Towards an evolutionary theory of language

Martin A. Nowak and Natalia L. Komarova

Language is a biological trait that radically changed the performance of one species and the appearance of the planet. Understanding how human language came about is one of the most interesting tasks for evolutionary biology. Here we discuss how natural selection can guide the emergence of some basic features of human language, including arbitrary signs, words, syntactic communication and grammar. We show how natural selection can lead to the duality of patterning of human language: sequences of phonemes form words; sequences of words form sentences. Finally, we present a framework for the population dynamics of grammar acquisition, which allows us to study the cultural evolution of grammar and the biological evolution of universal grammar.

Martin A. Nowak*
Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA.
*e-mail: nowak@ias.edu

Natalia L. Komarova
Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA.
and Dept of Applied Mathematics, University of Leeds, Leeds, UK LS2 PJT.

During evolution, a nervous system emerged that enabled animals to observe their world, learn behavioral patterns and communicate with one another. One lineage of the animal kingdom eventually produced a communication system with *infinite* expressibility. Human language is infinite, not because everything can be expressed, but because there are infinitely many sentences; no finite list can contain all possible sentences of a given language. Language allows the transfer of unlimited, non-genetic information among individuals and thus induces a new mode of evolution. Language gives rise to cultural evolution far beyond what is possible for non-speaking animals. Among all great evolutionary innovations that affected evolution itself, such as nucleic acids, cells, chromosomes, multi-cellular organisms, the nervous system etc, language is the only one (presently) confined to one species.

Humans and chimps diverged some 5 million years ago. Because humans have complex language, but chimps do not, the final components of the biological basis of human language must have arisen since then. It is clear, however, that evolution did not build the human language faculty *de novo* in the last few million years, but used material that had evolved in other animals over a much longer time. Many animal species have sophisticated cognitive abilities in terms of understanding the world and interacting with one another[1]. Evolution often uses existing structures for new and sometimes surprising purposes. Monkeys, for examples, appear to have brain areas similar to our language centers, but seem to use them for controlling facial muscles and for analyzing auditory input[2]. Evolution may have had an easy task here to reconnect these centers for human language. Hence the human language instinct should not be seen as the result of a sudden moment of inspiration of evolution's blind watchmaker, but rather the consequence of several hundred million years of 'experimenting' with animal cognition.

Language allowed our ancestors to share ideas and experiences, and to solve problems in parallel. The adaptive significance of human language is obvious. It pays to talk. Cooperation in hunting, making plans, coordinating activities, task sharing, social bonding, manipulation and deception all benefit from an increase in expressive power. Natural selection (we use it to include sexual selection) can certainly see the consequences of communication[3,4].

The linguist Ray Jackendoff outlines how human language reveals an architecture that seems to have been formed by distinct innovations that were added over time[5,6]. He also finds 'fossils' of earlier, more primitive communication systems in the grammar of modern language. Part of Jackendoff's program is an extension of Bickerton's idea that modern language evolved from 'protolanguage', which still can be found in our brain[7]. Protolanguage emerges whenever full-blown language is disrupted such as in pidgin languages or in children who were deprived of social interaction (the most famous case is Genie). These are some examples of biological and linguistic evidence which point towards a gradual evolutionary process that has shaped human language.

'Evolutionary dynamics are based on well-defined mathematical principles: mutation and selection…'

Evolution is based on well-defined, mathematical principles: mutation and selection. Hence in order to talk about language evolution, it seems essential to construct a precise mathematical framework. This is what we do in this article. We will discuss how a group of individuals (humans or other animals) can evolve a communication system where arbitrary signals become associated with specific referents. We will show that mistakes in communication lead to an error limit, which can be overcome by sequencing basic signal units (such as phonemes) into words. We discuss a necessary condition under which natural selection can see the advantage of syntactic communication. Finally, we present a general framework for the evolution of grammar acquisition and discuss how natural selection acts on universal grammar. The material presented here is part of a larger effort to establish a connection between evolutionary biology, linguistics, and cognitive science[8–21].

Arbitrary signs

First we ask how natural selection can design a simple communication system where certain, arbitrary signals become associated with specific referents[22–26].
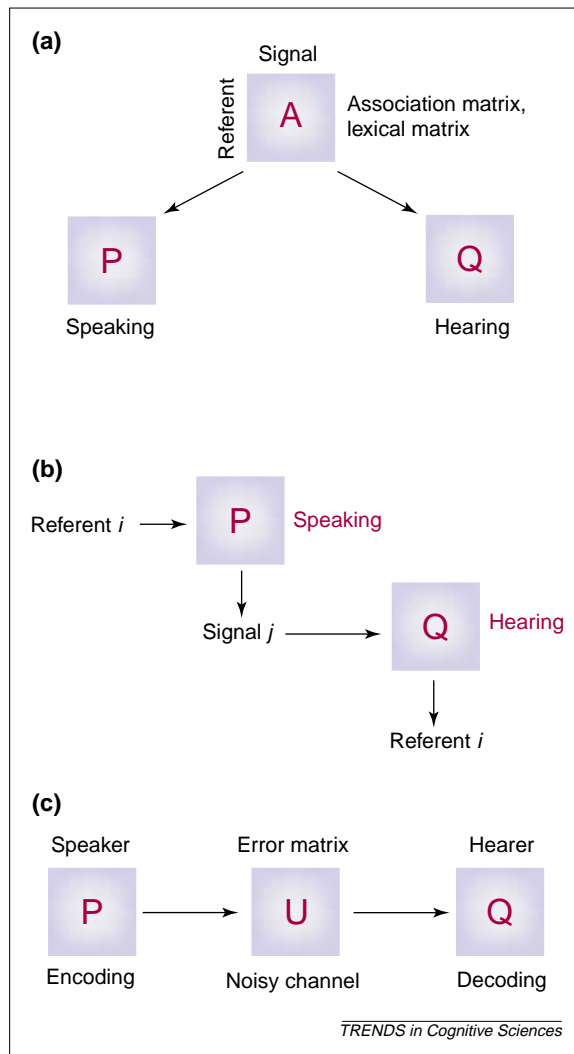
Fig. 1. (a) The association matrix, *A*, links signals to referents. The lexical matrix of human language links word form to word meaning. The elements of this matrix, a_{ij}, are non-negative real numbers and denote the strength of the association between referent *i* and signal *j*. A speaker is described by a *P* matrix. The element p_{ij} denotes the probability of using signal *j* for referent *i*. A hearer is described by a *Q* matrix. The element q_{ij} denotes the probability of interpreting signal *j* as denoting referent *i*. The *P* and *Q* matrices are derived from the *A* matrix by normalizing rows and columns respectively. (b) Suppose a speaker uses signal *j* for referent *i*. Correct communication occurs if the hearer receives signal *j* and associates it with referent *i*. Let us now consider two individuals *I* and *J* with association matrices A_I and A_J. We can define the payoff for *I* communicating with *J* as:

$$F(A_I,A_J) = (1/2)\sum_{i=1}^{n} \sum_{j=1}^{m} \left(p_{ij}^{(I)} q_{ji}^{(J)} + p_{ij}^{(J)} q_{ji}^{(I)} \right)$$

The term $\sum_{j=1}^{m} p_{ij}^{(I)} q_{ji}^{(J)}$ denotes the probability that individual *I* will successfully communicate referent *i* to individual *J*. This probability is summed over all objects and averaged over the situation where individual *I* signals to individual *J* and vice versa. Note that the payoff function assumes that communication about each object occurs with the same frequency. (c) We can also assume that signals can be mistaken for each other. There is an error matrix, *U*, between speaker and hearer. This model is based on Shannon's information theory. The *P*, *U*, and *Q* matrices describe, respectively, encoding, a noisy channel and decoding. In our evolutionary context, we observe that errors during communication often lead to a scenario where maximum fitness is achieved for systems with limited repertoire size. Sequencing of phonemes into words (that is increasing the code length) can extend this error limit[32].

Consider an association matrix, *A*, whose entries, a_{ij}, specify the strength of the association between referent *i* and signal *j* (see Fig. 1). The association matrix determines the probability that a speaker will use signal *j* when wanting to communicate referent *i*, and the probability that a hearer will interpret signal *j* as denoting referent *i*. Hence from the association matrix we can calculate the probability of correct information transfer between a speaker and a hearer. Such an association matrix is at the basis of any animal communication system. It is also a convenient description for the lexical matrix of human language. The lexical matrix specifies the arbitrary relations between word form and word meaning[27].

The arbitrariness of the association between signals and referents gives rise to the problem of coherence: if different individuals can assign different signals to the same referent (or vice versa), then how does the population achieve a coherent communication system where everybody uses the same association between signals and referents (or word forms and word meanings)?

Let us consider evolutionary dynamics. There is a group of individuals. In the beginning, each individual has a different, randomly chosen *A* matrix. Thus no signal is associated with a specific referent. For any given referent, there is only a small probability that a speaker-hearer pair will have a successful communication about it. Furthermore, we assume that individuals reproduce and generate offspring that inherit – genetically – a mechanism for learning the association matrix of their parents or others. After specifying some learning mechanism, we can simulate this evolutionary process. We will, however, observe that no coherent communication will evolve. In this case, there is no selection against individuals who do not learn any associations at all. The mechanism for learning the *A* matrix will eventually deteriorate.

For natural selection to act on language ability, there must be a reward for successful communication. We have to link language to biological 'fitness'. Let us therefore assume that successful communication leads to a 'payoff' for both the speaker and the hearer. In the spirit of evolutionary game theory, we link payoff to reproductive success[28,29]. Individuals that communicate more successfully have increased survival probabilities and leave more offspring.

Let us first assume that offspring learn their association matrix from some randomly chosen individuals or some population average of the *A* matrix regardless of payoff. Again no coherent language will evolve. The reason is that more successful *A* matrices do not proliferate faster than less successful ones.

If however we assume that offspring learn the *A* matrix of their parents or of other individuals proportional to their payoff, then a coherent communication system can emerge. In both cases, successful *A* matrices spread. Learning from the parents works, because more offspring are born to parents with successful *A* matrices. Learning

preferentially from other individuals with higher payoff gives a direct advantage to successful *A* matrices.

For either case, we can observe that after a few generations different individuals have similar *A* matrices and certain associations between specific signals and referents become stronger whereas other – conflicting – associations disappear. The evolutionary optimum is a state where each signal is uniquely associated with one referent and vice versa. The particular signal-referent pairing is arbitrary. For *n* signals and referents there are *n*! evolutionarily stable *A* matrices[30]. The population will converge to one arbitrary *A* matrix out of these *n*! possibilities. Hence the model describes the emergence of arbitrary signs.

The next task is to calculate the minimum cognitive requirements for the language learning device that leads to the evolution of a coherent association matrix. This task turns out to be difficult. We have partial results for specific cases, but as yet no general model that is analytically tractable.

A simplified model that provides analytical insights makes the following assumptions[31]: (1) the population size is large, the evolutionary dynamics are deterministic; (2) the *A* matrix has only binary entries: if $a_{ij} = 1$, there is an association between referent *i* and signal *j*; if $a_{ij} = 0$, there is none; (3) offspring learn the *A* matrix of one parent. If the parent's *A* matrix has a 1 entry in a particular place, the offspring's *A* matrix has a 1 entry in the same place with a probability $1 - u$. If the parent's *A* matrix has a 0 entry in a particular place, the offspring's *A* matrix will always have a 0 entry in this place. Therefore, offspring do not form new associations. For this model, unique signal-referent pairings are the only stable equilibrium solutions of the evolutionary dynamics. The maximum number of signal-referent pairs that can evolve and be maintained is given by $n = 1/(2u)$. On average an individual knows $n/\sqrt{e}$ signals, and two randomly chosen individuals have n/e signals in common. (Here, *e* is Euler's number.)

We do not have analytic results for the case where offspring can erroneously form associations that are not present in their parent's *A* matrix. Computer simulations for finite populations and stochastic dynamics suggest the following results[26,31]. If this type of mistake is too frequent then no coherent communication can evolve. If the error rate is below a threshold then coherent associations can emerge. The observed associations are close to the evolutionary optimum, but some signals may refer to more than one referent (homonymy) and some referents may be associated with more than one signal (synonymy). Associations are metastable. From time to time there are transitions among predominating *A* matrices. In the context of historical linguistics, this corresponds to spontaneous changes in lexicon.

Word formation

Association matrices are useful descriptions of both animal communication systems and the lexical matrix of human language. In the first case, they describe the association between animal signals and their meaning, in the second case they describe the association between word form and word meaning. There are, however, fundamental differences between animal signals and word forms. Animal communication appears to be based on fairly limited repertoires (perhaps 10–100 signals), whereas human languages use large numbers of words (of the order of 10 000 or more). Furthermore, human language makes extensive use of combinatorics: words are sequences of well-defined smaller building blocks, called phonemes. In this section, we formulate an argument for why it is necessary that words are made up of phonemes.

Suppose we have a communication system where certain signals are unambiguously associated with certain referents. Clearly the communicative potential and therefore the biological fitness of the system increases with the number of signals. However, as the number of signals increases, chances are that some of them will sound quite similar to others. If we admit the possibility that signals can be mistaken for each other, there is a limit to the increase in fitness. A general mathematical result shows that for any such signaling system there is a maximum fitness which cannot be overcome by adding more signals[32].

If in addition we assume that different referents have different fitness contributions, then usually there exists an intermediate number of signals which maximizes fitness. Adding further signals reduces fitness. In this case, natural selection favors limited repertoires where a small number of signals denote the most valuable referents[25].

This error limit can be extended if combinatorial sequences of signals are used. It can be shown that the maximum fitness of a communication system increases exponentially with the length of the sequence[33]. This observation is related to Shannon's 'noisy coding theorem'[34]. If natural selection acts on the rate of communication, then there is an optimum word length that maximizes fitness.

Human language makes use of this principle. Words are sequences of individual phonemes. Each human language uses only a small fraction of all possible phonemes, but sequences of phonemes give rise to large numbers of words. The same argument holds for sign language: sequencing of basic units increases the error limit.

A first step towards syntactic communication

Human language uses combinatorics on two levels: sequences of phonemes form words, sequences of words form sentences. The linguist Charles Hockett called this design 'duality of patterning'. The sequencing of words into sentences is a necessary component of syntactic communication. Let us define compound signals as those that consist of parts that have their own meaning. In contrast, elementary signals cannot be decomposed into parts that have their own meaning. The alarm calls of vervet monkeys for leopard, snake or eagle are examples of elementary

Fig. 2. A very simple model for exploring the emergence of syntactic communication. (a) Non-syntactic communication uses elementary signals refering, for example, to the events 'lion sleeping', 'monkey running'. Syntactic communication uses compound signals refering, for example, to the objects 'lion', 'monkey' and the actions 'running' and 'sleeping'. (b) An evolutionary model[36] shows that syntactic communication is only favoured by natural selection, firstly, if the number of relevant messages exceeds a certain threshold and, secondly, if elementary signals can be used in sufficiently many different messages.
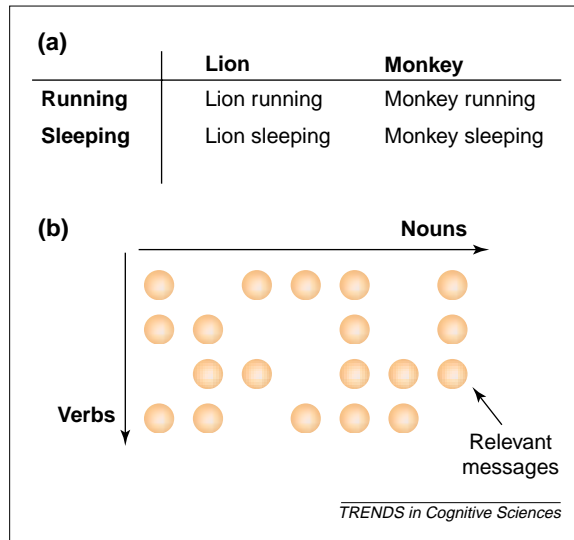
signals[35]. Word stems (listemes) of human languages are elementary signals, but phrases, sentences or any syntactic structures of human languages represent compound signals. The question which we would like to answer is how natural selection can guide the emergence of such syntactic structures (see Fig. 2).

Clearly, communication systems with compound signals have greater potential. The number of possible messages can greatly exceed the number of components (words) that make up these messages. For elementary communication, each message has to be learned, whereas compounding allows to express new messages that have not been encountered before. In human language, words have to be memorized, but most sentences are new constructions. Given these advantages it seems surprising that animals make little (or no?) use of compound signals.

An evolutionary model for the transition from elementary to compound communication shows that certain conditions have to be met before natural selection can see the advantages of compounding[36]. First, the total number of relevant messages has to exceed a critical value. Hence, only if the communication system has reached a certain size, can there be an advantage to using compound signals. Smaller systems are more efficiently encoded by elementary signals. Second, the compound signals must be able to encode the relevant messages in such a way that individual components occur in many different messages. If each component would only appear in one or a few messages then there is little chance that this system might out-compete non-syntactic communication.

Apart from combining elementary signals into compound messages, syntactic communication requires rules that specify how the parts of signals relate to each other to convey a certain meaning. This brings us to our next topic, the evolution of grammar.

Universal and other grammars

The most fascinating aspect of human language is grammar. Grammar is a computational system that mediates a mapping between linguistic form and meaning. Grammar is the machinery that gives rise to the unlimited expressibility of human language.

Children develop grammatical competence spontaneously without formal training. All they need is interaction with people and exposure to normal language use. The child hears a certain number of grammatical sentences and then constructs an internal representation of the rules that generate grammatical sentences. Chomsky pointed out that the evidence available to the child does not uniquely determine the underlying grammatical rules[37]. This phenomenon is called the 'poverty of stimulus'[38]. The 'paradox of language acquisition'[39] is that children nevertheless reliably achieve correct grammatical competence. How is this possible?

As Chomsky pointed out: 'To learn a language, then, the child must have a method for devising an appropriate grammar, given primary linguistic data. As a precondition for language learning, he must possess, first, a linguistic theory that specifies the form of grammar of a possible human language, and second, a strategy for selecting a grammar of the appropriate form that is compatible with the primary linguistic data.' (Ref. 37). Chomsky introduced the term Universal Grammar (UG) to denote the preformed 'linguistic theory', the initial pre-specification of the form of possible human grammars[40].

Hence, for language acquisition the child needs a mechanism for processing the input sentences and a 'search space' of candidate grammars from which to choose the appropriate grammar. Chomsky's original concept is that UG is a rule system that generates the search space. More recent views use UG to encompass both the search space and the mechanism for evaluating input sentences. Therefore, UG has become almost synonymous with 'mechanism of language acquisition'.

The notion of an innate, genetically encoded, UG is controversial[41–43]. Much of the discourse, however, focuses on which specific linguistic features are innate (for example, phrase structure rules of X-bar theory, or lexical categories such as nouns and verbs) and to what extent UG is a specific syntactic module or simply uses general purpose cognitive abilities. We do not participate in this controversy. Instead we choose a sufficiently general formulation of the process of language acquisition. Ultimately everybody agrees that human beings require some innate components for language acquisition. These innate components are what we call UG.

First of all, let us state that 'poverty of stimulus' has an elegant mathematical formulation known as Gold's theorem[44]. Suppose there is a rule that generates a subset of all integers. A person is provided with a sample of integers that are generated by the rule. After some time the person is asked to produce other integers that are compatible with the rule. Gold's theorem states that this task cannot be solved. Any finite number of sample integers is not enough to

(a)

Grammar

Phonological rules ⟷ Syntactic rules ⟷ Conceptual rules

Hearing and speaking   Perception and action

(b)

Set of all sentences | Grammatical | Ungrammatical

Set of all sentences:
0
1
00
01
10
11
000
001
...

G

Grammatical:
1
01
10
001
...

Ungrammatical:
0
00
11
000
...

(c)

Linguistic form

Meaning

1   1
1
  1   1
1
   1
     1

Grammar

Word form

Word meaning

1   1
1
  1
1   1
1
   1
     1

Lexical matrix

TRENDS in Cognitive Sciences

Fig. 3. (a) The grammar of human language is a rule system that encompasses phonological, syntactic and conceptual (semantic) rules. The phonological rules are linked to hearing and speaking, whereas the conceptual rules are linked to perception and action. The linguist Ray Jackendoff describes phonology, syntax and semantics as three independent combinatorial systems that are linked via interfaces. (b) Mathematically, a grammar can be seen by a rule system that divides a countable infinite number of sentences into two subsets, grammatical and ungrammatical. (c) More generally, a grammar should be seen as a rule system that generates a mapping between linguistic form and meaning. Note the formal similarity between such a 'grammar matrix' and the lexical matrix, which links word forms to word meanings. The important differences are size and compressability: the lexical matrix consists of a finite number of memorized items, whereas the grammar matrix has infinitely many entries that can be compressed into rules. Clearly the grammar matrix can also be interpreted as including the lexical matrix. Such a representation holds the starting point for a possible unified theory that describes both the acquisition of lexical items and grammatical rules.

determine uniquely the underlying rule. The person can only solve the task if she had a preformed expectation determining which rules are possible (or likely) and which are not. The sample integers correspond to the sentences presented to the child, the rule corresponds to the grammar used by the parents (or other speakers). The preformed expectation is universal grammar. Hence, in this

sense 'poverty of stimulus' and the necessity of an innate universal grammar are not controversial issues, but mathematical facts.

Let us now formulate a mathematical description of language acquisition[45–50]. The sentences of all languages can be enumerated. We can say that a grammar, G, is a rule system that specifies which sentences are allowed and which sentences are not allowed (see Fig. 3). Universal grammar, in turn, contains a rule system that generates a set (or a search space) of grammars, $\{G_1, G_2,…, G_n\}$. These grammars can be constructed by the language learner as potential candidates for the grammar that needs to be learned. The learner cannot end up with a grammar that is not part of this search space. In this sense, UG contains the possibility to learn all human languages (and many more). Figure 4 illustrates this process of language acquisition. The learner has a mechanism to evaluate input sentences and to choose one of the candidate grammars that are contained in his search space.

More generally, it is also possible to imagine that UG generates infinitely many candidate grammars, $\{G_1, G_2,…\}$. In this case, the learning task can be solved if UG also contains a prior probability distribution on the set of all grammars. This prior distribution biases the learner towards grammars that are expected to be more likely than others. A special case of a prior distribution is one where a finite number of grammars is expected with equal probability and all other grammars are expected with zero probability, which is equivalent to a finite search space.

A fundamental question of linguistics and cognitive science is what are the restrictions that are imposed by UG on human language. In other words, how much is innate and how much is learned in human language. In learning theory[51,52], this question is studied in the context of an ideal speaker–hearer pair. The speaker uses a certain 'target grammar'. The hearer has to learn this grammar. The question is, what is the maximum size of the search space such that a specific learning mechanism will converge (after a number of input sentences, with a certain probability) to the target grammar.

In terms of language evolution, the crucial question is what makes a *population* of speakers converge to a coherent grammatical system. In other words, what are the conditions that UG has to fulfill for a population of individuals to evolve coherent communication? In the following, we will discuss how to address this question[53,54].

Population dynamics of grammar acquisition

Imagine a group of individuals that all have the same UG, given by a finite search space of candidate grammars, $G_1,…,G_n$, and a learning mechanism for evaluating input sentences. Let us specify the similarity between grammars by introducing the numbers s_{ij} which denote the probability that a speaker who uses G_i will say a sentence that is compatible with G_j.

Fig. 4. Universal grammar specifies the search space of candidate grammars and the learning procedure for evaluating input sentences. The basic idea is that the child has an innate expectation of grammar (for example a finite number of candidate grammars) and then chooses a particular candidate grammar that is compatible with the input.

Input sentences

Learning procedure — Universal grammar

$G_1, G_2, ..., G_n$ — Search space

Candidate grammars

TRENDS in Cognitive Sciences

We assume there is a reward for mutual understanding. The payoff for someone who uses G_i and communicates with someone who uses G_j is given by:

$$F(G_i, G_j) = \frac{(s_{ij} + s_{ji})}{2}$$

This is simply the average taken over the two situations when G_i talks to G_j and when G_j talks to G_i.

Denote by x_i the relative abundance of individuals who use grammar G_i. Assume that everybody talks to everybody else with equal probability. Therefore, the average payoff for all those individuals who use grammar G_i is given by:

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

We assume that the payoff derived from communication contributes to biological fitness; individuals leave offspring proportional to their payoff. These offspring inherit the UG of their parents. They receive language input (sample sentences) from their parents and develop their own grammar. At first, we will not specify a particular learning mechanism but introduce the stochastic matrix, Q, whose elements, q_{ij} denote the probability that a child born to an individual using G_i will develop G_j. (In this first model, we assume that each child receives input from one parent. We are currently working on models that allow input from several individuals.) The probabilities that a child will develop G_i if the parent uses G_i is given by q_{ii}. The quantities q_{ii} measure the accuracy of grammar acquisition. If $q_{ii} = 1$ for all i, then grammar acquisition is perfect for all candidate grammars.

The population dynamics of grammar evolution are given by the following system of ordinary differential equations, which we call the 'language dynamics equations':

$$\frac{dx_j}{dt} = \sum_{i=1}^{n} f_i q_{ij} x_i - \phi x_j, \; j = 1, \ldots, n$$

The term $-\phi x_j$ ensures that the total population size remains constant: the sum over the relative abundances, $\sum_i x_i$, is 1 at all times. The variable

$$\phi = \sum_{i=1}^{n} f_i x_i$$

denotes the average fitness or 'grammatical coherence' of the population. The grammatical coherence is given by the probability that a randomly chosen sentence of one person is understood by another person. It is a measure for successful communication in a population. If $\phi = 1$ all sentences are understood and communication is perfect. In general, ϕ is a number between 0 and 1.

The language dynamics equation is reminiscent of the quasispecies equation of molecular evolution[55], but has frequency dependent fitness values: the quantities f_i depend on the relative abundances, $x_1, \ldots, x_n$. In the limit of perfectly accurate language acquisition, $q_{ii} = 1$, we recover the replicator equation of evolutionary game theory[29]. Thus, our model provides a connection between two of the most fundamental equations of evolutionary biology.

Evolution of grammatical coherence
In general, Eqn 1 admits multiple (stable and unstable) equilibria. For low accuracy of grammar acquisition (low values of q_{ii}), all grammars, G_i, occur with roughly equal abundance. There is no predominating grammar in the population. Grammatical coherence is low. As the accuracy of grammar acquisition increases, however, equilibrium solutions arise where a particular grammar is more abundant than all other grammars. A coherent communication system emerges. This means that if the accuracy of learning is sufficiently high, the population will converge to a stable equilibrium with one dominant grammar. Which one of the stable equilibria is chosen, depends on the initial condition.

The accuracy of language acquisition depends on UG. The less restricted the search space of candidate grammars is, the harder it is to learn a particular grammar. Depending on the specific values of s_{ij} some grammars may be much harder to learn than others. For example, if a speaker using G_i has a high probability of formulating sentences that are compatible with many other grammars (s_{ij} close to 1 for many different j) then G_i will be hard to learn. In the limit $s_{ij} = 1$, G_i is considered unlearnable, because no sentence can refute the hypothesis that the speaker uses G_i.

The accuracy of language acquisition also depends on the learning mechanism that is specified by UG. An inefficient learning mechanism or one that evaluates only a small number of input sentences will lead to a low accuracy and hence prevent the emergence of grammatical coherence.

We can therefore ask the crucial question: which properties must UG have such that a predominating grammar will evolve in a population of speakers? In other words, which UG can induce grammatical coherence in a population? As outlined above, the answer will depend on the learning mechanism and the search space. We can derive results for two learning mechanisms that represent reasonable boundaries for the actual, unknown learning mechanism used by humans.

The memoryless learning algorithm, a favorite with learning theorists, makes little demands on the

Questions for future research

- What are the consequences of small population sizes and stochasticity on the dynamics of grammar acquisition? How does the coherence threshold depend on population size?
- Can we formulate a tractable model, where each individual learns grammar or lexicon from several other individuals (not just from one other individual)? What difference does it make?
- In a spatial model, do the equations lead to different grammars in different regions? Are such patterns stable?
- Can we obtain exact results on the competition between variants of UG that differ in their search space?
- Can we formulate a unified model of language acquisition that includes both grammar and lexicon learning?
- What is the consequence of introducing specific assumptions about the rules that specify the candidate grammars?
- What are the language dynamics for infinitely large search spaces with prior probability distribution?
- An interesting difference between humans and animals is that human communication can be stimulus free: messages are often not prompted by environmental stimuli. How can we explain this behavioral difference? What is its adaptive significance?

cognitive abilities of the learner. It describes the interaction between a teacher and a learner. (The 'teacher' can be one or several individuals or the whole population.) The learner starts with a randomly chosen hypothesis (say G_i) and stays with this hypothesis as long as the teacher's sentences are compatible with this hypothesis. If a sentence arrives that is not compatible, the learner will at random pick another candidate grammar from his search space. The process stops after a certain number of sentences. The algorithm is called 'memoryless', because the learner does not remember any of the previous sentences nor which hypotheses have already been rejected. The algorithm works, primarily because once it has the correct hypothesis it will not change anymore (this is incidentally the definition of so called 'consistent learners').

The other extreme is a batch learner (resembling Jorge Louis Borges' man with infinite memory). The batch learner memorizes all sentences and at the end chooses the candidate grammar that is most compatible with the input.

For the memoryless learner, we can show that, under some assumptions on the values s_{ij}, grammatical coherence is possible if the number of input sentences, b, exceeds a constant times the number of candidate grammars, $b > C_1 n$. For the batch learner, the number of input sentences has to exceed a constant times the logarithm of the number of candidate grammars, $b > C_2 \log n$. These inequalities define a 'coherence threshold', which limits the size of the search space relative to the amount of input available to the child. A UG that does not fulfill the coherence threshold does not lead to a stable, predominating grammar in a population. The learning mechanism used by humans will perform better than the memoryless learner and worse than the batch learner; hence it will have a coherence threshold somewhere between $b > C_1 n$ and $b > C_2 \log n$.

Cultural evolution of grammar

The language dynamics equation describes deterministic dynamics for a large population size. Smaller population sizes can play a role if we consider stochastic language dynamics. Computer simulations suggest that the equilibrium solutions of the deterministic system correspond to metastable states. Individual grammars will dominate for some time and then be replaced by other grammars. Such transitions are more likely to occur between similar grammars.

In a small population, the requirements imposed on UG are also slightly stronger. Grammatical coherence in a population will require a larger number of input sentences or smaller search spaces. A detailed mathematical study of the stochastic dynamics of our system is still outstanding.

Individual candidate grammars, G_i, can also differ in their performance. Some grammars can be less ambiguous or describe more concepts than others. In such a context, the language dynamics equation can describe a cultural evolutionary optimization of grammar within the space of grammars generated by UG. It also provides a general framework for studying the dynamics of grammar change in the context of historical linguistics[56,57].

Biological evolution of universal grammar

So far we have assumed that all individuals have the same UG. Studying the biological evolution of UG, we need variation in UG and a system that describes natural selection among variants of UG.

At first, let us consider universal grammars with the same search space and the same learning procedure, the only difference being the number of input sentences, b (Ref. 58). This quantity is proportional to the length of the learning period. We find that natural selection leads to intermediate values of b. For small b, the accuracy of learning the correct grammar is too low. For large b, the learning process takes too long (and thus the rate of producing children that have acquired the correct grammar is too low). This observation can explain why there is a limited language acquisition period in humans.

Second, consider universal grammars, U_1 and U_2, that differ in the size of their search space, n, but have the same learning mechanism and the same value of b. In general, there is selection pressure to reduce n. Only if n is below the coherence threshold, can the universal grammar induce grammatical communication. In addition, the smaller n, the larger is the accuracy of grammar acquisition. There can, however, also be selection for larger n: suppose universal grammar U_1 is larger than U_2 (that is $n_1 > n_2$). If all individuals use a grammar, G_1, that is both in U_1 and U_2, then U_2 is selected. Now imagine that someone invents a new advantageous grammatical concept which leads to a modified

grammar G_2 which is in U_1, but not in U_2. In this case, the larger universal grammar is favored. Hence there is selection both for reducing the size of the search space and for remaining open minded to be able to learn new concepts. For maximum flexibility, we expect search spaces to be as large as possible but still below the coherence threshold.

An interesting extension of the above model is obtained by assuming that UG is only very roughly defined by our genes. Randomness during the developmental process could give rise to variation in neuronal patterns in the brain and consequently to variation in UG. Hence it might be a reasonable assumption that individuals have slightly different UGs. Each individual could have a personal 'universal' grammar. An interesting question is how similar these UGs have to be such that a population achieves grammatical coherence. In this case, there is again selection for maintaining a large search space of candidate grammars, as the target grammar should be contained in each of the UGs.

Conclusions

In summary, we have outlined how populations can evolve coherent communication, both in terms of lexical items and grammatical rules. We have described how arbitrary signals become associated with specific referents and have shown how natural selection can lead to the 'duality of patterning' of human language: words are sequences of phonemes, sentences are sequences of words. Finally, we have formulated a mathematical theory for the population dynamics of grammar acquisition. The key result here is a 'coherence threshold' that relates the maximum complexity of the search space to the amount of linguistic input available to the child and the performance of the learning procedure. The coherence threshold represents an evolutionary stability condition for the language acquisition device: only a universal grammar that operates above the coherence threshold can induce and maintain coherent communication in a population.

References

1 Smith, W.J. (1977) *The Behaviour of Communicating*, Harvard University Press
2 Deacon, T. (1997) *The Symbolic Species*, Penguin Books
3 Pinker, S. and Bloom, A. (1990) Natural language and natural selection. *Behav. Brain Sci.* 13, 707–784
4 Pinker, S. (1994) *The Language Instinct*, W. Morrow & Co.
5 Jackendoff, R. (1999) Possible stages in the evolution of the language capacity. *Trends Cognit Sci.* 3, 272–279
6 Jackendoff, R. *Foundations of Language* (in press)
7 Bickerton, D. (1990) *Language and Species*, University of Chicago Press
8 Cavalli-Sforza, L.L. and Feldman, M.W. (1981) *Cultural Transmission and Evolution: A Quantitative Approach*, Princeton University Press.
9 Aoki, K. and Feldman, M.W. (1987) Toward a theory for the evolution of cultural communication: Coevolution of signal transmission and reception, *Proc. Natl. Acad. Sci. U. S. A.* 84, 7164–7168
10 Newmayer, F. (1991) Functional explanation in linguistics and the origin of language. *Lang. Comm.* 11, 3–96
11 Lightfoot, D. (1999) *The Development of Language: Acquisition, Changes and Evolution*, Blackwell/Maryland Lecture in Language and Cognition
12 Brandon, R. and Hornstein, N. (1986) From icons to symbols: some speculations on the origins of language. *Biol. Philos.* 1, 169–189
13 Hurford, J.R. *et al.*, eds (1998) *Approaches to the Evolution of Language*, Cambridge University Press
14 Lieberman, P. (1991) *Uniquely Human: The Evolution of Speech, Thought, and Selfless Behavior*, Harvard University Press
15 Lieberman, P. (1984) *The Biology and Evolution of Language*, Harvard University Press
16 Maynard Smith, J. and Szathmary, E. (1995) *The Major Transitions in Evolution*, W.H. Freeman
17 Hawkins, J.A. and Gell-Mann, M. (1992) *The Evolution of Human Languages*, Addison-Wesley
18 Dunbar, R. (1996) *Grooming, Gossip, and the Evolution of Language*, Cambridge University Press
19 Aitchinson, J. (1987) *Words in the Mind. An Introduction to the Mental Lexicon*, Blackwell Science
20 Fitch, W.T. (2000) The evolution of speech: a comparative review. *Trends Cognit. Sci.* 4, 258–267

21 Hauser, M.D. (1996) *The Evolution of Communication*, Harvard University Press
22 Hurford, J.R. (1989) Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua* 77, 187–222
23 Cangelosi, A. and Parisi, D. (1998) The emergence of a 'language' in an evolving population of neural networks. *Connect. Sci.* 10, 83–97
24 Steels, L. (1996) Self-organizing vocabularies. *Proc. Artificial Life* (Vol. 5), MIT Press
25 Nowak, M.A. and Krakauer, D.C. (1999) The evolution of language. *Proc. Natl. Acad. Sci. U. S. A.* 96, 8028–8033
26 Nowak M.A., *et al.* (1999) The evolutionary language game. *J. Theor. Biol.* 200, 147–162
27 Miller, G. A. (1996) *The Science of Words*, Scientific American Library
28 Maynard Smith, J. (1982) *Evolution and the Theory of Games*, Cambridge University Press
29 Hofbauer, J. and Sigmund, K. (1998) *Evolutionary Games and Replicator Dynamics*, Cambridge University Press
30 Trapa, P.E. and Nowak, M.A. (2000) Nash Equilibria for an evolutionary language game. *J. Math. Biol.* 41, 172–188
31 Komarova, N.L. and Nowak, M.A. (2001) The evolutionary dynamics of the lexical matrix. *Bull. Math. Biol.* 63, 451–485
32 Nowak, M.A. *et al.* (1999) An error limit for the evolution of language. *Proc. R. Soc. London Ser. B.* 266, 2131–2136
33 Plotkin, J.B. and Nowak, M.A. (2000) Language evolution and information theory. *J. Theor. Biol.* 205, 147–159
34 Shannon, C.E. and Weaver, W. (1949) *The Mathematical Theory of Information*, University of Illinois Press
35 Cheney, D.L. and Seyfarth, R.M. (1990) *How Monkeys see the World*, University of Chicago Press
36 Nowak, M.A. *et al.* (2000) Evolution of syntactic communication. *Nature* 404, 495–498
37 Chomsky, N. (1965) *Aspects of the Theory of Syntax*, MIT Press
38 Wexler, K. and Culicover, P. (1980) *Formal Principles of Language Acquisition*, MIT Press

39 Jackendoff, R. (1997) *The Architecture of the Language Faculty*, MIT Press
40 Chomsky, N. (1972) *Language and Mind*, Harcourt Brace
41 Tomasello, M. (1995) Language is not an instinct. *Cognit. Dev.* 10, 131–156
42 Bates, E. (1984) Bioprograms and the innateness hypothesis. *Behav. Brain Sci.* 7, 188–190.
43 Langacker, R. (1987) *Foundations of Cognitive Linguistics* (Vol. 1), Stanford University Press
44 Gold, E.M. (1967) Language identification in the limit. *Inform. Control* 10, 447–474
45 Gibson, E. and Wexler, K. (1994) Triggers. *Linguist. Inquiry* 25, 407–454
46 Hornstein, N.R. and Lightfoot, D.W. (1981) *Explanation in Linguistics*, Longman
47 Lightfoot, D. (1991) *How to Set Parameters: Arguments from Language Change*, MIT Press
48 Manzini, R. and Wexler, K. (1987) Parameters, binding theory, and learnability. *Linguist. Inquiry* 18, 413–444
49 Niyogi, P. (1998) *The Informational Complexity of Learning*, Kluwer Academic Publishers
50 Osherson, D. *et al.* (1986) *Systems that Learn*, MIT Press
51 Vapnik, V. (1995) *The Nature of Statistical Learning Theory*, Springer-Verlag
52 Valiant, L.G. (1984) A theory of the learnable. *Commun. ACM* 27, 436–445
53 Nowak, M.A. *et al.* (2001) Evolution of universal grammar. *Science* 291, 114–118
54 Komarova, N.L. *et al.* (2001) Evolutionary dynamics of grammar acquisition, *J. Theor. Biol.* 209, 43–59
55 Eigen, M. and Schuster, P. (1979) *The Hypercycle: A Principle of Natural Self-Organisation*, Springer-Verlag
56 Niyogi, P. and Berwick, R.C. (1996) A language learning model for finite parameter spaces. *Cognition* 61, 161–193
57 Niyogi, P. and Berwick, R.C. (1997) Evolutionary consequences of language learning. *Linguist. Philos.* 20, 697–719
58 Komarova, N.L. and Nowak, M.A. (2001) Natural selection of the critical period for language acquisition. *Proc. R. Soc. London Ser. B.* 268, 1189–1196