Language Change in Multi-Generational Community

Peter Bodík

Institute of Informatics, Faculty of Mathematics, Physics and Informatics
Comenius University
email: Peter.Bodik@st.fmph.uniba.sk

Abstract

Steels in [4] claims that both flux of agents (changing of agents in an experiment) and stochasticity in communication of agents are necessary for a spontaneous change in language. This paper argues that flux of agents alone could be responsible for a spontaneous change in language. This hypothesis is demonstrated by modeling language use through language games played in a population of evolving agents.

1 Introduction

Human language is a complex communication system that has evolved during thousands of years among humans. *How* did language evolve is still a big puzzle that is constantly being tackled by linguists, biologists, psychologists, and in the past 10 years also by computer scientists. Apart from the sheer desire to understand the emergence of language, scientists are motivated also by possible applications. Teaching our computers to fully understand human language is an example of such an application.

The debate on the origins of language has been dominated by “informal” theories of linguists, psychologists or biologists. However, these theories often underestimate the intricacies of the evolutionary dynamics and rely too much on “evolution”. A different approach to understanding of language evolution are *mathematical models of language*, e.g. [2]. Typically, the language dynamics are described by a set of equations and the properties of the language are studied using mathematical methods. However, for the mathematical proofs to be feasible, the models have to make crucial simplifications that are linguistically poorly motivated. Yet another approach to modeling of language evolution is offered by artificial life. Artificial life has been used so far to simulate complex dynamic systems, where verbal theorizing often leads to incorrect predictions because our intuitions about the links between local interactions and global behavior are notoriously unreliable. This makes human language an ideal topic for exploration using A-life models.

One of the interesting characteristics of human languages is its spontaneous change. Human languages are constantly evolving; some words that were used 100 years ago are no longer used in the present and new words are created every day. An A-life model of spontaneous language change was proposed by Steels in [4]. His experiments show that both stochasticity in communication of agents and flux of agents (old agents leaving and new agents entering the experiment) is necessary for a change in language. We propose a simpler model where a flux of agents alone is sufficient for a change in language. Please refer to [1] for a more detailed description of the environment, the agents and other obtained results.

2 The agent

All experiments presented in this paper were conducted as simulations of a multi-agent system. The whole system consists of agents, objects, and a square playground, where the agents and objects are situated. Each experiment proceeds in turns; in every turn an agent can move and/or communicate with another agent (play a *language game*).

2.1 Lexicon

The lexicon is the most important part of an agent – it stores agent's language information. More specifically, it stores all concepts of the agent (see below) along with corresponding words. The lexicon is thus a set of "word/meaning" pairs. *Meaning* is the internal representation (the concept) and *word* is a string of characters that is used to communicate the meaning to other agents. The agent uses its lexicon a) to find a meaning of a word (word to concept), and b) to find a word to express a concept (concept to word).

Each word/meaning pair has its score – a real number expressing its success in communication. The score is modified during language games; increased after a successful communication and decreased otherwise. To express a meaning, the agents use the word with the highest score, the preferred word.

2.2 Concepts

Agents have three types of concepts: objects, distances and directions. All concepts are stored in the agent's lexicon along with the associated words. The concepts of distance and direction are innate in the agents. However, in the beginning they only have a very general notion of the *spatial* concepts. For example, the most general direction concept is "0 to 360 degrees". During an experiment the agents play *spatial games* (see below) where they describe objects using distances and directions (e.g. 20 meters to the north). If an agent can't disambiguate among several objects, it needs more specific spatial concepts. The agent can thus *divide* an existing spatial concept into two *more specific* concepts. The spatial concepts thus form a *concept tree* (see figure 1).

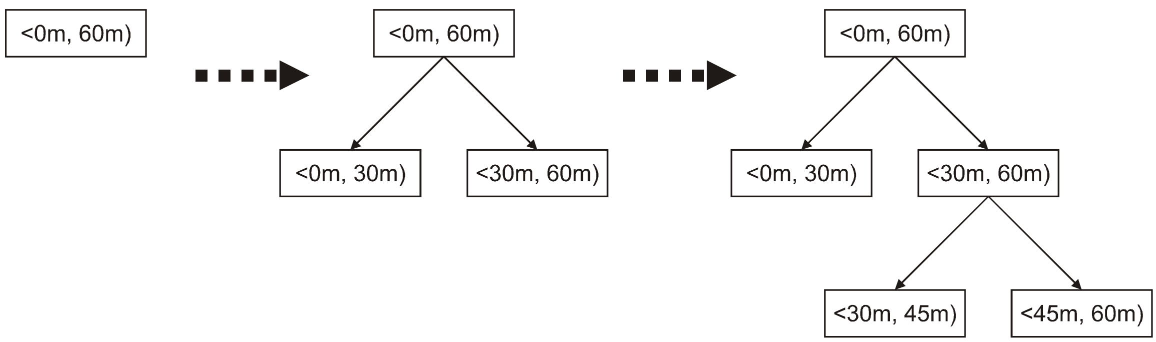

Figure 1: Conceptualization tree

2.3 Language games

During the experiments, the agents can communicate using three types of language games: pointing games, spatial games and evaluation games.

Pointing game involves two agents (A and B) and an object O in their vicinity. Agent A searches the lexicon for a word w describing O (a preferred word). If A has no word for O, the game fails and A creates a new word for O. Otherwise, A *points* to O and utters w. If agent B has word w associated to O in its lexicon, the game succeeds. Otherwise, the game fails and B learns the new word w. Finally, agent B updates the score of (w,O) in its lexicon. This game is used to develop a lexicon for objects.

Spatial game also involves two agents (A and B) and an object O, but the agents describe O by its relative position (e.g. 20 meters to the north), instead of pointing. Agent A first selects one or two spatial concepts (direction and/or distance) that unambiguously describe object O (if A has no such concepts, more specific ones are created). A then finds preferred words for the selected concepts (again, if there are no words for a concept, the game fails and the word for the concept is created). Finally, agent A utters preferred word for O and one or two words describing O's position (words w_1 and w_2). Agent B then finds object O on the playground (using its name) and

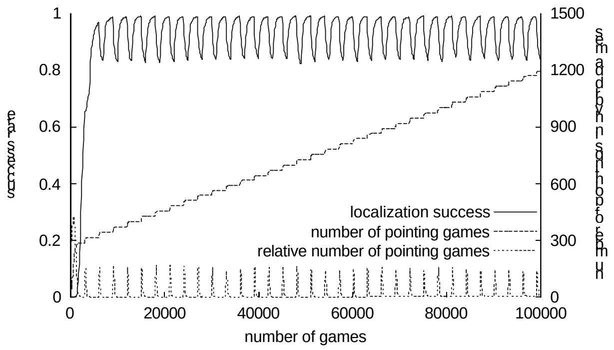

Figure 2: Experiment 1c, flux rate 3000

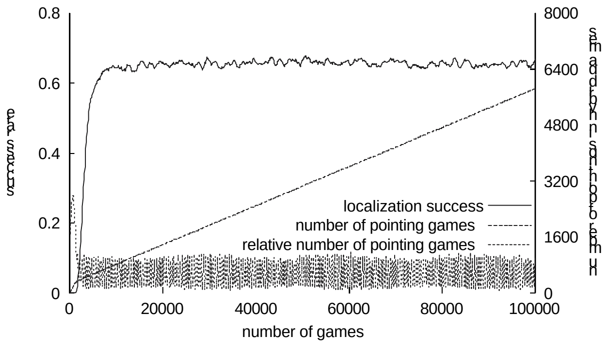

Figure 3: Experiment 1f, flux rate 500

describes its position from A's point of view (using spatial concepts c_1 and c_2). If (w_1,c_1) and (w_2,c_2) are in B's lexicon, the game is successful. Otherwise it fails and B learns the new word/meaning pairs. In the end, B updates the score of (w_1,c_1) and (w_2,c_2) in its lexicon. This game is used to develop a lexicon for spatial concepts.

Evaluation game is used only to assess the quality of the emerged language. It is very similar to spatial game, but the lexicons of agents are left unchanged. The average success rate of the last 200 evaluation games is called *localization success*.

2.4 The experiments

The experiments were conducted with 10 agents and 9 objects on the playground. 20% of the games played were evaluation games, the remaining 80% were spatial games. A pointing game was played only if a spatial game failed due to object word misunderstanding (agent B didn't understand the word for object O). We used multi-generation experiments; every *n* turns a random agent was removed from the experiment and a new one was introduced. The speed of flux of agents, *n*, is called the *flux rate* (FR).

3 Flux of agents

To understand the dynamics of an experiment with a flux of agents, we conducted a set of experiments with a varying flux rate. The flux rates were as follows: experiment 1a) 10000, 1b) 5000, 1c) 3000, 1d) 2000, 1e) 1000, 1f) 500, 1g) 200, 1h) 100, 1i) 50.

The results of experiments c and f are presented in figures 2 and 3. We can see from the graphs, that after an old agent leaves the experiment (and a new one enters), there is a drop in the localization success. This drop is caused by the fact that the new agent doesn't *share* the language of the other agents. First, it has to learn the object lexicon (notice the increase of the number of pointing games) and then the spatial lexicon.

We can divide the experiments into three groups: the *first group* contains experiments a to d (flux rate at least 2000). In these experiments, a shared language evolved and the agents were able to sustain it. The new agent had enough time (played enough games) to learn the established language from the other agents. The *second group* (e, f, g) contains experiments with flux rates between 200 and 1000. In these experiments, the localization success reached about 80%, 60% and 20% respectively and remained on this level during the whole run. The agents didn't have enough time to learn and the language couldn't develop fully. The *last group* (h and i) contains experiments with flux rate less than 200. The new agents were entering the environment very quickly and didn't have enough time to establish a shared language. The localization success was 0% during the whole experiment.

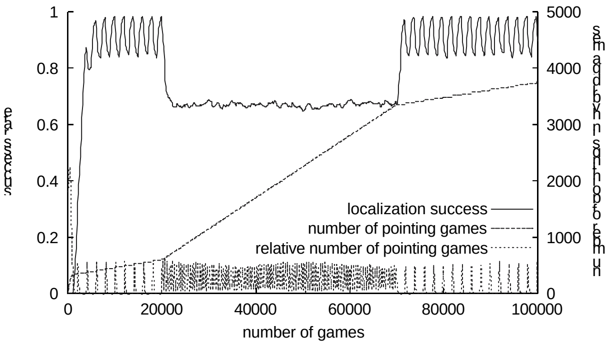

Figure 4: Experiment 2a, varying flux rate; FR 2000, 500, 2000

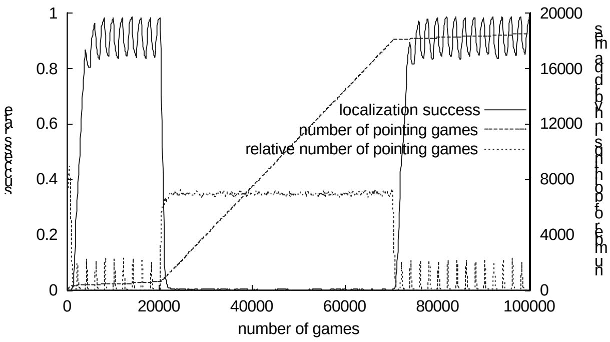

Figure 5: Experiment 2b, varying flux rate; FR 2000, 50, 2000

In the following two experiments, we look at how the language evolves when we change the flux rate during the experiment. Both experiments start with flux rate of 2000. After 20000 games the flux rate is changed to 500 (experiment 2a) or 50 (experiment 2b) and changed back to 2000 after another 50000 games (figures 4 and 5).

In the first phase of the experiment (games 0 to 20000), the flux was slow enough for the language to evolve. In the second phase (games 20001 to 70000), the flux became faster (flux rate of 500 and 50). Even though the language was fully evolved, the agents couldn't sustain it during the faster flux. During flux rate of 500, the quality of language deteriorated to about 60%. During flux rate of 50, there was no shared lexicon at all (localization success fell to 0%). During the third phase of the experiment (games 70001 to 100000), the flux rate was slowed down again to 2000. Despite the low quality of language in the previous phase, the slower flux rate allowed the agents to build the language anew. Localization success have risen again to about 99% in both experiments.

4 Spontaneous Language Change

In this section, we look at two experiments and examine how the language changes during each one of them. Experiment 3a starts with flux rate of 2000 (phase 1), after 20000 games the flux rate is changed to 50 (phase 2) and changed back to 2000 after another 160000 games (phase 3). Experiment 3b runs with flux rate 2000 for 5 million games.

4.1 Constant change at flux rate 50

We have learned from the experiments that the faster is the flux, the higher is the probability a word for concept will change. The most extreme case is experiment 3a with flux rate 50. A shared language emerged during the first phase (FR 2000), but the established words were forgotten soon after the second phase began (with FR 50). New agents were flowing in so quickly (one every 50 games) that new words were being constantly invented but soon forgotten.

We took a snapshot of the preferred words for objects of all 10 agents after 60000 games. The agents were in the middle of the second phase with FR 50 and many different words were used for every single object. Also, a lot of words (35 out of 90) didn't have *any* word assigned to them and thus the agents had many possibilities to invent new ones. With so many different words used for every object, the new ones could easily outcompete them. The original words were soon forgotten as the agents left the experiment. The competition of words for object 1 is depicted on figure 4 (the graph represents relative score of each word used between games 42700 and 54000 among all agents). During about 10000 games, 35 words were invented for object 1 and only two of them survived (both invented in game 53500). Both of these words remained in the population of agents for quite short time and were soon replaced by new ones.

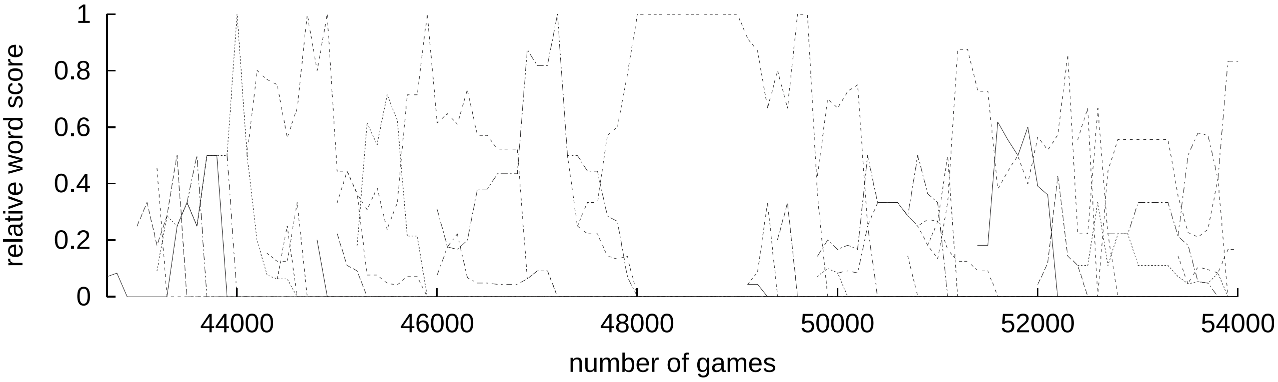

Figure 6: Experiment 3a, evolution of relative word scores, flux rate 50

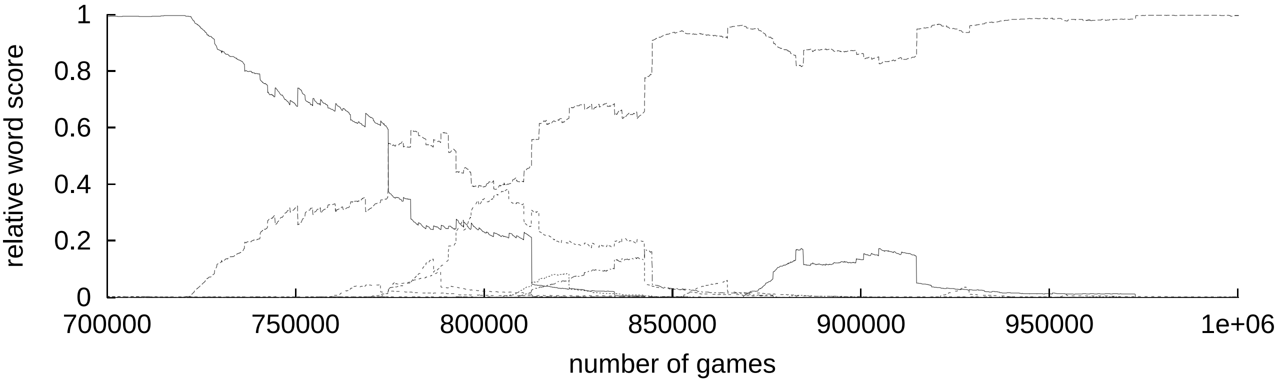

Figure 7: Experiment 3b, evolution of relative word scores, flux rate 2000

Another snapshot of the preferred words was taken after 200000 games. The flux rate was set to 2000 after 180000 games, the lexicon settled down and only one or two words were used for every object. Most of the words that survived were created between games 178000 and 181000 – close to the end of phase 2. None of the words created before game 178000 survived.

The flux rate in this experiment was extremely fast and the shared language that had evolved after 20000 games deteriorated very quickly in the second phase. After the flux rate slowed down to 2000, a new shared language emerged and localization success stayed close to 100%. This experiment is, however, not a very realistic simulation of language change in real languages, because the quality of language is very low and changes occur very often.

4.2 Gradual change at flux rate 2000

In experiments with flux rate 2000, the words changed very rarely. In an experiment with 10 objects and flux rate 2000 only 13 changes of words for objects took place during 5 million games. Competition of words during one of the changes can be seen on figure 7. Let's look at this particular change of word in detail.

Up to game 720800, there was only one word used for object 8 – word *svz0*. However, a few hundred games ago a new agent entered the experiment and invented a new word in game 720800 – word *nlf720*. The next agent that came (2000 games later) learned the new word and there were thus two agents using this new word. After 2000 games came another agent, also learned the new word, but switched back to the old one as the majority of agents still used it. An agent which came in game 726 thsd. made up a new word, *fwz726*, but changed it to *nlf720* after a while. In game 728 thsd. the new word was used by 4 agents, after 743 thsd. games by 5 agents, and after 746 thsd. by 6 agents. Then some new words appeared, *wnt752*, *shr758*, *vck768*, and *wdp772*, of which *vck768* was the strongest and was used by 5 agents after 800 thsd. games. Eventually these four words disappeared. A few thousand games later, *nlf720* regained its popularity, was used by 7 agents in game 816 thsd. and by all agents in game 845 thsd. The whole transition from *svz0* to *nlf720* took more than 120000 games. During the change, 60 new agents entered the experiment and the whole population thus changed 6 times.

We noticed that it's only the *young* agents that can learn the new word. An *old* agent which have been using an old word for a long time is unlikely to change because it's hard to outcompete the established word. Also, it sometimes happens that a new agent learns a new word, but after communicating with the *old* agents, it starts using the old word.

This experiment is a much more realistic simulation of a language change in real languages. With flux rate 2000, a shared language emerged and localization success was close to 100% during the whole experiment. The emerged language was very stable (as opposed to experiment 3a) and most of the time the new agents didn't invent any new words. However, very rarely a change of word for an object did occur and took several generations of agents to accomplish.

We also noticed that the most specific distance concepts changed faster than the more general ones. The most specific distance concepts were usually used only within a smaller group of agents (4 to 6 agents) and it was easier for the new words to spread in this smaller group.

5 Conclusion

This paper presents a new model of spontaneous language change in a population of artificial agents. The agents are situated in a 2D environment and use spatial concepts of distance and direction to describe the positions of objects around them. A shared spatial language emerges during the experiments. We used multi-generation experiments with agents *flowing* in and out of the experiment to examine the change of language. We found out that the flux of agents alone was sufficient to cause a change in language (we observed no language change in experiments with no flux of agents). Another factor influencing the change of language is the size of the group that is using the word. The words for the most specific distance concepts were used by a smaller group of agents (4 to 6 agents) and thus changed much more frequently than the other words.

References

[1] P. Bodík. Emergence of Language in Spatial Language Games. Diploma Thesis, Comenius University, April 2003. `http://pauli.fmph.uniba.sk/~8bodik/thesis.pdf`

[2] M. A. Nowak, N. L. Komarova, and P. Niyogi. Evolution of universal grammar. *Science*, 291:114–118, 2001.

[3] L. Steels and F. Kaplan. Stochasticity as a source of innovation in language games. In C. Adami, R. Belew, H. Kitano, and C. Taylor, editors, *Artificial Life VI*, Los Angeles, 1998. MIT Press.

[4] L. Steels and F. Kaplan. Spontaneous lexicon change. In *COLING-ACL98*, pages 1243–1249, Montreal, 1998. ACL.

[5] L. Steels, F. Kaplan, A. McIntyre, and J. Van Looveren. Crucial factors in the origins of word-meaning. In Alison Wray, editor, *The Transition to Language*, chapter 12. Oxford University Press, Oxford, 2002.