Language adaptation helps language acquisition

Willem Zuidema
AI Lab – Vrije Universiteit Brussel
Pleinlaan 2, B-1050, BRUSSELS, Belgium
jelle@arti.vub.ac.be

Abstract

Language acquisition is a very particular type of learning problem: it is a problem where the target of the learning process is itself the outcome of a learning process. Language can therefore adapt to the learning algorithm. I present a model that shows that due to this effect – and contrary to some claims from the Universal Grammar tradition – "unlearnable" grammars can be successfully acquired, and grammatical coherence in a population can be maintained.

1 Introduction

Human language is one of the most intriguing adaptive behaviors that has emerged in evolution. Language makes it possible to express an unbounded number of different messages, and it serves as the vehicle for transmitting knowledge that is acquired over many generations. Not surprisingly, the origins of language are a central issue in both evolutionary biology and the cognitive sciences.

The dominant explanation for the origins and nature of human language postulates a "Universal Grammar": an innate system of principles and parameters, that is universal, genetically specified and independent from other cognitive abilities. In this paper, I study an argument that lies at the heart of this dominant position: the argument from the poverty of stimulus. This argument states that children have insufficient evidence to learn the language of their parents without innate knowledge about which languages are possible and which are not. This claim is backed-up with a series of mathematical models. Here, we will focus our discussion on two such models: Gold (1967) and Nowak et al. (2001).

Gold (1967) introduced the criterion "identification in the limit" for evaluating the success of a learning algorithm: with an infinite number of training samples all hypotheses of the algorithm should be identical, and equivalent to the target. Gold showed that context-free grammars are in general not learnable by this criterion from positive samples alone. This proof is based on the fact that if one has a grammar G that is consistent with all the training data, one can always construct a grammar G' that is slightly more general: i.e. the language of G, $L(G)$ is a subset of $L(G')$.

Nowak et al. (2001) provide a novel variant of the argument from the poverty of stimulus, that is based on a mathematical model of the evolution of grammars. The first step of their argument is a "coherence threshold". This threshold is the minimum learning accuracy of an individual that is consistent with grammatical coherence in a population, i.e. with a majority of individuals to use the same grammar. The second step relates this coherence threshold to a lower bound (b_0) on the number of sample sentences that a child needs. They derive that b_0 is proportional to the total number of possible grammars N. From this and the fact that the number of sample sentences is finite, Nowak et al. conclude that only if N is relatively small can a stable grammar emerge in a population. I.e. the population dynamics require a restrictive Universal Grammar.

2 Model design

These models have in common that they implicitly assume that every possible grammar is equally likely to become the target grammar for learning. If even the best possible learning algorithm cannot learn such a grammar, the set of allowed grammars must be restricted. There is, however, reason to believe that this assumption is not the most useful for language learning. Language learning is a very particular type of learning problem, because the outcome of the learning process at one generation is the input for the next.

The model study I present here is motivated by this observation. The model consists of an evolving population of language learners, that learn a grammar from their parents and get offspring proportional to the success in communicating with other individuals in their generation. The grammar induction procedure is fixed; it is inspired by Kirby (2000). The details of the grammatical formalism (context-free grammars) and the population structure are deliberately close to Gold (1967) and Nowak et al. (2001) respectively.

I use context-free grammars to represent the linguistic abilities. In particular, the representation is limited to grammars G where all rules are of one of the following forms: $A \mapsto t$, $A \mapsto BC$, or $A \mapsto Bt$. Since every context-free grammar can be transformed to such a grammar, the restrictions on the rule-types above do not limit the scope of languages that can be represented. They are, however, relevant for the language acquisition algorithm that will be discussed below. Note that the class of languages that the formalism can represent is unlearnable by Gold's criterion.

The language acquisition algorithm used in the model consists of three operations: (i) incorporation (extend the language, such that it includes the encountered string), (ii) compression (substitute frequent and long substrings with a nonterminal, such that the grammar becomes smaller and the language remains unchanged), (iii) generalization (equate two nonterminals, such that the grammar becomes smaller and the language larger).

3 Results

The main result is in figure 1, which shows two curves: (i) the average communicative success of agents speaking with their parents which is the measure for the *learnability* of the language (labeled "between generation C"), and (ii) the average communicative success of agents speaking with other agents of the same generation (labeled "within generation C") which gives the fitness of agents and is a measure for the grammatical variation in the population.

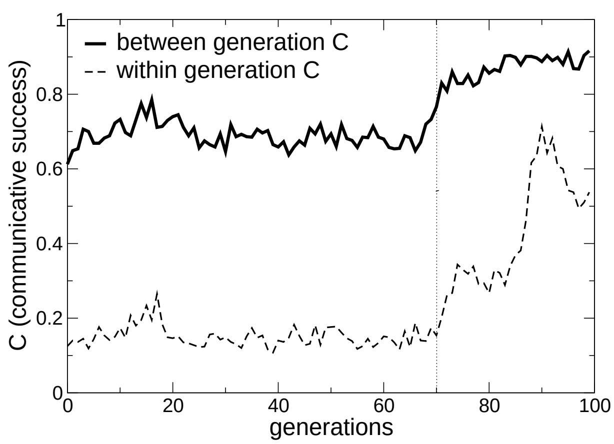

Figure 1: *Parameters are:* $V_t = \{0, 1, 2, 3\}$, $V_{nt} = \{S, a, b, c, d, e, f\}$, *P=20, T=100, M=100,* l_0*=12*

For a long period the learning is not very successful. The between generation C is low (grammars are unlearnable), and consequently the within generation C is also low (the dynamics are below the "coherence threshold" of Nowak et al. 2001). In other words, individuals are so bad at learning that members of the population can not understand each other. Around generation 70 this situation suddenly changes. The between generation C rises, and very quickly also the within generation C rises to non-trivial levels. With always the same number of sample sentences, and with always the same grammar space, there are regions of that space where the dynamics are apparently under the coherence threshold, while there are other regions where the dynamics are above this threshold. The language has adapted to the learning algorithm, and, consequently, the coherence does not satisfy the prediction of Nowak et al. In many runs (not shown here) I have also observed 100% learning accuracy of children. The grammars in this situations are thus learnable by Gold's criterion. In some, but not all cases, these emergent grammars are recursive.

4 Discussion

I believe that these results, simple and preliminary as they may be, have some important consequences for our thinking about language acquisition. In studies like the mathematical models of Gold and Nowak et al., one derives from the properties of the learning procedure (the search procedure), fundamental constraints on the nature of the target grammar (the search space). My results, like those of Kirby (2000) and others, indicate that in *iterated learning* it is not necessary to put the (whole) explanatory burden on constraints on the search space. In my model, the target grammars are learnable, not because the used formalism imposes restrictions on the grammars, but because the targets dynamically change and – in the iteration of learners learning from learners – adapt to the used learning algorithm. In other words, neither the search space nor the search procedure directly determine which grammars "exist"; the set of target grammars at the end of the simulation is the emergent result of iterating a search process over and over again.

Isn't this Universal Grammar in disguise? Learnability is – consistent with the undisputed proof of Gold (1967) – still achieved by constraining the set of targets. However, unlike in usual *interpretations* of this proof, these constraints are not strict (some grammars are better learnable than others, allowing for an infinite "Grammar Universe"), and they are not a-priori: they are the outcome of iterated learning. The poverty of stimulus is here no longer a problem; instead, the ancestors' poverty is the solution for the child's.

References

Gold, E. M. (1967). Language identification in the limit. *Information and Control (now Information and Computation)*, 10:447–474.

Kirby, S. (2000). Syntax without natural selection. In Knight, C., Hurford, J., and Studdert-Kennedy, M., (Eds.), *The Evolutionary Emergence of Language*. Cambridge University Press.

Nowak, M. A., Komarova, N., and Niyogi, P. (2001). Evolution of universal grammar. *Science*, 291:114–118.