Z. Phys. Chem. **216** (2002) 5–20
© by Oldenbourg Wissenschaftsverlag, München

From Quasispecies to Universal Grammar

By Martin A. Nowak*

Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA

Dedicated to Prof. Dr. Peter Schuster
on the occasion of his 60th birthday

(Received June 22, 2001; accepted July 13, 2001)

Molecular Evolution / Sequence Space / Fitness Landscape / Evolutionary Dynamics / Language Acquisition

The perspective of this paper is to compare mathematical models for the evolutionary dynamics of genomes and languages. The quasispecies equation describes the evolution of genetic sequences under the influence of mutation and selection. A central result is an error threshold which specifies the minimum replication accuracy required for maintaining genetic information of a certain length. The language equation describes the evolution of communication, including the cultural evolution of grammar and the biological evolution of universal grammar. A central result is a coherence threshold which specifies certain conditions that universal grammar has to fulfill in order to induce coherent communication in a population.

Languages and quasispecies have something in common. Quasispecies replicate under the influence of mutation and selection and so do languages. A quasispecies is an ensemble of related, but different chemical molecules carrying genetic information. Any two individuals of a biological population have very similar, but not necessarily identical genomic sequences. Similarily a group of humans of the same speech community speak almost the same language, but any two individuals may slightly differ with respect to certain linguistic features in terms of phonetics, syntax or semantics. Ultimately, what we call ‘English’ is a social construct and there may be no two English speakers with identical linguistic properties.

Quasispecies theory was originally invented, by Manfred Eigen and Peter Schuster, to describe chemical kinetics guiding the origin of life [1–5]. There are some parallels between the origin of life and the origin of language. Both are mysterious. Both are major events in evolutionary history, that led to something totally new. Language led to the origin of a new mode of evolution. For

---

* E-mail: nowak@ias.edu.

billions of years of life on earth, evolution was almost completely restricted to information in form of genetic sequences. Language allows the transfer of unlimited, non-genetic information among individuals and from one generation to the next. Thus language generates a new mode of evolution: cultural evolution. Certainly higher animals can also have some form of cultural evolution, but human language provides the machinery for unlimited cultural evolution [6].

When Eigen and Schuster conceived quasispecies theory they did not want to reconstruct the historical events of the origin of life, but instead formulate the natural laws that must have been operating. This is also my attempt with respect to the evolution and origin of language. Let us understand the fundamental principles that guide the evolution of language and animal communication [7–20].

This paper is in two parts. First, I review some aspects of quasispecies theory, then I will discuss mathematical models for the evolutionary dynamics of language.

1. Quasispecies

The word, quasispecies, is confusing for many biologists, possibly because it was invented by chemists. Manfred Eigen and Peter Schuster wanted to develop a chemical theory for the origin of life. They described how populations of RNA molecules could reproduce themselves. They noted that the spontaneous chemical reproduction of such comparatively simple molecules was much less accurate than the genetic replication of any organism currently alive. The first proliferation of biological information was thus an extremely error-prone process. Random events would lead to mutations. The reproduction of such molecules by base-pairing is not completely accurate, so a certain degree of mis-matching is to be expected. Consequently a population of RNA molecules that was the result of such an inaccurate replication process would not be homogeneous, but a mixture of RNA molecules with different nucleotide sequences.

Chemists refer to an ensemble of equal molecules as “species”. For example, the “species” of all H_2O molecules. In contrast, a species of RNA molecules, derived by inaccurate reproduction, is not an ensemble of *identical* molecules. Hence the term “quasispecies”. For biologists, the term is confusing because a biological species is a complicated and in some sense losely defined concept. How much more unspecific is then a “quasispecies”!

Eigen and Schuster were primarily interested in the origin of life. They assumed that RNA was the first biological replicator. In the primordial soup, the four nucleotides (adenin, guanosin, cytidin and thymidin) would spontaneously form short chains, so called polymers. These polymers could in principle reproduce by base-pairing. Alongside of each polymer another polymer would form which consisted of the complimentary sequence of nucleotides. Subsequently, the double-helix could split up and the two single strands would go on to form new double helices, thereby imprinting their sequences on new polymers. This process, which is now the basis of all life and is conducted by highly sophisticated and accurate enzymes inside the cell, might have been occuring initially very slowly and subject to high error rates.

This primitive genetic replication is a chemical process and can be described by chemical kinetics, that is by equations specifying how the concentration of certain molecules change over time. The kinetics of RNA self-replication assumes that molecules have different replication rates according to their sequence. This means that some may produce “offspring” faster than others; they are “fitter”. In addition the theory takes into account that replication is inaccurate. An offspring sequence need not be identical to its parent, but may differ in certain positions. A substitution of one base for another is called a “point mutation.” The equations do not lead to a population of identical sequences but to an ensemble of related but different sequences. This ensemble is called quasispecies.

More precisely, Eigen and Schuster refer to the *equilibrium distribution* of sequences that is formed by this mutation and selection process as quasispecies. “Mutation” because reproduction is subject to errors. “Selection” because sequences have different fitnesses. Eigen and Schuster go on to argue that the target of natural selection is not the fittest sequence, but the quasispecies. Natural selection will not just chose the fittest type, but an ensemble of different variants. The fittest sequence may only represent a very small fraction of the quasispecies; it may indeed not be present at all.

An important concept of quasispecies theory is the *error-threshold*. If the mutation rate is too high, that is if too many mistakes occur in any one replication event, then the population will be unable to maintain any genetic information. In the long run the composition of the quasispecies will only be determined by randomness. The abundance of individual sequences will be independent of their fitness. Thus the error rate must be below a critical threshold level for the system to maintain information. If the error-rate is expressed as a per base probability to make a mistake, then the error-threshold can be written as a condition which limits the maximum length of the RNA sequence. For any given per-base fidelity the sequence can only reach a certain length. Eigen and Schuster estimated that this critical length should be roughly 100 bases for self-replicating RNA molecules in the primordial soup.

Since every biological reproduction is error prone, the quasispecies concept can be readily applied to genetic processes other than RNA self-replication. Populations of viruses, bacteria, plants or animals are quasispecies. Their genetic reproduction is of course more complicated than a simple copying of the sequence and will include more sophisticated mutational events (such as recombination, or sexual reproduction), but the underlying principles remain the same. Any natural biological population will be mix of genomes, a quasispecies.

The population of human immunodeficiency viruses in an infected individual represent a quasispecies that can escape from drug treatment and immune responses. In my opinion, HIV quasispecies evolution in individual patients holds the key to understanding disease progression [21].

The error-threshold concept can also be expanded to more complicated organisms. Very roughly, the genome length should not exceed one over the per-base mutation rate. Viral RNA replication, in the absence of any error-correcting mechanisms (proof-reading), has a per base mutation rate of about 10^{-4}. As predicted, the genome length of such viruses is thus about 10^4. The human genome is about $3 \cdot 10^9$ bases long, and the high quality DNA replication enzymes in human cells ensure an error rate of about 10^{-9} per base. The rule of thumb that the genome lenght equals one over the per-base mutation rate holds for many different organisms.

1.1 More than atoms in our universe

Consider an RNA or DNA sequence of length L denoting the number of bases in the sequence. There are 4^L different variants. This means that even for moderate lengths a ‘hyper-astronomically’ large number of different variants can be formed. For example, for a polynucleotide of length $L = 300$, which is just large enough to encode for one of the smallest proteins, there are more than 10^{180} different variants. There are only 10^{80} protons in our universe.

1.2 Quasispecies live in sequence space

The correct geometry for quasispecies is given by the sequence space. In the sequence space all possible variants of a given length are arranged such that neighbours differ by only one base substitution. More generally, the distance between two sequences equals the number of substitutions between them. For example, the sequence AATCG differs from ATCCG in two postions. The dimension of the space is given by the length of the sequence. In each dimension there are 4 possibilities, corresponding to the 4 nucleotides, A, T, C and G. The sequence space that contains all sequences of length 5 has 5 dimensions and $4^5 = 1024$ points (different sequences).

The important features of this sequence space are (1) its high dimensionality, (2) the large number of shortest mutational routes between two distant mutant sequences (for two sequences seperated by d point mutations there are $d!$ shortest mutational routes) and (3) that many sequences are confined to a close neighbourhood of each other. The diameter of a sequence space that contains 10^{80} points is only 133 “length units”, *i.e.* point mutations. This means that relatively few point mutations can lead from one region in the sequence space to a completely different region providing there exists something like a guiding gradient to avoid going into ‘wrong directions’. In evolution, this gradient is provided by natural selection.

1.3 Quasispecies explore fitness landscapes

Every point in sequence space can also be assigned a fitness value (representing the reproduction rate of this sequence). This leads to the concept of a fitness landscape. Fitness landscapes have one more dimension than the corresponding sequence space, because for every point there is a “height”. Quasispecies wander over the fitness landscapes searching for peaks, which represents region of high fitness values. Under the guidance of natural selection quasispecies climb the mountains in the high dimensional fitness landscape [22, 23].

Here again we can easily envision how natural selection does not simply choose the fittest sequence, but the quasispecies. Imagine two sequences A and B. Assume A has a higher replication rate than B, thus it has a higher intrinsic fitness value. Suppose A is surrounded by mutants with very low fitness, while B is sourrounded by mutants with high fitness. (Both A and B are local optima, but A is a very sharp peak in the fitness landscape, whereas B is the top of a flat mountain.) In the absence of mutation, A will be selected and B will disappear. With mutation, however, the situation can change and B could be the winner. In fact, the mathematical equations will show a critical mutation rate, below which A is the winner, but above which B and its neighbours are favoured [24, 25].

1.4 The quasispecies equation

Suppose there are n different nucleic acid sequences, $x_1, ..., x_n$, with replication rates $a_1, ..., a_n$. These quantities represent the selective values of the individual mutants. In the absence of mutation the variant with the highest replication rate will grow fastest and reach fixation. The result of selection in this world without errors is a homogeneous population consisting of the fastest replicating variant. But replication is not error-free. Thus it is necessary to define the probabilities q_{ij} that (erroneous) replication of template x_j results in the production of the sequence x_i. The quantities q_{ij} form the $n \times n$ mutation matrix, Q.

If we consider binary sequences and point mutations, we obtain

$$q_{ij} = p^{H_{ij}}(1-p)^{(L-H_{ij})} .$$

Here p is the mutation rate per bit, L is the length of the bitstring, and H_{ij} is the Hamming distance between strains i and j, that is the number of bits in which the two strains differ. Error free replication is given by $q_{ii} = (1-p)^L$.

The quasispecies equation is given by

$$\dot{x}_i = \sum_{j=1}^{n} a_j q_{ij} x_j - \phi x_i \qquad i = 1, ..., n . \tag{1}$$

The variants x_j replicate at rate a_j and generate mutants, x_i, with probability q_{ji}. Hence the sum determines the total rate at which x_i variants are being produced by the population. The death term, ϕx_i, is chosen to keep the total population size constant. This is the case if ϕ denotes the average fitness of the population, given by $\phi = \sum_{i=1}^{n} a_i x_i$. Without loss of generality, we set $\sum_{i=1}^{n} x_i = 1$.

Eq. (1) is nonlinear, because of the term, ϕx_i. Interestingly, however, this term can be dropped. For mathematical purposes, we can also study the linear system

$$\dot{x}_i = \sum_{j=1}^{n} a_j Q_{ij} x_j \qquad i = 1, ..., n \, .$$

Mathematical physicists prefer to write

$$\dot{x}_i = \sum_{j=1}^{n} Q_{ij} x_j + a_i x_i \qquad i = 1, ..., n$$

which is reminiscent of a random Schroedinger equation. Here, mutation acts like a disrete Laplace operator, while selection acts like a potential. Therefore, studying deterministic quasispecies dynamics reduces to analyzing the spectral properties of the linear mutation-selection operator.

The frequency of a given variant within the quasispecies does not depend on its replicative value alone, but also on the likelihood with which it is produced by erroneous replication of other templates and their frequencies in the quasispecies distribution. This is important to the understanding of the structural organization of a quasispecies. The consequence of this effect is that the individual sequence x_i with its replicative value a_i no longer serves as the unit (or target) of selection. The quasispecies itself is the target of selection in a mutation-selection process. This fact has important implications. Evolution is normally thought of as the interaction between mutation and selection. Selection is a factor that favours advantageous mutants that have been generated by pure chance; indeed, it is normally considered a mistake to think of mutations as being guided other than by chance. A quasispecies, however, can guide mutations. This does not mean that there is any correlation between the (intrinsically stochastic) act of mutation and the selective advantage of the mutant. But selection operates on the structure of the whole quasispecies which is adapted to its *fitness lanscape* (this term is originally from Sewall Wright). Therefore evolution can be guided towards the peaks of this fitness landscape. This happens because more successful mutants (that may be in closer neighborhood to the peaks of the landscape) will produce more offspring than less successful mutants (which may be further away of the peaks). Evolutionary optimization can be viewed as a hill-climbing process of the quasispecies that occurs along certain pathways in sequence space [26–28].

1.5 Error threshold

An important concept in quasispecies theory is the error threshold of replication. If replication were error free, no mutants would arise and evolution would stop. Evolution would, however, also be impossible if the error rate of replication were too high (only some mutations may lead to an improvement in adaptation, but most will lead to deterioration). The quasispecies concept allows us to quantify the resulting minimal replication accuracy that maintains adaptation [29–32].

Let us assume a population consists of (1) a fast replicating variant x_1 – the wild type sequence – with replication rate a_1 and (2) its mutant distribution (error tail) x_2 with a lower average replication rate a_2. Let q denote the per base accuracy of replication, *i.e.* the probability that a single base is accurately replicated. Thus the probability that the whole sequence (of length L) is replicated without errors is given by $Q = q^L$. Neglecting the small probability that erroneous replication of a mutant gives rise to a wild type sequence leads to the equations

$$\begin{aligned}\dot{x}_1 &= a_1 Q x_1\\ \dot{x}_2 &= a_1(1-Q)x_1 + a_2 x_2\,.\end{aligned}$$

Here the ratio of wild type over mutants converges to

$$x_1/x_2 \to \frac{a_1 Q - a_2}{a_1(1-Q)}\,.$$

Therefore the wild type can only be maintained in the population if $Q > a_2/a_1$. This means that the single digit replication accuracy, q, must be larger than a certain critical value. This error threshold relation is obtained as

$$q > q_{\mathrm{crit}} = (a_2/a_1)^{1/L}\,.$$

For replication accuracies lower than q_{crit} the wild type sequence will be lost from the population although it has the highest replication rate. This leads to an important relationship between the replication accuracy and the sequence length

$$L < 1/(1-q)\,.$$

Here I have used the approximation that the logarithm of a_1/a_2 is about 1. This represents an approximation for the upper genome length L that can be maintained by a given single digit replication accuracy without loosing adaptation.

2. Universal grammar

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. Ray Jackendoff proposed that grammar has a parallel architecture consisting of three, largely independent combinatorial systems, phonetics, syntax and semantics. All three subsystems are generative and are linked via interfaces [11, 33].

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 all grammatical sentences. Chomsky pointed out that the evidence available to the child does not uniquely determine the underlying grammatical rules [34]. This phenomenon is called the ‘poverty of stimulus’ [35]. The ‘paradox of language acquisition’ [36] is that children nevertheless reliably achieve correct grammatical competence. How is this possible?

As Chomsky [34] 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.’ Chomsky introduced the term Universal Grammar (UG) to denote the preformed ‘linguistic theory’, the initial pre-specification of the form of possible human grammars [37].

Hence, for language acquisition the child needs an innate learning mechanism for processing the input sentences and an innate search space of candidate grammars from which to choose the appropriate grammar. UG can be defined as the grammar that generates the whole search space. The more recent convention, however, is that UG encompasses both the search space and the learning mechanism. In this scenario, UG is essentially synonymous with ‘language acquisition device’. Note that both the search space and the learning mechanism could change during language acquisition, but this does not negate the necessity of innate compoments.

This ‘biolinguistic approach’ was controversial when introduced by Chomsky about 30 years ago, and remains so [10, 38]. Note, however, that the ‘poverty of stimulus’ argument has a precise, mathematical formulation known as Gold’s theorem [39]. 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 new 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 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 logical necessities.

Hence, the controversy concerning a genetically determined UG cannot be whether there is one (this is a logical necessity), but what form it has. This is largely an emprirical question and will be resolved eventually.

Let us now formulate a mathematical description of language acquisition [40–43]. 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. Therefore, a grammar can be represented as a mapping from the set of all integers into $\{0, 1\}$. 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).

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 or an ordering on the set of all grammars. The learner needs to know which of the infinitely many grammars are apriori 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 [44–46], 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 [20].

2.1 Language quasispecies

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.

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

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

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 then given by the following system of ordinary differential equations, which we call the 'language dynamics equations'

$$\dot{x}_i = \sum_{j=1}^{n} f_j(\boldsymbol{x}) q_{ji} x_i - \phi(\boldsymbol{x}) x_j \qquad j = 1, ..., n \,.$$

As before, the term $\phi(\boldsymbol{x})x_i$ 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 similar to the quasispecies equation, but has frequency dependent fitness values: the quantities f_i depend on the relative abundances $x_1, ..., x_n$. In the limit of perfectly accurate language acquisition, $q_{ii} = 1$, we recover the *replicator equation* of evolutionary game theory [47–49]. Thus, the language equation is a replicator equation with mutation (sometimes called mutation selection equation) or a quasispecies equation with frequency dependent fitness. In contrast to the quasispecies equation, the language equation cannot be reduced to a linear equation.

2.2 Evolutionary dynamics of grammatical coherence

The language equation admits multiple (stable and unstable) equilibria. For low accuracy of grammar acquisition (low values 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 high probabilities 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_j.

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 UG must 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 employed by humans.

The memoryless learning algorithm, a favorite with learning theorists, makes little demands on the cognitive abilities of the learner (or the learning theorist). 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.

Consider a very symmetric case for the similarity between candidate grammars: let $s_{ij} = s$ for all $i \neq j$ and let $s_{ii} = 1$. In this special case, we can show for the memoryless learner that 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 $b > C_2 \log n$.

Fig. 1 shows the grammatical coherence, ϕ, of stable solutions of the the language equation for the case where the s_{ij} values are taken from a uniform distribution on [0, 1] and $s_{ii} = 1$. The memoryless learning algorithm is used. Again, if children consider a number of input sentences, b, below a certain threshold then the only stable solution consists of a nearly uniform mixture of all the grammars. As b increases, stable equilibria emerge that are dominited by individual grammars. Komarova and Rivin [50] showed that, for this case, the coherence threshold is given by $b > Cn \log n$.

2.3 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 meta-stable states. Individual grammars will dominate for some time and then be replaced by other grammars.

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

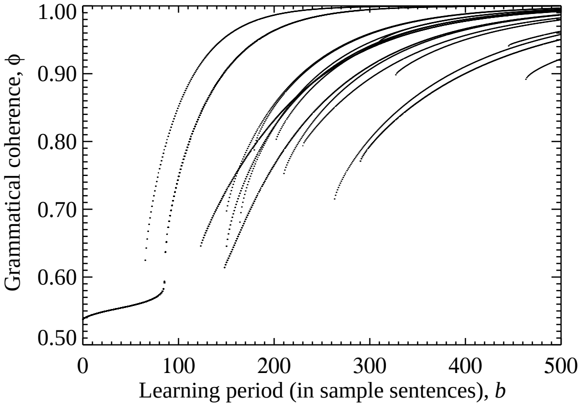

Fig. 1. Grammatical coherence of a population versus the number of sample sentences, b, per individual for stable equilibrium solutions of the language equation. There are $n = 20$ grammars with randomly chosen pairwise distances; the s_{ij} values (for $i \neq j$) are taken from a uniform distribution on [0, 1], and $s_{ii} = 1$. Children learn the grammar of their parents according to a memoryless learning algorithm. The grammatical coherence (or average fitness) of the population is given by $\phi = \sum_i x_i f_i$, where $f_i = (1/2) \sum_j x_j (s_{ij} + s_{ji})$. It is a measure of mutual understanding in the population. For small b all grammars occur at roughly similar frequency; the coherence is low. For larger values of b, stable equilibria appear with the majority of the population adopting the same grammar. The first critical transition occurs roughly at $b = 3.5n$ which is Eq. (2) with $a = 1/2$. Some grammars lead to stable equilibrium solutions only for large numbers of sample sentences. In the limit $b \to \infty$, there are n stable equilibria corresponding to all people using one of the n grammars.

UG. It also provides a general framework for studying the dynamics of grammar change in the context of historical linguistics [51].

2.4 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. 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, since the target grammar should be contained in each of the UGs. All of this requires more work.

3. Epilogue

This paper is dedicated to the 60th birthday of Peter Schuster. When Peter returned from Goettingen to Vienna, he persuaded Karl Sigmund and Josef Hofbauer to work on replicators and hypercycles. Hence, Peter founded the Vienniese School of mathematical biology. I met Peter in 1985 in the desolate lecture hall 1 of the Chemistry Department in Vienna. I was studying biochemistry. He gave a lecture course on theoretical chemistry. I picture him wearing a green suit and a red-chequered tie. I could be wrong. Meeting Schuster (in this outfit) changed my life. I immediately knew that I wanted to work with him. I did a Diploma thesis under his guidance on a stochastic formulation of quasispecies dynamics. Much of my work over the years has been inspired by quasispecies theory and replicator equations. Peter stood at the beginning of both. On the first day of my diploma thesis, I had a modified verse from Goethe’s Faust in my mind: ‘Das preisen die Schueler aller Orten, sind aber keine Schuster geworden’.

Acknowledgement

Support from the Packard Foundation, the Leon Levy and Shelby White Initiatives Fund, the Florence Gould Foundation, the Ambrose Monell Foundation, the Alfred P. Sloan Foundation, the National Science Foundation and Jeffrey Epstein is gratefully acknowledged.

References

1. M. Eigen, Naturwissenschaften **58** (1971) 465–523.
2. M. Eigen and P. Schuster, Naturwissenschaften **64** (1977) 541–565.
3. M. Eigen, J. McCaskill and P. Schuster, Adv. Chem. Phys. **75** (1989) 149–263.
4. P. Schuster, Origins Life **14** (1984) 3–14.
5. P. Schuster, Physica D **22** (1986) 100–119.
6. L. L. Cavalli-Sforza and M. W. Feldman, *Cultural transmission and evolution: a quantitative approach*, Princeton Univ. Press, Princeton, NJ (1981).
7. W. T. Fitch, Trends Cognit. Sci. **4** (2000) 258–267.
8. M. D. Hauser, *The Evolution of Communication*, Harvard University Press, Cambridge, MA (1996).
9. S. Pinker and A. Bloom, Behavioral Brain Sci. **13** (1990) 707–784.
10. S. Pinker, *The Language Instinct*, W. Morrow and Company, New York (1994).
11. R. Jackendoff, Trends Cognit Sci. **3** (1999) 272–279.
12. K. Aoki, M. W. Feldman, Proc. Nat. Acad. Sci. USA **84** (1987) 7164–7168.
13. F. Newmayer, Language Commun. **11** (1991) 3–96.
14. J. R. Hurford, M. Studdert-Kennedy, C. Knight (Eds.), *Approaches to the Evolution of Language*, Cambridge University Press, Cambridge (1998).
15. P. Lieberman, *Uniquely Human: The Evolution of Speech, Thought, and Selfless Behavior*, Harvard University Press, Cambridge, MA (1991).
16. M. J. Smith and E. Szathmary, *The Major Transitions in Evolution*, W. H. Freeman Spektrum, Oxford, New York (1995).
17. J. A. Hawkins, M. Gell-Mann, *The Evolution of Human Languages*, Addison-Wesley, Reading MA (1992).
18. M. A. Nowak and D. C. Krakauer, Proc. Nat. Acad. Sci. USA **96** (1999) 8028–8033.
19. M. A. Nowak, J. B. Plotkin, V. A. A. Jansen, Nature **404** (2000) 495–498.
20. M. A. Nowak, N. L. Komarova, P. Niyogi, Science **291** (2001) 114–118.
21. M. A. Nowak, R. M. Anderson, A. R. McLean, T. F. W. Wolfs, J. Goudsmit and R. M. May, Science **254** (1991) 963–969.
22. P. F. Stadler, J. Math. Chem. **20** (1996) 1–45.
23. P. F. Stadler, J. Mol. Structure **463** (1999) 7–19.
24. J. Swetina, P. Schuster, Biophys. Chem. **16** (1982) 329–345.
25. P. Schuster and J. Swetina, Bull. Math. Biol. **50** (1988) 635–660.
26. W. Fontana, W. Schnabl and P. Schuster, Phys. Rev. A **40** (1989) 3301–3321.
27. W. Fontana and P. Schuster, Biophys. Chem. **26** (1987) 123–147.
28. W. Fontana and P. Schuster, Science **280** (1998) 1451–1455.
29. J. S. McCaskill, J. Chem. Phys. **80** (1984) 5194–5202.
30. J. S. McCaskill, Biol. Cybernetics **50** (1984) 63–73.
31. M. Nowak and P. Schuster, J. Theor. Biol. **137** (1989) 375–395.
32. S. Bonhoeffer and P. F. Stadler, J. Theor. Biol. **164** (1993) 359–371.
33. R. Jackendoff, *Foundations of language*, Oxford University Press, Oxford (2001).
34. N. Chomsky, *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA (1965).
35. K. Wexler and P. Culicover, *Formal Principles of Language Acquisition*, MIT Press, Cambridge, MA (1980).
36. R. Jackendoff, *The Architecture of the Language Faculty*, MIT Press, Cambridge, MA (1997).
37. N. Chomsky, *Language and Mind*, Harcourt Brace Jovanovich, NY (1972).
38. M. Tomasello, Cogn. Development **10** (1995) 131–156.

39. E. M. Gold, Information Control **10** (1967) 447–474.
40. E. Gibson and K. Wexler, Linguistic Inquiry **25** (1994) 407–454.
41. D. Lightfoot, *How to Set Parameters: Arguments from Language Change*, MIT Press, Cambridge, MA (1991).
42. R. Manzini and K. Wexler, Linguistic Inquiry **18** (1987) 413–444.
43. P. Niyogi, *The Informational Complexity of Learning*, Kluwer Academic Publishers, Boston (1998).
44. V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York (1995).
45. L. G. Valiant, Commun. ACM **27** (1984) 436–445.
46. D. Osherson, M. Stob and S. Weinstein, *Systems That Learn*, MIT Press, Cambridge, MA (1986).
47. J. Maynard Smith, *Evolution and the Theory of Games*, Cambridge University Press, Cambridge, U.K. (1982).
48. P. Schuster, K. Sigmund, J. Theor. Biol. **100** (1983) 533–538.
49. J. Hofbauer, K. Sigmund, *Evolutionary games and replicator dynamics*, Cambridge University Press (1998).
50. N. L. Komarova, Rivin, I. Mathematics of learning, submitted (2001).
51. P. Niyogi and R. C. Berwick, Cognition **61** (1996) 161–193.