Emergent syntax: the unremitting value of computational modeling for understanding the origins of complex language

Willem H. Zuidema

Artificial Intelligence Laboratory
Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussels, Belgium
jelle@arti.vub.ac.be
http://arti.vub.ac.be/~jelle

Abstract. In this paper we explore the similarities between a mathematical model of language evolution and several A-life simulations. We argue that the mathematical model makes some problematic simplifications, but that a combination with computational models can help to adapt and extend existing language evolution scenario's.

1 Introduction

The debate on the origins of language has been dominated by "verbal" theories, both in scientific publications (see e.g. [3]) and in popular, best-selling books (e.g. [1]). Recently also mathematical models of the evolution of language, especially those of Martin Nowak et al., have received much attention (e.g. [6]). These models are sometimes seen as a validation of the earlier verbal theories. Steven Pinker, e.g., writes in the accompanying news story of [7] that the paper shows *"the evolvability of [one of] the most striking features of language"*, i.e. its compositionality.

Although we appreciate the major contributions in these books and papers, we still observe many shortcomings in the proposed theories. Both the verbal and the mathematical accounts tend to overlook many crucial details. Verbal theories often underestimate the intricacies of the evolutionary dynamics and take "evolution" too much as a general problem solver. The mathematical models often make crucial simplifications that are linguistically poorly motivated. In particular, both types of theories have shown little appreciation for the importance of the "frequency dependency" of language evolution and the role of selforganization there-in.

A-life models, on the contrary, have shed light on both the dynamics of language evolution and the explanatory role of selforganization. However, A-life models are too often studied as relatively isolated cases, and too seldomly systematically compared with each other and with mathematical models (the review papers [8, 4] are exceptions, although they unfortunately do not discuss mathematical models). In this paper we explore the similarities between a recently published mathematical model [6], our own A-life simulations [9] and the model of Kirby [5]. We believe that such an approach can eventually both avoid the problematic simplifications of mathematical models, and the *ad hoc-ness* of many A-life models. In the conference presentation we will also discuss some shortcomings of "verbal" theories as revealed by A-life models.

2 The mathematical model

Nowak et al. use in [6] an elegant formalism that is in line with our view that one should study both the cultural dynamics of language and the evolutionary dynamics that operate on the parameters of the cultural process. We will discuss here only the model for cultural dynamics.

Nowak et al. assume that there is a finite number of states (grammar types) that an individual can be in. Further, they assume that newcomers (infants) learn their grammar from the population, where more successful grammars have a higher probability to be learned and mistakes are made in learning. The system can now be described in terms of the changes in the relative frequencies x_i of each grammar type i in the population:

$$\dot{x_i} = \sum_{j=0}^{N} x_j f_j Q_{ji} - \phi x_i \qquad (1)$$

In this differential equation, f_i is the *relative fitness* (quality) of grammars of type i and equals $f_i = \sum_j x_j F_{ij}$, where F_{ij} is the expected communicative success from an interaction between an individual of type i and an individual of type j. The relative fitness f of a grammar thus depends on the frequencies of all grammar types, hence it is *frequency dependent*. The proper way to choose F depends on the characteristics of *language use* (production and interpretation).

Q_{ij} is the probability that a child learning from a parent of type i, will end up with grammar of type j. The probability that the child ends up with the same grammar, Q_{ii}, is defined as q, the copying fidelity. The proper way to choose Q depends on the characteristics of *language acquisition* (learning and development). (ϕ is the average fitness in the population and equals $\phi = \sum_i x_i f_i$. This term is needed to keep the sum of all fractions at 1).

The main result that Nowak et al. obtain is a "coherence threshold": they show mathematically that there is a minimum value for q to keep coherence in the population. If q is lower than this value, all possible grammar types are equally frequent in the population and the communicative success in minimal. If q is higher than this value, one grammar type is dominant; the communicative success is much higher than before and reaches 100% if $q = 1$. Further, Nowak et al. derive an upper and a lower bound on the number of sample sentences that a child needs to acquire its parents' language with the required fidelity q.

3 A-life models

We argue that computational models that we [9] and others [2] have studied fit the general format of equation 1 well, but differ significantly in the particular choices for the representation of language use and language acquisition, i.e. the functions F and Q. In the limited space that is available here we will only shortly mention two examples of interesting, qualitative differences that these choices bring.

First, for sake of simplicity Nowak et al. assume that all grammars are *equally expressive*, and are all *equally similar* to each other. This has the unrealistic consequence that the benefits of interacting with another individual (F) are either maximal or minimal. We studied a computational model [9] were we used context-free grammars to represent the linguistic abilities of agents. This formalism can represent "languages" of many different types and levels of expressiveness. In that study, we did not model learning explicitly, but in stead assumed (as in equation 1) that children end up with a slightly different grammar than their parents.

One of the surprising findings was that once a certain type of language was established in the population, the language kept changing but remained of the same type. The language types formed "self-enforcing regimes", because the language present at time t determines which agents will be successful and reproduce to the next generation, and therefore indirectly determine the language at time $t + 1$. We found three such regimes: (i) idiosyncratic, non-syntactic languages, (ii) compositional languages and (iii) recursive languages. In a population where a rich but idiosyncratic language is established, syntax could not emerge. This phenomenon is important for understanding the consequences of the frequency dependency of language evolution, but is excluded in the simplifications of the mathematical model.

Second, Nowak et al. consider two extreme possibilities for the learning algorithm, and claim to have found a lower and a upper bound on the number of training samples that a learning algorithm needs to reach the coherence threshold. However, in their analysis they have not taken into account that the choice of the grammar that a child has to learn is biased by how well previous generations have been able to learn and maintain it.

In a follow-up of the study above, we have implemented a variant of the "iterated learning model" of Kirby [5], in which agents are endowed with a language-acquisition algorithm to learn the context-free grammars. Kirby found that in the process of iterated cultural transmission the language adapts itself to be better learnable by individual agents. Concretely, this means that the language becomes compositional (syntactic) and that agents are more successful in learning it than would be expected a priori. We replicated this finding, and can show that agents in fact need less training samples than Nowak et al. calculate as a lower bound for maintaining a stable language in the population. The reason is that not only do individuals evolve to be better at language-learning, but also do languages evolve to be better learned [1]. Again, this phenomenon is important for our understanding of the origins of language, but excluded in the simplifications of the mathematical model.

4 Conclusions

Research on the evolution of language faces two aspects of language that are particularly important: (i) it is transmitted, at least in part, culturally, and learned by one individual from the other; (ii) it is a group phenomenon, that occurs only between individuals and has no apparent value for an individual in isolation. These aspects make that the fitness of individual is not a function of its language acquisition system alone, but is dependent on the cultural dynamics and the composition of the group it is in as well. This observation brings *restrictions* and *opportunities* for language evolution scenario's that are deemed to be overlooked in both verbal and mathematical theorizing. We conclude that A-life models can help to evaluate the validity of these scenarios and help to adapt them, while at the same time mathematical models can help to compare computational models and to identify common themes between them.

Acknowledgements Part of the work reported here has been done in close collaboration with Paulien Hogeweg. I thank her and other members of the Theoretical Biology group in Utrecht, the Netherlands, and members of the AI-Laboratory in Brussels, Belgium, for many helpful discussions.

References

1. Terrence Deacon. *Symbolic species, the co-evolution of language and the human brain.* The Penguin Press, 1997.
2. Takashi Hashimoto and Takashi Ikegami. The emergence of a net-grammar in communicating agents. *BioSystems*, 38:1–14, 1996.
3. J. Hurford, M. Studdert-Kennedy, and C. Knight, editors. *Approaches to the evolution of language.* Cambridge University Press, 1998.
4. James R. Hurford. Expression / induction models of language. In Ted Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models.* Cambridge University Press, 2000.
5. Simon Kirby. Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, J. Hurford, and M. Studdert-Kennedy, editors, *The Evolutionary Emergence of Language: Social function and the origins of linguistic form.* Cambridge University Press, 2000.
6. Martin A. Nowak, Natalia Komarova, and Partha Niyogi. Evolution of universal grammar. *Science*, 291:114–118, 2001.
7. Martin A. Nowak, Joshua B. Plotkin, and Vincent A.A. Jansen. The evolution of syntactic communication. *Nature*, 404:495–498, 2000.
8. Luc Steels. The synthetic modeling of language origins. *Evolution of Communication*, 1:1–35, 1997.
9. Willem H. Zuidema and Paulien Hogeweg. Selective advantages of syntactic language: a model study. In *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*, pages 577–582. Lawrence Erlbaum Associates, 2000.