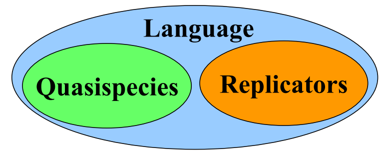
From quasispecies to universal grammar

Martin Nowak
Institute for Advanced Study, Princeton

- Quasispecies
- Replicators
- Language



deterministic evolutionary dynamics

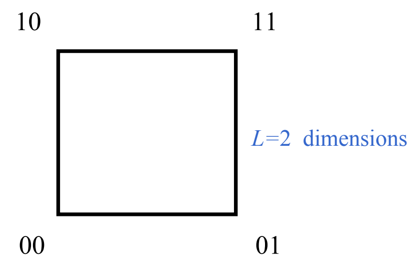
A quasispecies

is a population of RNA or DNA molecules:

ATCAGGACTCA	0000110011000110
ATCGGGACTCA	0000110011100110
ATCAGGAATCA	1000110011000010
….	….

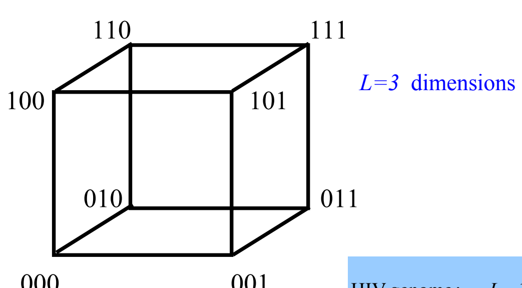
Manfred Eigen & Peter Schuster

Quasispecies live in sequence space



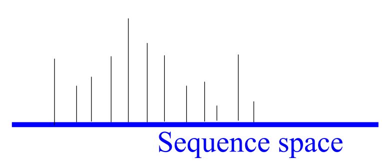
Hamming (or Manhattan) metric

Quasispecies live in sequence space



HIV genome: $L = 10^4$

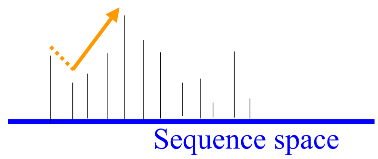
Fitness landscape



Each sequence has a reproduction rate (=fitness).

Evolution

… is adaptation of the quasispecies on the fitness landscape.



The Quasispecies equation

$$\dot{x}_i = \sum_{j=1}^{n} x_j f_j Q_{ji} - \phi(\vec{x}) x_i$$

Frequency of type i

Mutation from j to i

Fitness of type j

Average fitness

$\phi(\vec{x}) = \sum_i f_i \, x_i$

$\sum_i x_i = 1$

Error threshold

Adaptation is only possible if the mutation rate is below a certain threshold.

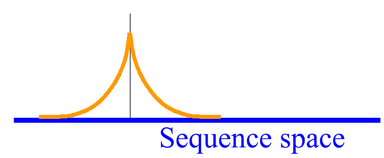
$$u < 1 / L$$

Mutation rate per base

Genome length

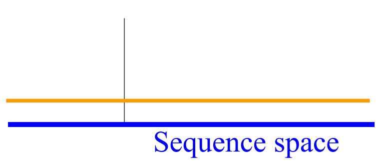
Adaptation = localization in sequence space

$$u < 1 / L$$

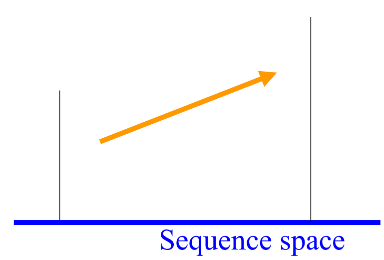


No adaptation

$$u > 1 / L$$



The rate of transition between peaks is maximized close to the error threshold



Applications of quasispecies theory

- Origin of life
- Experimental evolution
- Viruses
- Bacteria
- Cancer

Replicators

Evolutionary game theory

Game Theory

John von Neumann

Oskar Morgenstern

Evolutionary game theory

- Application of game theory to evolution of animal behavior.
- Successful strategies spread by natural selection.

John Maynard Smith

Evolutionarily stable strategy

If every individual of a population adopts the evolutionarily stable strategy, then no mutant can invade.

Nash equilibrium

Replicator equation

Frequency of type i

Average fitness

$$\dot{x}_i = x_i[f_i(\vec{x}) - \phi(\vec{x})]$$

Fitness of type i is frequency dependent

$$f_i = \sum_j a_{ij} x_j \qquad \phi = \sum_i f_i x_i$$

$$\sum_i x_i = 1$$

Payoff matrix

Josef Hofbauer & Karl Sigmund

Lotka-Volterra equation

$$\dot{x}_i = x_i(r_i + \sum_j a_{ij} x_i) \quad i = 1,..,n$$

describes the competition of species in an ecosystem.

Ecology

Lotka-Volterra and replicator equations

$$\dot{x}_i = x_i(r_i + \sum_j a_{ij} x_i) \quad i = 1,..,n$$

$$\dot{x}_i = x_i[\sum_j a_{ij} x_i - \phi(\vec{x})] \quad i = 1,..,n+1$$

… are equivalent.

Evolutionary game theory

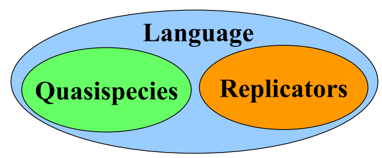
=

Frequency dependent selection

=

Ecology

Language



Natalia Komarova (Princeton)
Partha Niyogi (Chicago)

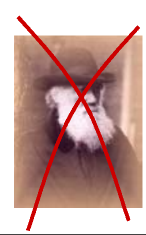
Why work on language evolution?

- … because the linguistic society of Paris officially banned any work on language evolution at a meeting in 1866.

Why work on language evolution?

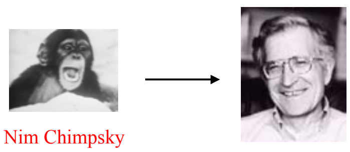
- … because some people said language might not have arisen by Darwinian evolution.

Noam Chomsky



Why work on language evolution?

- … because one view is that language came as the by-product of a big brain.



Nim Chimpsky — Noam Chomsky

Language is the most interesting thing that evolved in the last 600 million years

?	Origin of life
3.5	Prokaryotes
1.5	Eukaryotes
0.6	Multicellular Organisms
0.001	Language

(billions of years ago)

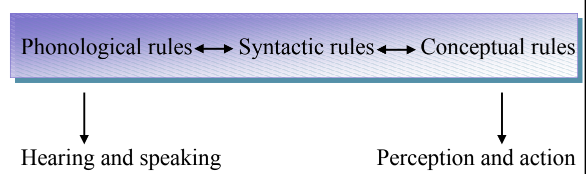
Language gives rise to a new mode of evolution.

Goals of this research program

- Formulate an evolutionary theory for how language changes over time
- Describe the emergence of the basic design features of human language: arbitrary signs, words, syntactic signals, grammar
- Make connection between formal language theory, learning theory and evolutionary dynamics

Grammar is

- the computational system of language
- a rule system that links phonetic forms and semantic forms



Grammar acquisition

- Children acquire the grammar of their native language by hearing sentences.
- This information does not uniquely determine the underlying grammatical rules.

Poverty of stimulus

Universal grammar

- Children could not guess the correct grammar if they had no preformed, innate expectation.
- This innate expectation is **universal grammar**.

Noam Chomsky

Formal language theory

- Alphabet: {0,1}
- Sentences:{0,1,00,01,10,11,000,…}
- a language is a subset of sentences: e.g. *L*= {01, 0011, 000111,...}

Grammar

- A grammar is a finite list of rules that specifies a language.

$$S \rightarrow 0S1$$
$$S \rightarrow \varepsilon$$

$$L = 0^n 1^n$$

Machines

- There is an equivalence between languages, grammars and machines.

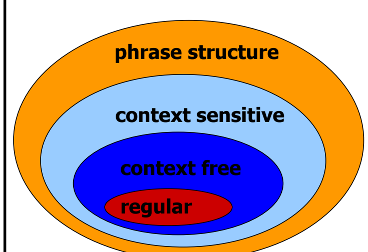
Chomsky hierarchy

Language	Grammar	Machine
Regular language	Finite state grammar	Finite state automata
Context free language	Context free grammar	Push down automata
Decidable language	Context sensitive grammar	Linear bounded automata
Computable language	Phrase structure grammar	Turing machine

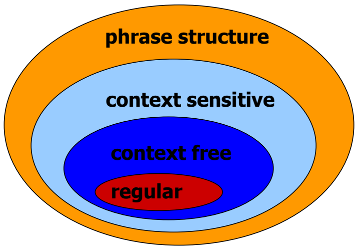
Chomsky hierarchy

Language	Grammar	Machine	
Regular language	Finite state grammar	Finite state automata	$0^m 1^n$
Context free language	Context free grammar	Push down automata	$0^n 1^n$
Decidable language	Context sensitive grammar	Linear bounded automata	$0^n 1^n 2^n$
Computable language	Phrase structure grammar	Turing machine	(M, w)

Chomsky hierarchy



Chomsky hierarchy



Human languages are not regular.

Swiss German is not context free.

Learning theory

- A text, *T*, of language, *L*, is a list of sentences, s1,s2,s3,…, which contains each sentence of *L* at least once.
- Denote by T_n the first *n* sentences of *T*.

Algorithm

- An algorithm for language learning is a mapping from text to language.
- It receives text as input and specifies a language as output

$$A(T) \to L$$

Learnability

- A language, *L*, is learnable by an algorithm, *A*, if for all texts of *L*, the algorithm will eventually specify the correct language:

$$A(T_n) \xrightarrow{n \to \infty} L$$

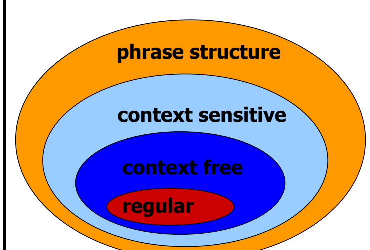
$$\forall T\, \exists m_T \text{ s.t. } \forall n > m_T \quad A(T_n) = L$$

Learnability

- A set of languages, **L**={L1,L2,…} is learnable by an algorithm, *A*, if each language is learnable by this algorithm.

Gold's theorem

- The set of all regular languages cannot be learned by any algorithm.



4 problems with Gold

- Learn exactly the target language
- Receive only positive evidence
- No informational complexity
- No computational complexity

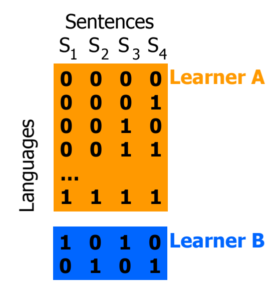
Statistical learning theory (Probably almost correct)

- The algorithm must converge with high probability to a language that is close to the correct language.
- Positive and negative evidence.
- Informational (and computational) complexity.

Statistical learning theory (Probably almost correct)

- The algorithm must converge with high probability to a language that is close to the correct language.
- Positive and negative evidence.
- Informational (and computational) complexity.
- The set of all regular languages is not learnable.
- The set of all finite languages is not learnable.

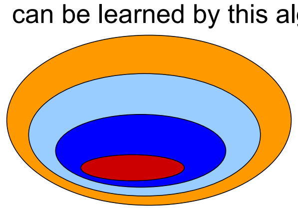
Memorization - generalization



Generalization (looking for rules) requires a sparse search space

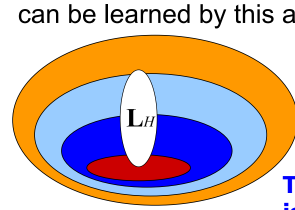
Human language learning

- The human brain contains an algorithm, A_H, that can learn language.
- The question is what is the set, $\mathbf{L}_H$, that can be learned by this algorithm?



Human language learning

- The human brain contains an algorithm, A_H, that can learn language.
- The question is what is the set, $\mathbf{L}_H$, that can be learned by this algorithm?



The theory for this set is universal grammar.

Two aspects of language evolution

time

UG

$\rightarrow L_1 \rightarrow L_2 \rightarrow L_3 \rightarrow$

Cultural evolution of language within the same UG.

Languages change:
randomly (neutral evolution)
as byproduct of other processes (cultural or military success)
adaptively (selected for acquisition and communication)

Two aspects of language evolution

time

$\rightarrow UG_1 \rightarrow UG_2 \rightarrow UG_3 \rightarrow$

Biological evolution of UG.

UG changes:
1. randomly (neutral evolution)
2. as byproduct of adaptation of other cognitive function
3. adaptively (for enhanced language acquisition and communication; selected by language)

Two aspects of language evolution

time

$\rightarrow UG_1 \rightarrow UG_2 \rightarrow UG_3 \rightarrow$

At some time in the last 5 million years, a UG arose that allowed recursion, discrete infinity, making infinite use of finite means

Evolution of universal grammar

- What criteria does UG have to fulfill
  - to induce linguistic coherence in a population
  - to allow language adaptation
  - to admit localization in language space

Grammar acquisition

Environmental input
(sample sentences)

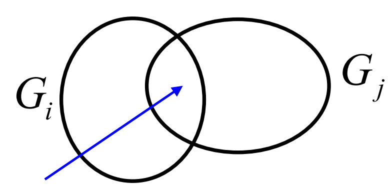
Learning procedure

Choose one of *n* candidate grammars

G1, G2, G3, ..., Gn **Search space**

Universal grammar

Compatibility of two grammars



$$a_{ij} = \mu_i(G_i \cap G_j)$$

probability that a speaker of Gi says a sentence that is compatible with Gj

Payoff for successful communication

$$F(G_i, G_j) = \frac{1}{2}(a_{ij} + a_{ji})$$

Language equation

$$\dot{x}_i = \sum_{j=1}^{n} x_j f_j(\vec{x})\ Q_{ji} - \phi(\vec{x}) x_i$$

x_i ... frequency of G_i $\quad \sum_{i=1}^{n} x_i = 1$

Language equation

$$\dot{x}_i = \sum_{j=1}^{n} x_j f_j(\vec{x})\ Q_{ji} - \phi(\vec{x}) x_i$$

Fitness of G_i : $\quad f_i(\vec{x}) = \sum_{j=1}^{n} x_j F(G_i, G_j)$

Language equation

$$\dot{x}_i = \sum_{j=1}^{n} x_j f_j(\vec{x})\ Q_{ji} - \phi(\vec{x}) x_i$$

Q_{ij}...probability that a learner will acquire G_j from a teacher with G_i.

Language equation

$$\dot{x}_i = \sum_{j=1}^{n} x_j f_j(\vec{x})\ Q_{ji} - \phi(\vec{x}) x_i$$

$\phi\ (\vec{x}) = \sum_i x_i f_i(\vec{x})$... average fitness, grammatical coherence

Quasispecies equation $\quad \dot{x}_i = \sum_{j=1}^{n} x_j f_j Q_{ji} - \phi(\vec{x}) x_i$

Language equation

$$\dot{x}_i = \sum_{j=1}^{n} x_j f_j(\vec{x})\ Q_{ji} - \phi(\vec{x}) x_i$$

↑ constant fitness

↓ perfect learning

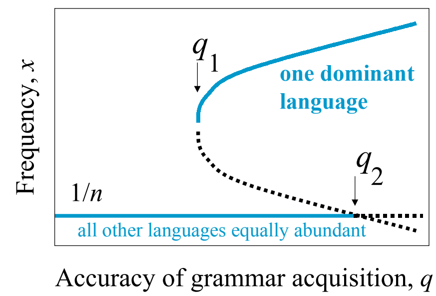
Replicator equation $\quad \dot{x}_i = x_i[f_i(\vec{x}) - \phi(\vec{x})]$

A very symmetric case

$$A = \begin{pmatrix} 1 & a & a & a & a \\ a & 1 & a & a & a \\ a & a & 1 & a & a \\ a & a & a & 1 & a \\ a & a & a & a & 1 \end{pmatrix} \qquad Q = \begin{pmatrix} q & p & p & p & p \\ p & q & p & p & p \\ p & p & q & p & p \\ p & p & p & q & p \\ p & p & p & p & q \end{pmatrix}$$

All grammars are equally good.
All grammars have the same distance from each other.
q…accuracy of grammar acquisition

Bifurcation diagram



Coherence threshold

- If $q>q1$ then the universal grammar can induce coherent communication in a population.
- $q1$ does not depend on n.

Memoryless learner

- start with a randomly chosen grammar
- stay with current grammar as long as sentences are compatible
- change to a different grammar if a sentence is not compatible
- stop after N sentences

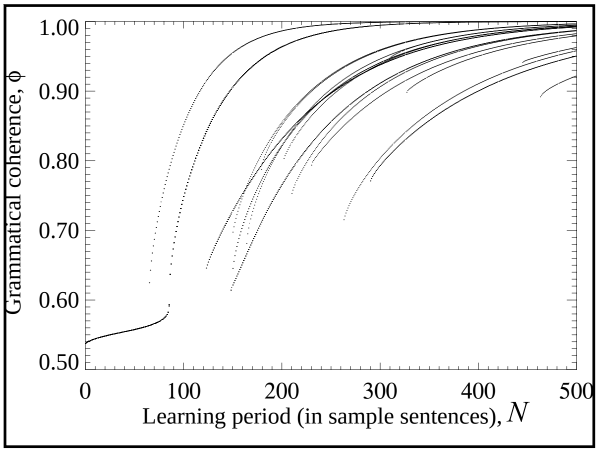
Batch learner

- memorize N sentences
- at the end, decide which grammar is most consistent with all those sentences

A more general case

a_{ij} … is a random number from $[0,1]$

$a_{ii} = 1$



Coherence threshold

- Memoryless learner

$$N > C_1 n \log n$$

- Batch learner

$$N > C_2 n$$

Coherence threshold

- Coherence threshold specifies a relationship between the duration of language acquisition and the complexity of UG.
- It determines if UG can induce coherent language in a population.
- It determines if UG allows adaptation of language for increased communicative function and learnability.

Evolution of recursion

When is generalization (=learning rule based grammars) better than memorization?

$$R > C \log n$$

relevant sentences (R) — candidate grammars (n)

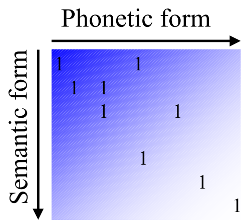
People

- Natalia Komarova (Princeton)
- Erez Lieberman (Princeton)
- Garrett Mitchener (Princeton)
- Joshua Plotkin (Princeton)
- Peter Trapa (Harvard)
- David Krakauer (Santa Fe)
- Andreas Dress (Bielefeld)
- Vincent Jansen (London)
- Partha Niyogi (Chicago)

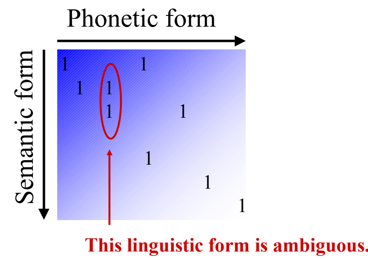
Data on language evolution

- Animal communication
- Historical linguistics
- Design of current human languages
- Empirical studies of language acquisition: describe language acquisition device and UG
- Creolization

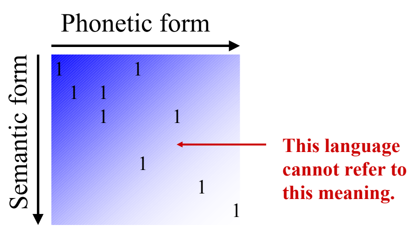
An extended model of grammar



Ambiguity



Poverty



Communicative potential

$$F(L_I, L_J) = \frac{1}{2}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sigma_i (p_{ij}^I q_{ij}^J + p_{ij}^J q_{ij}^I)$$

$$F(L_I, L_I) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sigma_i p_{ij}^I q_{ij}^I$$

$$F(L_I, L_I) = (1 - \text{poverty})(1 - \text{ambiguity})$$