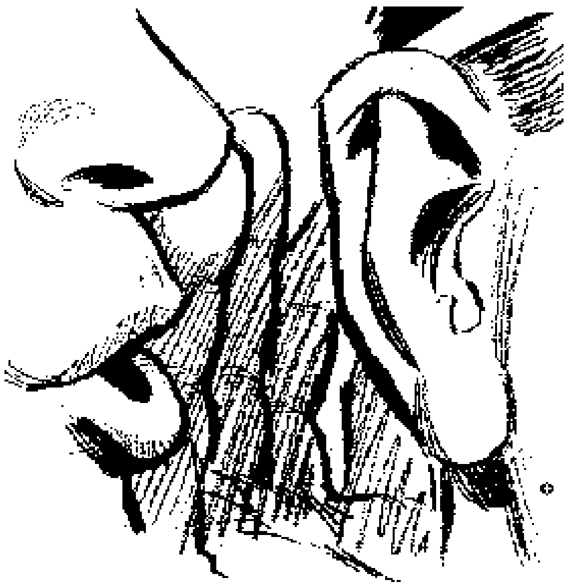
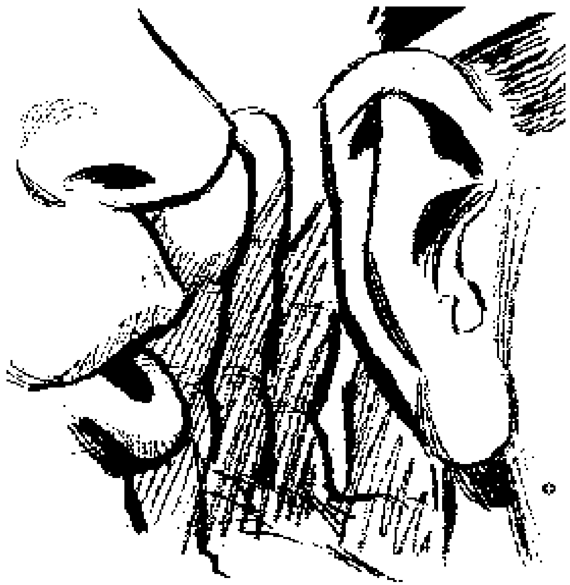
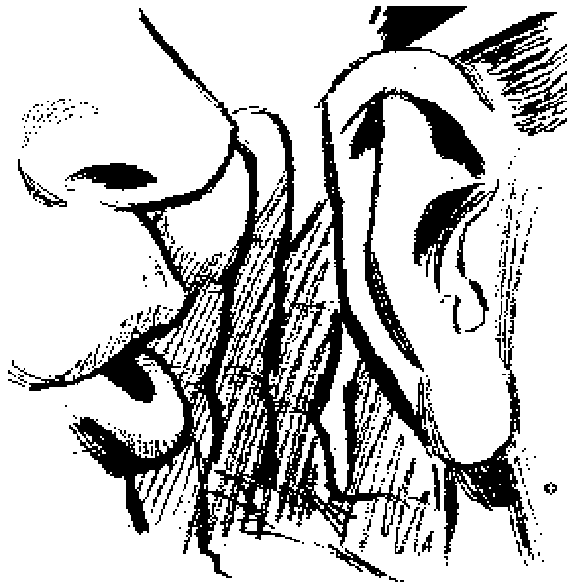
How the poverty of stimulus solves the poverty of stimulus

Willem Zuidema, VUB – AI Lab, Brussels, Belgium

The Poverty of Stimulus

"Chomsky's hypothesis is that many aspects of the formal structure of language are encoded in the genome."

"There are many arguments for the innateness hypothesis. But the most significant one in Chomsky's writings, and the one that has most affected the field, is the argument from the POVERTY OF STIMULUS."

“The basic results of the field [of learnability theory] include the formal, mathematical demonstration that without serious constraints on the nature of human grammar, no possible learning algorithm can in fact learn the class of human grammars.”

(Wexler, 1999, “Innateness of Language”, MITECS)

Formal proof I: Gold’s learnability

“the results [...] show that only the most trivial class of languages considered is learnable (in the sense of identification in the limit) from text [...].”

(Gold, 1967, “Language identification in the limit”)

"A class of language will be called identifiable in the limit *[...] if there is an effective learner [...] with the following property: Given any language of the class and given any allowable training sequence for this language, the language will be identified in the limit."*

Is this really the relevant definition of "learnability" for human languages?

Formal proof II: The coherence threshold

(Nowak, Komarova & Niyogi, 2001)

1. There is a minimum learning accuracy q_1 necessary to maintain grammatical coherence in the population.
2. The best possible learner (the “batch learner”) needs a minimum amount b_c of sample sentences to reach q_1
3. b_c is proportional to the number of possible grammars n

. . . . The Universal Grammar can only be of small size.

“If all the b sentences happen to be consistent with more than one grammar (say, with r grammars), then the [batch] learner can pick any of the r grammars with probability 1/r.”

“We assume that the learning mechanism employed by humans lies somewhere between these two extremes [i.e. the batch and memory-less learners].”

Is the batch learner really the best possible learner?

Language adaptation

1. Languages are transmitted culturally and are subject to change
2. Languages will change more if they are difficult to learn
3. Over time, languages that are easy to learn are more likely to occur

(Cavalli-Sforza & Feldman, 1980; Deacon, 1996; Kirby 2000)

The Uniformity Fallacy

Assuming every *possible* language is equally likely is NOT the proper assumption for natural languages.

- The average learnability of a class can be very good, if easy languages occur much more often than difficult languages
- Biased learning algorithms can be more successful than the batch learner

. . . . .The challenge is to show this in computational model

Search space (Grammar Universe)

Context-free grammars of the form: $A \mapsto t$, $A \mapsto BC$ or $A \mapsto Bt$. Convenient, but no restrictions on expressiveness.

ç

Parsing: depth-first search (Norvig's optimization), with maximum depth d

Derivation: random string from parsable language

Interaction: speaker derives random string s, and hearer checks if s is a string from its own language. If so, the interaction is a success, otherwise it is a failure.

Search procedure (Grammar Induction)

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

Compression: substitute frequent and long substrings with a nonterminal (the grammar becomes smaller and the language remains unchanged)

Generalization: equate two nonterminals if they occur frequently in the same context

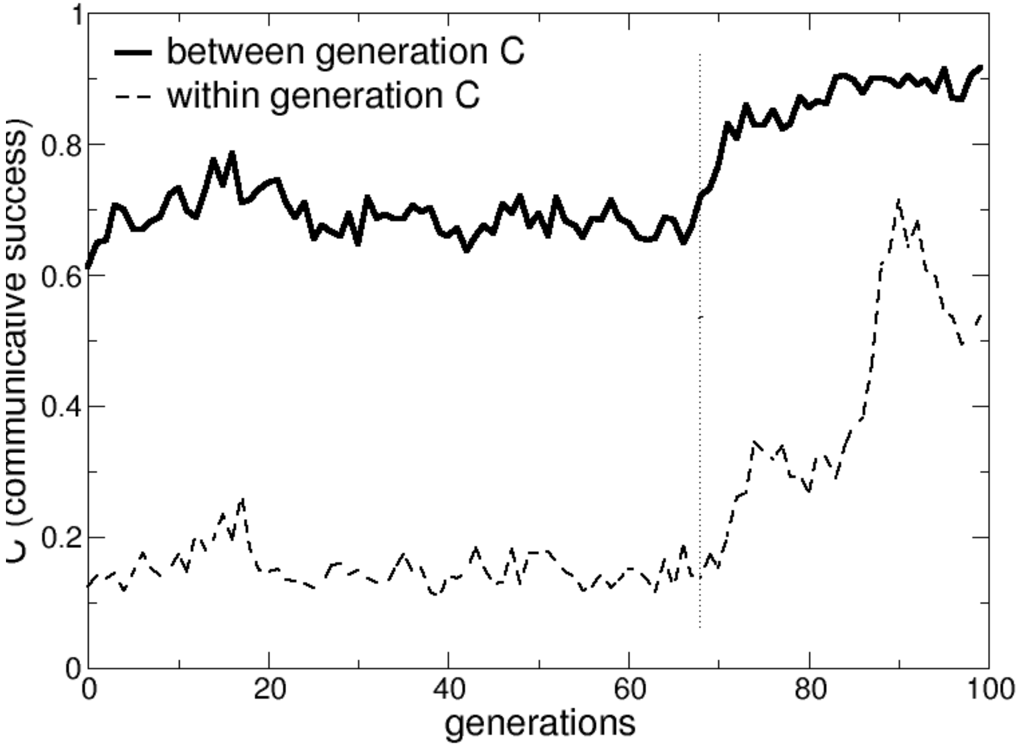
- off-line learning
- sufficient expressiveness enforced

Transmission

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

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

Results



Parameters: hearer benefit condition, $V_t = \{0, 1, 2, 3\}$, $V_{nt} = \{S, a, b, c, d, e, f\}$, P=20, T=100, M=100, d=8, 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) 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 “poverty of stimulus” does not make binary and a priori restrictions of the search space necessary. The child’s problem of a poverty of stimulus, can be solved the poverty of stimulus of previous generations.