. \tag{21}$$

For details taking into account the insights of coding theory (Shannon & Weaver 1949; Hamming 1980; Welsh 1988), see Plotkin & Nowak (2000).

As the number of words increases, the fitness approaches the maximum value $(1/s)^k$. Therefore, the maximum fitness scales exponentially with the effective word length k.

(b) *Words describing objects with different values*

In § 4, we have seen that if objects have different values a_i the maximum fitness may be achieved by describing only a finite (possibly quite small) number of objects. Here we study how word formation can increase the optimum size of the repertoire. Consider the fitness function

$$F = \frac{1}{1 + (N+1)s^k}\sum_{i=1}^{N} a_i. \tag{22}$$

As before, N is the number of words (i.e. the size of the lexicon) and k is the effective word length.

Consider the specific example $a_i = \max\{0, 1 - 2\delta i\}$, that is, the values of objects form a linearly decreasing sequence. We obtain

$$F = \frac{n[1 - \delta(n+1)]}{1 + (n-1)s^k}. \tag{23}$$

For $n \gg 1$ this is

$$F \approx \frac{n[1 - \delta n]}{1 + ns^k}. \tag{24}$$

F is a one-humped function, which obtains its maximum at

$$N_{\max} \approx \frac{1}{s}\{\surd[1 + (s^k/\delta)] - 1\}. \tag{25}$$

If s^k is small (specifically $s^k \ll \delta$), then we obtain from equation (24) that $N_{\max} = 1/(2\delta)$, that is, all objects with a positive value will be described. It essentially means that the repertoire size is not limited by the language, but by the availability of adaptively salient objects in the world. If $s^k \gg \delta$ we have $N_{\max} \approx 1/\surd(s^k\delta)$. Here the optimum lexicon size scales as $(1/s)^{k/2}$; hence, the lexicon size increases exponentially with the effective word length (until s^k becomes smaller than δ, in which case the evolutionary optimum is to describe all objects with a positive value).

7. CONCLUSION

We analysed the evolution of a simple symbolic communication system in the presence of noise and obtained the following results.

A general feature of models for language evolution is an error limit: the fitness of a language (that is the total amount of information that can be transferred) cannot be increased beyond a certain threshold by simply adding more signals. We show this result for a number of specific cases and present a general formulation.

The error limit, however, can be extended by combining signals (that is phonemes in the context of spoken language) into words. The maximum fitness scales exponentially with the length of words. Thus word formation (or 'sequencing'; Miller 1981) appears to be an essential step for the evolution of human language. It is the transition from an analogue to a digital communication system.

If communication about all objects is equally valuable, then the fitness of a language increases monotonically with the number of objects being described (converging to the value given by the error limit). If objects have different values, then describing only a small number of objects can yield maximum fitness.

The evolutionary language game studied in this paper has the unusual characteristic that appearantly all randomly chosen languages when employed by the whole population are evolutionarily stable strategies or Nash equilibria (Maynard Smith 1982; Nash 1996). This means that language evolution, in the context of this model, can only occur if new variations are adopted by (small) clusters of individuals. If there is a tendency to adopt new variations then 'mutation' in the language game may differ in an important way from mutation in genetic systems.

There is an interesting parallel between our error limit and the observation that word formation can extend this limit, and Shannon's noisy coding theorem. Shannon states that the maximum error in comunication declines exponentially as the word length increases linearly (see Shannon & Weaver 1949). In a forthcoming paper we will show the formal similarity between these two results.

Finally, we note that word formation in most human languages consists of two different processes: there is the (somewhat arbitrary) concatenation of phonemes that gives rise to multiple-phoneme words such as banana or *Australopithecus*, and there is the rule-based variation of word-stems (morphology) to convey different meanings such as *walk* and *walked*. The first of these two processes is what is important for overcoming the error limit described in this paper. The second and more difficult process belongs to theories for the evolution of grammar.

Support from The Florence Gould Foundation, J. Seward Johnson, Sr Charitable Trusts, The Ambrose Monell Foundation and The Alfred P. Sloan Foundation is gratefully acknowledged.

REFERENCES

Bickerton, D. 1990 *Species and language*. Chicago University Press.

Cheney, D. L. & Seyfarth, R. M. 1990 *How monkeys see the world: inside the mind of another species*. Chicago University 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*. London: Allen Lane, Penguin.

Frisch, K. 1967 *The dance language and orientation of bees*. Cambridge, MA: Harvard University Press.

Hamming, R. W. 1980 *Coding and information theory*. New Jersey: Prentice-Hall.

Hauser, M. 1996 *The evolution of communication*. Cambridge, MA: MIT Press.

Hofbauer, J. & Sigmund, K. 1998 *Evolutionary games and population dynamics*. Cambridge University Press.

Lieberman, P. 1991 *Uniquely human*. Cambridge, MA: Harvard University Press.

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 University Press.

Metz, J. A. J., Geritz, S. A. H., Meszena, F. G., Jacobs, F. J. A. & Van Heerwaarden, J. S. 1996 Adaptive dynamics: a geometrical study of the consequences of nearly faithful reproduction. In *Stochastic and spatial structures of dynamical systems* (ed. S. J. Van Strien & S. M. Verduyn Lunel), pp. 183–231. Amsterdam: North Holland.

Miller, G. A. 1981 *Language and speech*. San Francisco: Freeman.

Nash, J. 1996 *Essays on game theory*. Cheltenham, UK: Elgar.

Nowak, M. A. & Krakauer, D. C. 1999 The evolution of language. *Proc. Natl Acad. Sci. USA* **96**, 8028–8033.

Nowak, M. A. & Sigmund, K. 1990 The evolution of reactive strategies in iterated games. *Acta Applicandae Math.* **20**, 247–265.

Nowak, M. A., Plotkin, J. & Krakauer, D. C. 1999 The evolutionary language game. *J. Theor. Biol.* (In the press.)

Pinker, S. 1995 *The language instinct*. New York: Penguin.

Pinker, S. & Bloom, P. 1990 Natural language and natural selection. *Behav. Brain Sci.* **13**, 707–784.

Plotkin, J. & Nowak, M. A. 2000 Language evolution and information theory. (In preparation.)

Shannon, C. E. & Weaver, W. 1949 *The mathematical theory of communication*. University of Illinois Press.

Smith, W. J. 1977 *The behaviour of communicating*. Cambridge, MA: Harvard University Press.

Wahl, L. M. & Nowak, M. A. 1999 The continuous prisoner's dilemma. *J. Theor. Biol.* (In the press.)

Welsh, D. 1988 *Codes and cryptography*. Oxford University Press.