Self-organising vocabularies

Luc Steels
Artificial Intelligence Laboratory
Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussels, Belgium
E-mail: steels@arti.vub.ac.be

November 14, 1995

1 Introduction

Linguistics has traditionally focused on developing a description of all the possible sentences in a particular language (grammar) and on finding constraints that the grammars of all natural languages satisfy (universal grammar). However, a true theory of language should strive for more: It should explain why language has originated in the first place, how language can be learned, why language keeps evolving, and why there are so many diverse languages. The 'artificial life' approach has already shed some light on these issues. Several researchers have carried out experiments to investigate the origin of communication [5], the origin of vocabulary [6], and the growth in complexity of syntax [3]. These researchers assume that genetic evolution is the main driving force towards new structure, coherence, and more complexity. But humans learn the languages present in their environment during their life time. There is no evidence that a *particular* natural language is innate (although the thesis has been advanced that there is an innate language acquisition device [1]). Moreover languages are continuously evolving and expanding and speakers must consequently adapt.

This paper applies another mechanism than genetic evolution for structure formation to the problem of language formation, namely self-organisation. Self-organisation occurs in complex dynamical systems which are coupled in a particular way. Standard examples are the formation of a termite nest [2] or a path in an ant society.

This paper focuses on the formation of vocabularies, i.e. a set of couplings between words and meanings. A common vocabulary will be viewed as a self-organising phenomenon similar to a path in an ant society. Each agent is assumed to create his own vocabulary in a random fashion. But agents are coupled because they must share vocabularies in order to obtain the benefit of cooperating through communication. Agents therefore keep changing their own private vocabulary until it is conform to the common vocabulary. It will be shown that under certain conditions a coherent but still evolving vocabulary emerges.

The rest of the paper is in three parts. The first section introduces the kernel mechanism responsible for self-organisation. The second section introduces a spatio-temporal dynamics which ensures that the kernel mechanism copes with combinatorial explosions. The final section reports the results of some simulation experiments.

2 The self-organising kernel

We want to study how language, and more specifically vocabulary, may evolve in a group A of agents of size $\sharp A$. An agent $a \in A$ has an associated set of meanings M_a. For example an agent a_1 may want to communicate to another agent a_2 the presence of a particular vital resources r. This communication is in the interest of a_2 because it helps a_2 to replenish its vital resouces at reduced cost. It is also in the interest of a_1 when a_2 performs a similar communication in the future (repricocity). Therefore both agents benefit from communicating but only if they use the same language.

Each agent has a set of words W_a. The agent can randomly associate a word $w \in W_a$ to a particular meaning $m \in M_a$. When the agent expresses the meaning m he uses w. When the agent later hears the same word, he assumes that it has the same meaning. An association between w and m in a is denoted as $a : w = m$. Each agent can use a word only once. The communicative success of a coupling between w and m, denoted as $c_{w,m}$, is strongly related to the percentage of agents in the total group that use the same word for the same meaning. In the experiments in this paper we assume that $c_{w,m}$ is equal to this percentage, i.e. $c_{w,m} = (1-K)K/N$ with K the

Figure 1: The results of a typical experiment with 10 agents, 5 possible words, and 1 meaning. It plots the communicative success of each word (y-axis) over time (x-axis). We see a search period in which different words compete until one gains complete dominance.

number of agents using $c_{w,m}$.

Changes in the coupling between w and m should depend on communicative success. When all agents use the same coupling, i.e. $c_{w,m} = 1.0$, there should be no more changes. The simplest possible kernel mechanism for forming a language is therefore one in which agents randomly couple words to meanings, engage in communication, and then evaluate after a certain period of time how successful various communicative acts have been. Depending on this success the coupling between the word and its meaning is maintained or a new coupling is randomly created. The decision is based on a sigmoid function so that the probability of change decreases quickly as more than a majority adopts the same coupling. Figure 1 shows results for 10 agents, 5 words and 1 meaning. We see that a coherent association between a word and a meaning arises.

Figure 2 shows results when several meanings are involved. There are 10 agents, 5 possible words and 5 possible meanings. Contrary to what might be expected, there is no combinatorial explosion but an implosion. Once a word has a consistent meaning it is no longer available and so the set of possible choices for the others are shrinking. Consequently we see a rapid

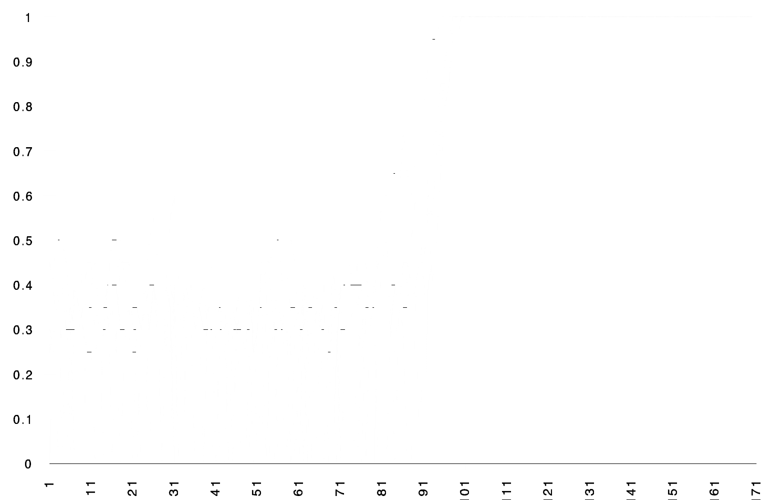

Figure 2: The graph plots the highest communicative success of all words over time for an experiment with 20 agents, 5 different meanings and 5 possible words. The combinatorial implosion happens after 90 time steps.

evolution towards coherence as soon as a sufficient part of the vocabulary has been established. This phenomenon is similar to the combinatorial implosions pointed out by Kauffman in the clustering and interconnection of autocatalytic networks [**?**].

The model also supports language learning. A new agent which is introduced in the group will create and try couplings. Only those that are conform to the rest of the community give any communicative success and are retained (see figure 4). The agent is forced to adopt the existing common vocabulary which is by now well entrenched.

The proposed mechanism clearly works for small-scale groups and limited vocabularies. But any mechanism will only be valid if it supports a significant scale up along three dimensions: the number of agents, the number of meanings, and the number of words. It is easy to see that a significant scale up is however not possible because increases along each of the dimensions will decrease the probability that a sufficient number of agents will adopt a common word.

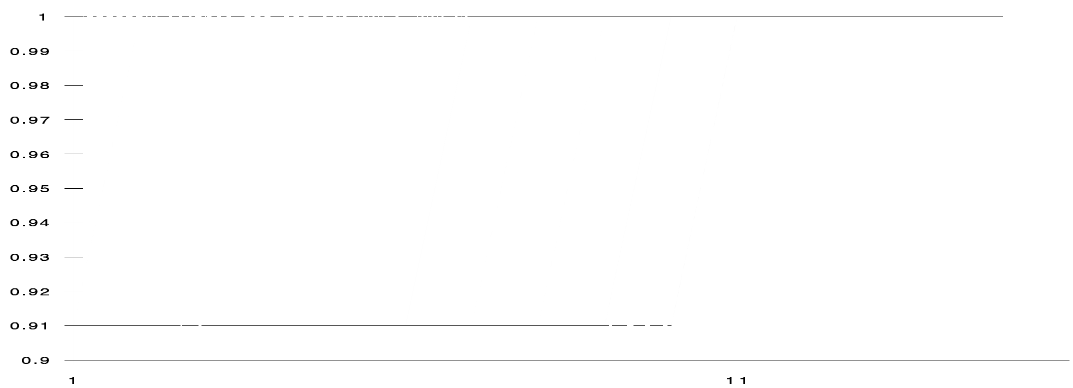

Figure 3: The graph plots how a new agent quickly adopts the common vocabulary maintained by a group of 10 agents with 10 meanings and 10 different words. The y-axis shows the communicative success of each word. Communicative success moves down to 0.91 as long as the new agent has not adopted the word.

3 Spatio-temporal dynamics

First of all, in normal circumstances, agents have conversations involving a limited set of words. The conversations are embedded in specific contexts in which only a limited set of meanings arise. It will consequently take more time before a coherent global language develops. But the search problem is also reduced because only a limited set of words and meanings are considered in a particular conversation. Second, it cannot be assumed that all agents already know all meanings and all words. Instead, a meaning recognised by one agent might be adopted by another agent and a word already possessed by one agent might be imitated and thus incorporated in the vocabulary of the other agent. Let me now discuss some investigations based on these conditions.

A *conversation* is defined as an event that involves a set of agents, a set of meanings, and a set of words. To create a conversation, a set of agents is selected randomly from the complete set of agents, a set of meanings is selected from the set of meanings of each of the agents, and the set of words is set equal to those the agents already associate words with these meanings. To determine the communicative success of an association k for a given conversation, the same formula as before, namely $c_k = (K-1)K/N$, is used. N is now the number of agents involved in the conversation, and K the number of agents that use the association k in the conversation.

For a particular agent and a particular meaning, there are now three possibilities and consequent actions:

1. 1 [1] The agent had a word for this meaning. This association k will be changed based on the probability p_k which depends on c_k using the same sigmoid function as before.

2. 2 [2] The agent did not have a word for this meaning. In that case, the agent selects a word from the words taking part in the conversation and associates it with the meaning. If there are not enough words left, the agent may create a new word and then create a new association.

3. 3 [3] The agent did not know about this meaning yet. In this case, the meaning is adopted and a word is chosen for it, using the same principle as in [2].

Steps [2] and [3] ensure that meanings and words propagate in the population. For the time being, new meanings are created externally and sprinkled at random over the population in a kind of innoculation process.

Here are some snapshots of conversations taken from a simulation: The first example shows a conversation with 3 agents. Two meanings are raised: m3 and m5. All agents are already familiar with all meanings. But only one agent has a word: agent-32 uses (A D A C C) for m5. The communicative success of this association is therefore 0.0. The agents change associations and end up to all select (A D A C C) for m3. Also a new word is created for m5.

```
Conversation 42
Agents: agent-35, agent-32, agent-105
Language used:
m5:
No word for 5 in agent-105
No word for 5 in agent-35
  (A D A C C): agent-32 [0.00]
m3:
No word for 3 in agent-105
```

```
No word for 3 in agent-32
No word for 3 in agent-35
Language generated:
m5:
No word for 5 in agent-105
No word for 5 in agent-32
  (D C A B A): agent-35
m3:
  (A D A C C): agent-105 agent-32 agent-35
```

The second example shows a situation much later involving 4 agents and 3 meanings. Three agents use (B D D D B) and receive 0.75 as communicative success. Two use (D C A B A) and therefore get 0.67. Otherwise success is either equal to 0.0 or no word is available yet. After the change, only one improvement can be seen because the third agent has adopted (D C A B A) for m2. Note that agent-80 already has words for every meaning but uses them in each case differently from the rest.

```
Conversation 98
Agents: agent-17, agent-80, agent-5, agent-250
Language used:
m5:
No word for 5 in agent-5
No word for 5 in agent-80
No word for 5 in agent-17
  (D C A B A): agent-250 [0.00]
m1:
  (B D D D B): agent-250 [0.75] agent-5 [0.75] agent-17 [0.75]
  (A D A C C): agent-80 [0.00]
m2:
No word for 2 in agent-250
  (D C A B A): agent-5 [0.67] agent-17 [0.67]
  (B D D D B): agent-80 [0.00]
Language generated:
m5:
No word for 5 in agent-250
No word for 5 in agent-5
No word for 5 in agent-17
```

```
  (B D D D B): agent-80
m1:
  (B D D D B): agent-250 agent-5 agent-17
  (D C A B A): agent-80
m2:
  (D C A B A): agent-250 agent-5 agent-17
  (A D A C C): agent-80
```

The third example shows a situation later still involving 4 agents and three meanings m4, m0 and m5. m4 and m0 are already coherently associated to the words (A D A C C) and (C C A) and therefore have a communicative success of 1.0. m5 has now also a coherent association except in one agent (agent-29) which does not have a word yet. The word (D C) which is 'floating around' in the conversation is picked up by agent-29 and associated with m5.

```
Agents: agent-29, agent-80, agent-17, agent-5
Language used:
m4:
  (A D A C C): agent-5 [1.00] agent-17 [1.00] agent-80 [1.00]
                  agent-29 [1.00]
m0:
  (C C A): agent-5 [1.00] agent-17 [1.00] agent-80 [1.00]
                  agent-29 [1.00]
m5:
No word for 5 in agent-29
  (D C): agent-5 [1.00] agent-17 [1.00] agent-80 [1.00]
Language generated:
m4:
  (A D A C C): agent-5 agent-17 agent-80 agent-29
m0:
  (C C A): agent-5 agent-17 agent-80 agent-29
m5:
  (D C): agent-5 agent-17 agent-80 agent-29
```

Note that a coherent choice among a group of agents does not mean that the whole community has adopted the same word. There could be distinct groups of agents that temporarily reach local coherence but for different associations. Global coherence will only be reached in a 'well-stirred' community where there are enough encounters between the agents to result in a globally coherent language.

References

[1] Chomsky, N. (1980) Rules and Representations. Vol 3. pp. 1-15.

[2] Deneubourg, J-L. (1977) Application de l'ordre par fluctuations a la description de certaines etapes de la construction du nid chez les termites. Insectes Sociaux, Tome 24, 2, p. 117-130.

[3] Hashimoto, and Ikegami (1995)

[4] Kauffman, S. (1993) The origins of order. Oxford University Press. Oxford.

[5] MacLennan, B. (1991) Synthetic Ethology: An Approach to the Study of Communication. In: Langton, C., et.al. (ed.) Artificial Life II. Addison-Wesley Pub. Co. Redwood City, Ca. p. 631-658.

[6] Werner, G. and M. Dyer (1991) Evolution of Communication in Artificial Organisms. In: Langton, C., et.al. (ed.) Artificial Life II. Addison-Wesley Pub. Co. Redwood City, Ca. p. 659-687.