

How grammar emerges to dampen combinatorial search in parsing

Pieter Wellens

Vrije Universiteit Brussel, Artificial Intelligence Laboratory
Pleinlaan 2, B-1050 Brussel

1 abstract

According to the functional approach to language evolution (inspired by cognitive linguistics and construction grammar), grammar arises to deal with issues in communication among autonomous agents, particularly maximisation of communicative success and expressive power and minimisation of cognitive effort. Experiments in the emergence of grammar should hence start from a simulation of communicative exchanges between embodied agents, and then show how a particular issue that arises can be solved or partially solved by introducing more grammar. In the talk I will elaborate on a case study of this approach, focusing on the issue of search during parsing. Multiple hypotheses arise in parsing when the same syntactic pattern can be used for multiple purposes or when one syntactic pattern partly overlaps with another one. It is well known that syntactic ambiguity rapidly leads to combinatorial explosions and hence an increase in memory use and processing power, possibly to a point where the sentence can no longer be handled. Additional grammar, such as syntactic or semantic subcategorisation or word order and agreement constraints can help to dampen search because it provides information to the hearer which hypotheses are the most likely. The paper shows an operational multi-agent experiment where avoiding search is used as the driver for the introduction and negotiation of syntax as can be seen in figure 1. The experiment is also a demonstration of how Fluid Construction Grammar is well suited for experiments in language evolution.

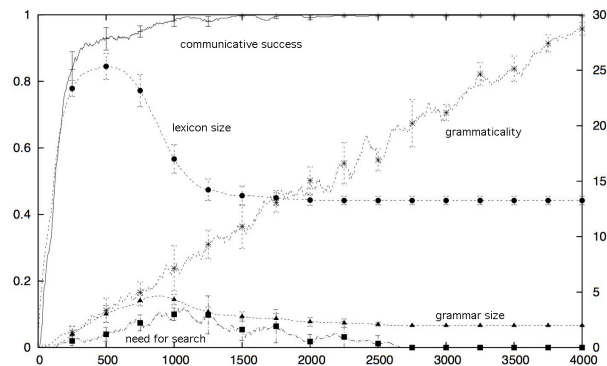


Fig. 1. Experiments where 5 agents play 4000 language games and are able to tighten grammatical constructions with additional syntax in order to avoid combinatorial search. We see that only little search is needed, even till the point (after the 2700th game) that parsing becomes deterministic in all cases. The search space consists of all grammatical constructions and no search means that the during parsing it is 100% clear for the hearer to know which construction to apply next.