Social patterns guide evolving grammars

W.H. Zuidema and P. Hogeweg
Theoretical Biology / Bioinformatics*
Utrecht University

Abstract

We study the interaction between between evolutionary dynamics and group dynamics in a computational model of the evolution of syntax. We find that social patterns can guide the evolution in unexpected directions, and can both facilitate and hinder the development of compositionality and recursion. We believe these results in some sense challenge the traditional picture of the transition towards syntactical language.

1 Introduction

The transition from short, finite communication systems found in many animal species, to the open ended language system of humans, is considered to be one of the major transitions in evolution (Maynard-Smith & Szathmáry, 1995). There is large agreement that the main qualitative difference is the syntax of human language: the compositional and recursive nature allows for a systematic production and interpretation of a tremendous amount of different messages. Syntax therefore reconciles the need for a large expressiveness with the limitations in learning and memory. This is, according to speculations about the origin of human language, what makes syntax *selectively advantageous*, and caused the transition from an extensive non-syntactical "protolanguage" to a more efficient, syntactical language system (Pinker & Bloom, 1990, Nowak & Krakauer, 1999).

However, empirical evidence on e.g. animal communication, innateness and language universals, remains controversial and inconclusive. An intriguing alternative approach has emerged: mathematical and computational modeling of language origins (Hurford, 1989, Steels, 1997, Hashimoto & Ikegami, 1996, Nowak & Krakauer, 1999). In this line of research an effort is made to understand the dynamics of language evolution by studying simple models ("minimal models") of communicating agents. These models help to generate new hypotheses, to evaluate how generic certain properties are, to tackle the supposed self-evidence of arguments and to find a minimal set of assumptions sufficient to explain a phenomenon. Their main contribution so far is, that they have shown the plausibility of *cultural evolution* as a mechanism in the development of more complex languages (Steels, 1997, De Jong, 1998, De Boer & Vogt, 1999, Batali, 1997, Kirby, 1999a,b).

Fewer studies exist that model genetic transmission of language capabilities. Following Hashimoto & Ikegami (1996), the model reported in this paper studies the dynamics of genetic transmission of language. It takes an extreme position, as it ignores learning mechanisms and semantics, and models genetic adaptation of *particular* grammars. Language capabilities are described with "context free grammars", that make compositional and recursive structures very easy to obtain. However, unlike some other studies of genetic transmission (e.g. Batali, 1994, Nowak & Krakauer, 1999), no static fitness function is defined; the grammars of all individuals in a group determine the environment in which an agent must survive.

Under these simplified conditions, the interaction between evolutionary dynamics and group dynamics is studied. We will show that even without learning and cultural transmission, social patterns can influence the evolutionary dynamics. We observe that a social embedding can yield powerful, recursive grammars, but it can also prevent a population from obtaining them. Interestingly, because of these group effects, rules in one agent's grammar can influence the persistence of rules in other grammars, even though the mechanism of cultural evolution is excluded. We will show, that the results in some sense challenge the traditional picture of the transition towards syntactical language.

2 Model description

The model consists of a small set of agents that play a language game. They communicate in a lan-

*Padualaan 8, 3584 CH Utrecht, the Netherlands, jelle.zuidema@phil.uu.nl; information about most recent work can be found at http://www-binf.bio.uu.nl/~jelle

guage of short sequences ($max_l = 6$) of 0's and 1's. Agents speak ("derive") and understand ("parse") these strings using a Chomskyan rewriting grammar, which they inherited – with some random mutations – from their parent. In each language game, all agents can speak once and try to understand each of the spoken strings. Agents receive scores depending on their success in speaking, understanding and (not) being understood. After a number of language games, scores are evaluated and offspring is produced. Successful agents have a higher chance of survival and reproduction.

The grammars of the agents are context free grammars, with a small terminal ($V_{te} = \{0, 1\}$) and non-terminal alphabet ($V_{nt} = \{S, A, B\}$). As an extra restriction, the start symbol is not allowed on the right-hand side of rules. At the start of most simulations, grammars are randomly initialized with either $S \mapsto 1$ or $S \mapsto 0$.

Derivation always starts with the start symbol, and applies iteratively random fitting rules for some maximum number of steps ($max_d = 60$; failure), until no fitting rule exists (failure), or until a string of only terminal symbols is reached (success). In *parsing* rules are tried in the order they are stored, and fitting rules are applied recursively until the maximum number of steps ($max_p = 500$) is reached (failure), no other fitting rules exist to any intermediate string (failure), or the start symbol is reached (success).

The model is similar to the model introduced by Hashimoto & Ikegami (1996). They use a very specific scoring scheme, where *not* being recognized is rewarded and scores depend on population size, the length of a string, and a string's novelty in the population. Section 3.1 shows new results obtained with similar parameters. Section 3.3 shows results obtained with much simpler scoring schemes.

Hashimoto & Ikegami discuss their results in terms of the Chomsky hierarchy of grammars and languages. In a domain of changing grammars and finite languages, we believe it is much more convenient to use a classification in terms of "routes". A *route* is a sequence of rewriting steps that connects the start symbol S to a string of terminal symbols. Routes can be categorized as *indexical* (directly from S to a terminal string), *compositional* (via non-terminal symbols from S to a terminal string) or *recursive* (leading from a non-terminal symbol via one or more rewriting steps to the same non-terminal symbol). The number of routes, can be divided in three components R_I, R_C, R_R, that depend on each of these categories of routes. Similarly, expressiveness (the number of distinct strings a grammar can parse) can be divided in E_I, E_C, E_R. routes. Grammars can be characterized by these values, and classified according to the largest component (Zuidema, 1999).

3 Results

3.1 three types of behavior

To evaluate some general properties of the model, we studied the behavior with the parameter settings of Hashimoto & Ikegami (1996), and a number of variations. Similar to their results, we find that evolution can quickly lead to grammars that can parse a large fraction of the 126 possible strings. However, under slightly different parameter settings we also find quite different results. We observe three types of behavior:

i The most frequent behavior is a quick growth of expressiveness, from 1 at initialization, to over 100 after about a 1000 generations. In the first stage the expressiveness depends only on indexical routes. Soon, however, compositional routes and recursive routes become more important. Eventually, recursive routes dominate the grammar's expressiveness.

ii Sometimes, it takes much longer to reach the high level of expressiveness, ranging from 2000 to many thousands of generations. In these type of runs, compositional routes quickly become important, but recursive routes are infrequent.

iii Least frequent are runs that show very little growth in expressiveness. After 3000 generations, only around 20 words can be parsed. In these runs, expressiveness depends almost exclusively on indexical routes.

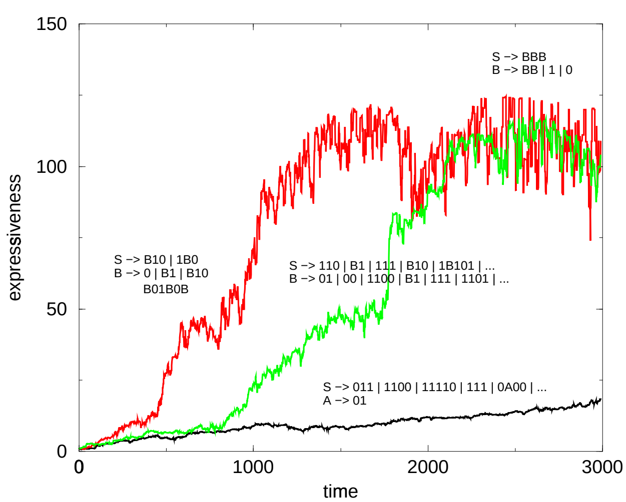

Figure 1: *Three runs, typical for the indexical, compositional and recursive regimes, and some example grammars.*

These types of behavior also differ in their robustness against mutations and generalization abilities. With some particular parameter settings, each of the three types of behavior can occur, solely depending on the “seed” for the random generator. At different generations we restarted runs with original grammars but a different random seed. In early generations, a change of type of behavior occurs frequently. However, in restarts from later generations, the type of behavior seems fixed and a change of type becomes increasingly improbable. The types of behavior thus form *self-enforcing, dynamical regimes.*

3.2 context and group effects

These results crucially depend on the fact that the fitness of an agent is evaluated with respect to its performance in the group, rather than with respect to some static fitness function. In a random population, agents with more expressive grammars speak more novel strings, understand more strings and are less likely to be understood, and thus should receive higher scores. The existence of the dynamical regimes, in a non-random population, can be explained by two mechanisms: a *context effect* (rules are generally most successful in a context of similar rules) and a *group effect* (agents are most successful in a group of similar agents).

The derive-languages of individuals, jointly constitute a group language, that in turn determines the success of agents in parsing. This indirect feedback can best be described as a *social pattern* that emerges from individual behaviors, and in turn restricts individual success. Initial similarities (in terms of our classification) are enforced by these social patterns.

Apparently, the larger an indexical grammar is, the less likely it is that evolution can lead to compositional and recursive grammars. This in some sense contradicts the traditional picture of the evolution of syntax, that states that only when indexical grammars became too large, syntax emerged.

A simple analysis can lead to some qualitative predictions on how, given the existence of these regimes, different variables in the model should relate. One can show, that the number of routes grows linearly with grammar size in the indexical regime. In a compositional regime it grows faster, and in a recursive regime extremely fast[1]. A rough estimate of how expressiveness[2] depends on R, gives a qualitative explanation for the trajectories in the phase space in figure 2. If a linear growth of grammar size over time is assumed, the shape of the curves in figure 1 can also be explained.

[1] Take for example the simple case of grammars with $V_{nt} = \{S, A\}$, and at most one non-terminal and at least one terminal symbol at all right-hand sides of rules. Estimates of R in each of the regimes are: $R_I \approx N$, $R_C \approx \left(\frac{1}{2}N\right)^2$, $R_R \approx \left(\frac{1}{3}N\right)^{max_c+2}$, where max_c is the maximum number of cycles.

[2] $E \approx E_{max}\left(1 - \left(1 - (\frac{1}{E_{max}})\right)^R\right)$, here $E_{max} = 126$

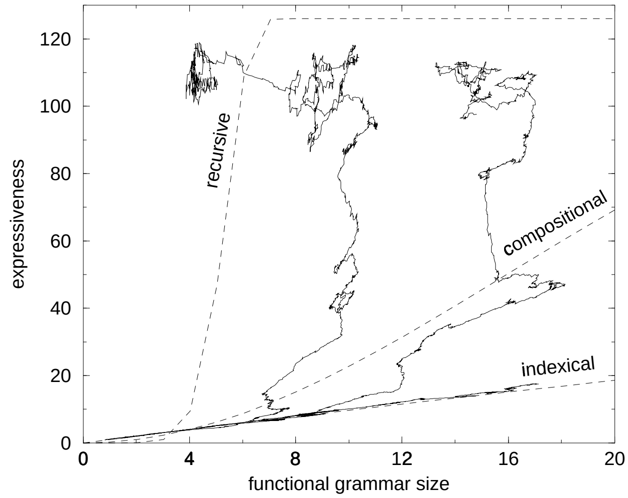

Figure 2: *Trajectories of the same three runs in a phase space of functional grammar size (the number of rules that are actually used in communication) vs. expressiveness*

3.3 selective advantages

With the scope of possible behaviors sketched, we can now turn to the question under what selection pressures the different dynamical behaviors are likely to arise. To study this, we designed several simple scoring schemes. These include: *(i)* communication, where both speaker and receiver benefit from exchanging information; and *(ii)* perception, where only the recognizing agent benefits from the information it receives.

Although recursive structures are always only a few mutations away, the development of recursive and expressive grammars is not trivial at all. With the default initial grammar (randomly $S \mapsto 1$ or $S \mapsto 0$), the communication scheme shows no increase in expressiveness, unless an explicit pressure is put on innovation. In that case sometimes recursive structures develop, but slowly and to a limited extent. If grammars are initialized with a longer indexical grammar, even this explicit innovation pressure can not force the simulation out of the indexical regime.

In contrast, the perception scheme leads to recursive grammars under all circumstances considered. However, when the population has been at a high level of expressiveness for some time, agents start to develop grammars that are just as expressive, but have a high probability of failing in derivation. The asymmetry in parsing and derivation, makes this possible.

These results yield an interesting paradox. Under the parameter settings that lead to expressive grammars, the willingness to speak is absent, while in cases where communication is mutually beneficial, no increase in expressiveness occurs.

3.4 expressiveness

Intuitively, we expect expressive grammars to be generally selectively advantageous. However, the model points at the need for a more differentiated analysis. The explicit role of expressiveness, is that *(i)* scores (can) depend on the novelty of a strings. Therefore, in general, more expressive grammars generate higher scores.

Implicitly, expressiveness also influences the scores in other ways. *(ii)* Expressive speakers are more likely *not* to be understood, while *(iii)* expressive listeners are more likely to understand. In the case of perception highly expressive grammars are beneficial, because they yield higher recognizing scores and fewer opportunities to gather scores for an agent's competitors. In the case of communication, recognition benefits from high expressiveness, but being recognized suffers from it (competition effects can be ignored).

If the net-effect of the three roles of expressiveness is positive, we expect, in a random population, recursive grammars to arise. This explains the path to recursive grammars with "perception" and, conversely, the minimal level of expressiveness in the case of "communication" without a novelty pressure.

However, as argued in section 3.2, social patterns sometimes guide the dynamics in different directions than expected from such fixed-fitness approximations. The difference in behavior between simulations with small and large initial indexical grammars can not be explained with the analysis of the roles of expressiveness alone.

4 Conclusions

This study concerns the interaction between group dynamics and evolutionary dynamics. We have seen that social patterns influence the course of evolution. Under some conditions powerful, recursive grammars develop (Hashimoto & Ikegami, 1996). Obtaining this type of grammars is in fact a hard problem; the fact that we nevertheless find recursive grammars, appears to be due to the social embedding that yields a dynamical fitness landscape. This relates to work showing the benefits of cultural transmission (Batali, 1997, Kirby, 1999a), "starting small" (Elman, 1991) and sparse fitness evaluation (Pagie & Hogeweg, 1997).

However, in other circumstances social patterns hinder the development of such grammars. These results are particularly interesting, as these specific circumstances in some sense resemble the situation that is thought to precede the emergence of syntax: large indexical grammars and mutually beneficial communication. In the model we arrive at a paradox, where those selection pressures that lead to syntactical languages, also lead to unwillingness to speak. Preliminary results indicate, that this paradox can be solved if a spatial distribution of agents and local communication is assumed.

Relaxing the idea of explicit selection pressures for syntax, an observation in section 3.2 points at an alternative mechanism for the development of recursion. The fact that recursive expressiveness (E_R) grows very fast with the number of rules (N), shows that the larger N (i.e. the "storage capacity"), the larger the expected *relative* fraction of recursive expressiveness. Whereas the traditional view emphasizes that *cognitive limitations* create the need for syntax, this observation indicates that larger *cognitive abilities* in fact make recursive expressiveness more likely to dominate. This might explain the apparent paradox, that the species with the most extended cognitive abilities, is the only species that developed "efficient", recursive communication.

References

BATALI, J. (1994). Innate biases and critical periods. In: *Artifical Life IV* (Brooks, R. & Maes, P., eds.), pp. 160–171. MIT Press.

BATALI, J. (1997). Computational simulations of the emergence of grammar. In: *Approaches to the evolution of language* (Hurford, J. & Studdert-Kennedy, M., eds.). Cambridge University Press.

DE BOER, B. & VOGT, P. (1999). Emergence of speech sounds in changing populations. In: *ECAL'99 Advances in artificial life* (Floreano, D., Nicoud, J.-D. & Mondada, F., eds.), pp. 664–673.

DE JONG, E. (1998). The development of a lexicon based on behavior. In: *Proc. of BNAIC'98* (La Poutr, H. & Van den Herik, J., eds.), pp. 27–36. Amsterdam: CWI.

ELMAN, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning* **7**, 195–225.

HASHIMOTO, T. & IKEGAMI, T. (1996). The emergence of a net-grammar in communicating agents. *BioSystems* **38**, 1–14.

HURFORD, J. (1989). Biological evolution of the saussurean sign. *Lingua* **77**, 187–222.

KIRBY, S. (1999a). Learning, bottlenecks and the evolution of recursive syntax. In: *Linguistic evolution through language acquisition* (Briscoe, T., ed.). Cambridge University Press.

KIRBY, S. (1999b). Syntax without natural selection. In: *The emergence of language* (Knight, C., Hurford, J. & Studdert-Kennedy, M., eds.). (forthcoming).

MAYNARD-SMITH, J. & SZATHMÁRY, E. (1995). *The major transitions in evolution.* Morgan-Freeman.

NOWAK, M. A. & KRAKAUER, D. C. (1999). The evolution of language. *Proc. Nat. Acad. Sci. USA* **96**, 8028–8033.

PAGIE, L. & HOGEWEG, P. (1997). Evolutionary consequences of coevolving targets. *Evolutionary computation* **5**, 401–418.

PINKER, S. & BLOOM, P. (1990). Natural language and natural selection. *Behavioral and brain sciences.*

STEELS, L. (1997). Synthesising the origins of language and meaning. In: *Approaches to the evolution of language* (Hurford, J. & Studdert-Kennedy, M., eds.).

ZUIDEMA, W. H. (1999). Describing changing grammars and languages. (in prep.).