# Modeling language evolution

October 16, 2002

## Linguists' view on language & evolution

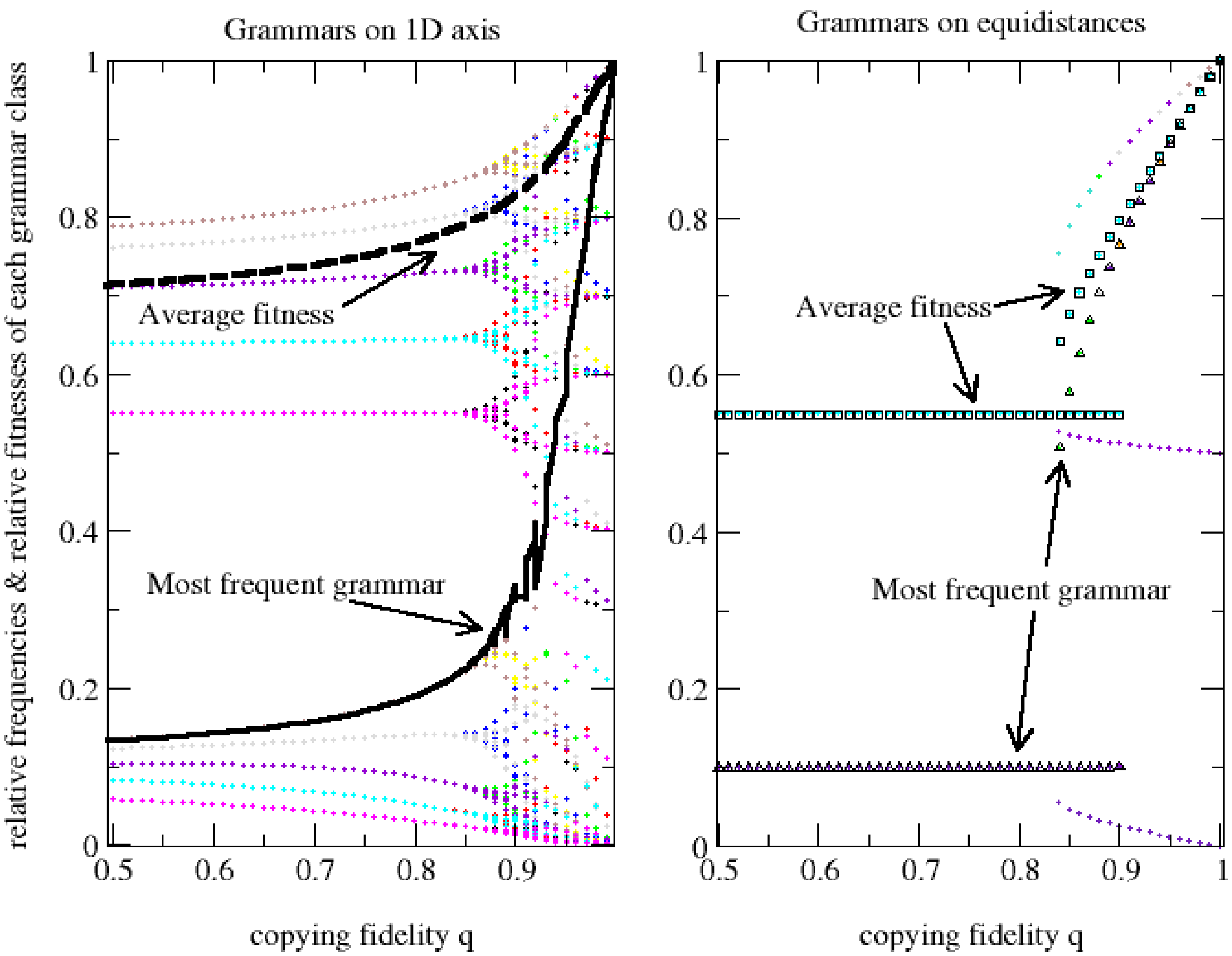
1. Human language is productive and infinite
2. It is so by virtue of compositionality and recursion
3. Therefore, it must consist of symbols and rules (i.e. rewriting grammars)
4. Non-trivial classes of rewriting grammars are not learnable from positive data alone
5. Therefore, crucial aspects of human language must be innate (Universal Grammar)
6. Human language is extremely useful
7. Evolution leads to useful things
8. Universal Grammar has evolved by Natural Selection

## Formal models of language evolution (Nowak et al, 2001)

- assume there are $N$ different possible grammars
- $x_i$ is the relative abundance of grammar $i$
- $f_i$ is the fitness of an individual using grammar $i$
- $Q_{ji}$ is the probability that a parent $j$ gets offspring $i$
- $\phi$ is the average fitness

$$\dot{x}_i = \sum_{j=0}^{N} \left( x_j f_j Q_{ji} \right) - \phi x_i \tag{1}$$

- Successful grammars are more likely to be transmitted (fitness, role model) $(f_i = \sum_j (x_j F_{ij}))$
- Every grammar is equally likely, is equally expressive and is equally similar to every other grammar $Q_{ii} = q$


Grammars on 1D axis
Grammars on equidistances
relative frequencies & relative fitnesses of each grammar class
Average fitness
Most frequent grammar
Average fitness
Most frequent grammar
copying fidelity q
copying fidelity q
0
0.2
0.4
0.6
0.8
1
0.5
0.6
0.7
0.8
0.9
1

## Evolution Universal Grammar

- There is a minimum learning accuracy $q_0$ necessary to maintain grammatical coherence in the population. Because the best possible learner (the "batch learner") needs a minimum amount $b_c$ of sample sentences to reach $q_0$, and $b_c$ is proportional to the number of possible grammars $n$, the Universal Grammar can only be of small size (Nowak et al, 2001).

- A Universal Grammar that leads to more efficient communication will be selected for

## The role of natural selection in language evolution

- If language is for transmitting information, there is no apparent advantage for the speaker to share this information
- If there is an advantage for the speaker of successful communication, there is a disadvantage of introducing innovations that are not understood.

## The role of cultural evolution in language evolution

- Language are transmitted culturally from generation to generation
- Mistakes are made in learning
- Aspects of languages that make them better learnable, are more likely to be retained
- Over time, languages will be better learnable than one would expect from a random sample

## Representation: context-free grammars

$$
\begin{array}{rcll}
S & \mapsto & NP\ VP & (1) \\
NP & \mapsto & Art\ N & (2) \\
N & \mapsto & N\ SP & (3) \\
VP & \mapsto & V\ NP & (4) \\
SP & \mapsto & \text{that } VP & (5)
\end{array}
\qquad
\begin{array}{rclcll}
Art & \mapsto & \text{the} & \mid & \text{a} & (6ab) \\
N & \mapsto & \text{cat} & \mid & \text{dog} & (7ab) \\
V & \mapsto & \text{chases} & \mid & \text{admires} & (8ab) \\
S & \mapsto & \multicolumn{3}{l}{\text{the cat fears the dog}} & (9) \\
S & \mapsto & \multicolumn{3}{l}{\text{the dog fears the cat}} & (10)
\end{array}
$$

- terminal alphabet: $\{a, b, c, d\}$
- nonterminal alphabet: $\{S, T, U, V, W\}$

# Mutations / Learning

1. Random replacement/addition of symbols, deletion/duplication of rules
2. Learning algorithm

**Incorporation:** extend the language, such that it includes the encountered string

**Compression:** substitute frequent and long substrings with a nonterminal, such that the grammar becomes smaller and the language remains unchanged

**Generalization:** equate two nonterminals, such that the grammars becomes smaller and the language larger

```
teach(i, j)
   repeat T times
      teacher i generates random string s from L_i
      student j calls incorporate(s)
   repeat until G_j doesnot change anymore
      student j calls compress()
   repeat until G_j doesnot change anymore
      student j calls generalize()
      repeat N times, or until generalization is rejected
         student j generates random string s' from L_j
         if (s' ∉ L_i) reject generalization
   if (size(L) < M)
      repeat M − size(L) times
         generate random string s of maximum size l_0
         student i calls incorporate(s)
```

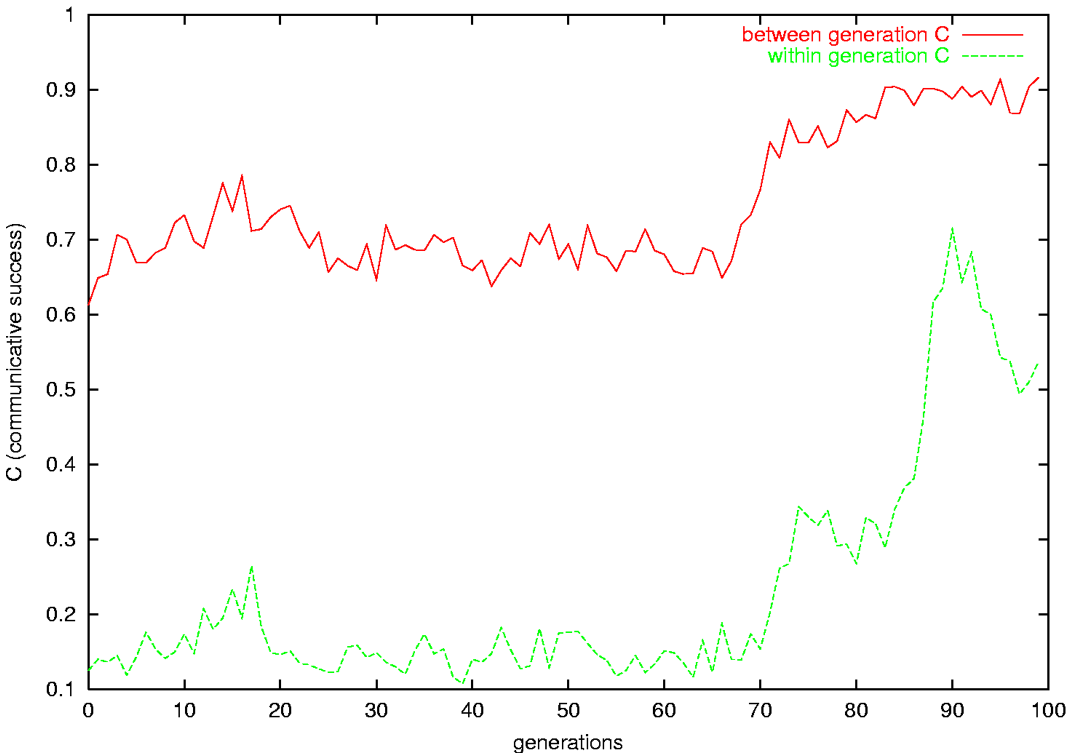
## Transmission

**Iterated learning** : agents in a *chain* learn from the previous agent and teach to the next;

**Fitness proportional selection** : the fitness of an agent is determined by its success in communicating with the agents of its own generation. The expected number of offspring is proportional to this fitness.

- *Mutual benefit:* the fitness is calculated by counting the number of times an agent is successful as a hearer and as a speaker
- *Hearer benefit:* the fitness is calculated by counting the number of times an agent is successful only as a hearer.

# Results




Parameters are: symbiotic condition, $V_t = \{0, 1, 2, 3\}$, $V_{nt} = \{S, a, b, c, d, e, f\}$, P=20, T=100, M=100, $l_0$=12

There are regions of grammar space where the dynamics are apparently under the "coherence threshold", while there are other regions where the dynamics are above this threshold. The parameters, including the number of sample sentences $T$, are still the same, but the language has adapted itself to the **bias** of the learning algorithm.

## Conclusions

- In this model, language adapts to the bias of the learning algorithm. The algorithm therefore needs less training samples than Nowak et al. predict as a lower bound.

- Results that “prove” the need for Universal Grammar (i.e. restrictions of the search space / representation bias) are based on the assumption that any target grammar from that space is equally likely. Here we show that in iterated learning that assumption is not reasonable.

- Limitations of the learning procedure make the learning in future generations easier. The collective dynamics give “emergent” restrictions; the “logical problem of language acquisition” can be solved without binary and a priori restrictions of the search space.