The Spontaneous Self-organization of an Adaptive Language.

Luc Steels
Sony Computer Science Laboratory
6 Rue Amyot 75005 Paris
Artificial Intelligence Laboratory
Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussels, Belgium
E-mail: `steels@arti.vub.ac.be`

December 5, 1997

Abstract

The paper studies how a group of distributed agents may spontaneously and autonomously develop a language to refer to other agents in their environment by engaging in a series of language games. The language is adaptive in the sense that it expands or adjusts to the entry of new agents and new meanings. The paper describes the language formation mechanisms and details the results of computational simulations.

Keywords: origins of language, self-organization, distributed agents, open systems.

1 Introduction

The paper proposes a set of mechanisms by which a group of distributed agents may develop autonomously a language for identifying other agents in their environment. The set of agents and the set of features used for making distinctions are open-ended. The language autonomously adapts by the individual actions of agents with only local interactions.

Concretely, three mechanisms are proposed: (1) Agents adopt word-meaning associations from others and thus words *propagate* in the population, (2) Agents may *generate* a new word and associate it with an uncovered feature set, and (3) there is a positive feedback mechanism between the selection of a word in a conversation and the success so far in using that word, thus leading to self-organized *coherence*. The mechanisms have been implemented and validated in computational experiments. The emergent languages do not have the full complexity of natural languages, for example because there is no syntax. However they involve expressions with multiple words, expressions with multiple meanings (ambiguity), and meanings with alternative expressions (synonymy).

The research discussed here has primarily a scientific motivation. The experiment is seen as a step in the investigation how language may originate and evolve in distributed agents. The origins of language must have been one of the crucial steps in the origins of intelligence [14]. Understanding these origins forms part of recent investigations in biology to understand the major transitions towards more complexity in living systems, beginning with the origin of life itself,[5] and of 'artificial life' experiments seeking to synthesise a spontaneous increase in complexity [4]. Of particular interest here are related experiments on the origin of communication [6], the origin of vocabulary [20] [12], and the growth in complexity of syntax [3].

There is also a secondary motivation. Understanding the mechanisms by which a language self-organises, may make a bottom-up approach to artificial intelligence possible [10]. Although much progress has been made recently on the synthesis and acquisition of sensori-motor behavior [9] [13], the bottom-up approach still needs to be shown to be effective for cognitive capabilities. A focus on language formation is one possible route to explore this.

The next section briefly and informally sketches the main ideas proposed in the paper. Then these ideas are presented more formally and illustrated with examples. Next a number of sections detail simulation results. Section 4 shows the formation of a language from scratch, section 5 how the language adapts to new agents entering the group, and section 6 how the language adapts to new distinctions becoming available for lexicalisation. Some further results are discussed in section 7.

2 Basic principles

Consider a set of agents capable to exchange messages with each other. Each agent has a set of features which can be used to distinguish one agent from the other agents. The agents can engage in *language games* [21]. One agent, the *initiator*, identifies an agent, the *topic*, out of a set of other agents which constitutes the *context*. Another agent, the *recipient*, must identify the agent chosen as topic. There are two possible ways to do so: either the initiator points to the agent so that identification is direct, or the initiator uses language. Language formation and acquisition is only possible when the initiator first uses pointing and then language. When more and more language becomes available, purely linguistic means suffice.

To use language, the initiator must first identify which possible sets of features distinguish the agent chosen as topic from the other agents present in the context. There could be several such distinctive feature sets. He then chooses one and encodes this set into an expression. An expression contains one or more words. Each word encodes one or more features. Words are allowed to be ambiguous and there is not necessarily a unique way in which a given set of meanings gets encoded into an expression (synonymy).

Next, the recipient decodes the expression. If the recipient already knows which object is intended, he can identify which possible sets of features distinguish the object. From that, the recipient can confirm that the expression encodes one of the expected distinctive feature sets or infer and possibly adopt new associations between words and meanings. This feedback then enables both agents to adjust their lexicons. Each language game has therefore two dimensions. On the one hand, there is the functional dimension of identifying an object using linguistic or extra-linguistic means. On the other hand, there is the linguistic dimension in which negotiations take place about the shape of the language itself.

As we will soon see, this mechanism enables word-meaning pairs to propagate in a population of agents. But this is not enough in itself. There must also be a way in which agents can extend the language whenever the existing language is not adequate, i.e. when no word exists to express a set of features. This is achieved by allowing the initiator to make up a new word and associate it with an uncovered feature set. This action is called *lexicalisation*. The creation of a new word-meaning pair happens with very low probability because the more words exist in a population the longer it takes to reach coherence.

Coherence is achieved through self-organization, in the sense of spontaneous

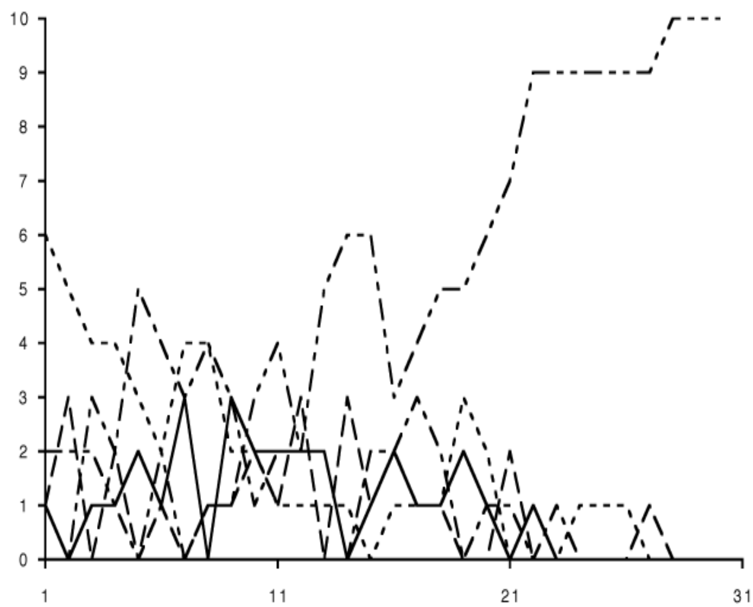

Figure 1: This figure shows the results of an experiment with 10 agents, 5 possible words, and 1 meaning. It plots the average communicative success of each word (y-axis) over a series of language games (x-axis). We see a search period in which different associations compete until one gains complete dominance.

formation of (dissipative) structures through the enforcement of random variations, known as fluctuations [8]. The fluctuations are here caused by the different associations floating around in the population. An agent records how many times a word-meaning association has been used in a game in which he was involved and how many times it was successful. When meanings need to be encoded, the agent picks the most commonly used successful association. This introduces a positive feedback loop: The more a word gets used, the more successful it will be and therefore the more it solidifies. When there are multiple possibilities a temporary struggle goes on until one association wins (see fig. 1) [11].

So we see three mechanisms: *propagation*, when agents adopt word-meaning associations from others, *creation*, when an agent generates a new word and associates it with an uncovered feature set, and *self-organisation*, due to the positive feedback mechanism between association selection and success in use.

Before I discuss the details of these mechanisms more formally, some interesting properties must be pointed out.

1. There is no single agent with a complete view of the language. Each agent has his own private set of associations between words and meanings and can only experience the associations used by others by engaging in language games. There is not a single agent in charge of creating new language. Each agent is allowed to make new words whenever it is needed. We therefore have a fully *distributed* system.

2. We also have an *open* system. At any point in time new agents are allowed to enter the group of existing agents or agents are allowed to leave. A new agent acquires the language of the group using the same mechanisms as those giving rise to the language in the first place. Because new agents are allowed to leave or enter, the set of topics of a conversation and the set of possible contexts are also open. The language continuously adapts, in a distributed fashion, to allow the necessary distinctions to be expressed. The system is also open with respect to features. At any point in time new features are allowed to enter into consideration for distinguishing between different agents. When these features are relevant they become lexicalised.

3. The languages derived using the proposed mechanisms have some important simularities with natural languages:

1. The same word may have different meanings, and therefore an expression may be ambiguous. The context or additional words are then used to determine which meaning is intended.

2. The same meaning may be expressable by different words, and therefore we see synonyms emerge.

3. There is never complete coherence among agents due to different developmental pathways. Nevertheless almost complete communicative success is observed after a sufficient number of language games, because the language has become adequate enough to deal with all possible contexts of use.

3 Detailed description of the mechanisms

3.1 Terminology

Let there be a set of agents $A = \{a_1, ..., a_n\}$. Each agent a is assumed to have a set of features $F_a = \{f_1, ..., f_m\}$. $\mathcal{F}$ is the set of all possible feature sets. A feature

f_k consists of a pair $(p\ v)$ where p is called an attribute and v a value. A set of features $D^B_{a_1}$ distinguishes an agent a_1 from a set of other agents $B = \{a_2, ..., a_n\}$ iff $D^B_{a_1} \subset F_{a_1}$ and $\forall a \in B, D^B_{a_1} \not\subset F_a$. $D^B_{a_1}$ is called a *distinctive feature set* with respect to a_1 and B. There can be several distinctive feature sets for the same a_1 and B. There can also be none if $F_{a_1} \subseteq F_{a_k} \in B$.

A *word* is a sequence of letters drawn from a finite shared alphabet. In the experiments reported later, a word is a consonant-vowel sequence, such as "(t a)" or "(k i)". An *expression* is a sequence of words. Word order is not assumed to play a role.

A *lexicon* $L \subset F \times W$ is a relation between a possible feature set $K \subset F$ and a word $w \in W$. Each member of this relation is called an *association*. Each agent $a \in A$ is assumed to have a single lexicon L_a which is initially empty.

$u(<K, w>, a)$ is the number of times the association $<K, w> \in L_a$ has been used by a. $s(<K, w>, a)$ is the number of times the association $<K, w> \in L_a$ was used successfully by a, i.e. when it is part of a language game which ended with communicative success (defined shortly).

3.2 Language Games

A language game $g =< C, i, r, o >$ includes a context $C = \{a_1, a_2, ..., a_j\} \subseteq A$, an initiator $i \in C$, a recipient $r \in C$, and a topic $o \in C$. The language game involves the following steps:

1. Both the initiator i and the receiver r determine the distinctive feature sets $\mathcal{D}^C_o = \{D^B_o \mid B = C \backslash \{o\}\}$. It is assumed that both agents share the same distinctive feature set.

2. The initiator chooses one feature set $D_j \in \mathcal{D}^C_o$ and constructs an expression e which *covers* D_j.

3. The recipient *uncovers* from e the feature sets $\mathcal{H} = \{H_1, ..., H_p\}$

4. g ends in communicative success when $H \bigcap \mathcal{D}^C_o \neq \emptyset$, otherwise in failure.

The *cover* and *uncover* functions are at the heart of the language encoding and decoding process. They are defined as follows:

- Given an agent $a \in A$, $cover(D, L_a) = \{e \mid e = \{w \mid D = \bigcup K\ with\ < K, w> \in L_a\}\}$.

- Given an agent $a \in A$, $uncover(e, L_a) = \{H \mid H = \bigcup K \; with \; <K, w> \in L_r, w \in e\}$.

The cover function yields a set of possible expressions. Only one expression is selected for use in the communication, based on two criteria: (1) the smallest set is preferred and (2) in case of equal size, an expression is preferred for which the implied associations score better. The score $m(<K, w>, a) = s(<K, w>, a)/u(<K, w>, a)$ for $<K, w> \in L_a$.

The overall result of a language game is either

1. that there are not enough distinctions to identify the topic in this context, $\mathcal{D}_o^C = \emptyset$.
2. That the game ends in communicative success, i.e. $\mathcal{H} \bigcap \mathcal{D}_o^C \neq \emptyset$.
3. That the game ends in communicative failure, which could take many different forms:
   (a) The initiator i may not have enough words to cover all the features, $cover(D_j, L_i) = \emptyset$.
   (b) The recipient r may not have enough associations to uncover all the meanings, $uncover(e, L_r) = \emptyset$.
   (c) There is a mismatch between expected and uncovered meanings, $\mathcal{H} \bigcap \mathcal{D}_o^C = \emptyset$.

These different types of results and the corresponding steps in language formation are discussed in more detail in the next subsection.

I have implemented the mechanisms needed to engage in language games as a computer program written in LISP. The program creates agents, assigns randomly values for features from the set of possible features, and then starts a series of language games. For each game, a context, initiator, recipient and topic are randomly chosen. A language game is printed out by the program as follows:

```
Dialog 47 between a-2 and a-4 about a-2.
    Context: {a-2 a-6 a-5 a-4 }
    a2 = (((size tall))
          ((shape square)))
 a-2: ((size tall)) => ((d o))
 a-4: => ((size tall))
(success)
```

In this game a-2 plays the role of initiator, a-4 that of recipient. The topic is a-2. The context is $\{$ a-2, a-6, a-5, a-4 $\}$. There are two distinctive feature sets: $((size\, tall))$ and $((shape\ square))$. a-2 has picked the first one and has translated this into the expression "((d o))". a-4 uncovers $((size\ tall))$ from "((d o))". The language game ends in communicative success because $((size\ tall))$ is one of the expected feature sets. The indication of the result of a game often contains additional information.

3.3 Language formation rules

This section details the different rules that agents follow in the adoption or formation of language.

```
0. No distinctions possible
```

The first case is one where there are not enough features available to distinguish the topic from the other agents present in the context. This should put pressure on a meaning creation process to introduce a new distinction (see section 6.).

```
1. Lexicon inadequate for initiator.
```

The initiator may not have enough associations in his lexicon to cover all the meanings in the chosen feature structure D_j. In this case the game ends in failure and it is indicated for which features no words were available. The usage u of all associations that were used by the initiator are incremented, but not the success s. The initiator may create a new word (with probability 0.05 in the present experiments) and associate it in his lexicon with the non-covered meanings. This is happening in the next example:

```
Dialog 54 between a-6 and a-5 about a-6.
    Context: {a-5 a-6 a-4 a-3 }
    a6 = ((weight light))
 a-6: ((weight light))  => ?
 a-5: => nil
     !! a-6: ((weight light)) -> (t u)
(failure-no-words-initiator nil ((weight light)))
```

a-6 has created the word "(t u)" to express $((weight\ light))$. Note that the word is not yet used in the conversation and there is no impact on the recipient a-5. If there are several features that are not covered yet by the lexicon, this rule will automatically lead to words that are associated with a set of features instead of just one.

The same situation may also occur when only a subset of D_j cannot be covered. This is illustrated in the next example. a-7 has already a word for $((weight\ light))$ but not yet for $((size\ medium))$. a-7 creates a new word "(v i)".

```
Dialog 39 between a-7 and a-2 about a-6.
    Context: {a-2 a-7 a-5 a-3 a-6 }
    a-6 = ((weight light) (size medium))
 a-7: ((weight light) (size medium)) => ?
 a-2: => nil
     !! a-7: ((size medium)) -> (v i)
(failure-no-words-initiator
    (((d o) ((weight light)) (3 . 0))) ((size medium)))
```

```
2. Lexicon inadequate for recipient
```

The recipient may not have enough associations in his lexicon to uncover all the meanings from e. In this case the game ends in failure and it is indicated for which words no associations were available. Several possibilities can be distinguished.

```
2.1. No words at all could be uncovered.
```

In this case, the recipient can deduce that the expression must be associated with one of the feature sets in the distinctive feature sets, and the association is consequently constructed. This is shown in the following example.

```
Dialog 53 between a-3 and a-4 about a-7.
    Context: {a-4 a-3 a-6 a-7 }
    a-7 = ((shape square))
 a-3: ((shape square)) => ((s o))
 a-4: => nil
     !! a-4: ((shape square)) -> (s o)
(failure-no-words-recipient ((s o)))
```

This operation may also lead to ambiguities because there might be more than one way in which the topic is distinguished from the context, as shown in the next example. The word "(z u)" comes to mean both $((weight\ light))$ and $((size\ tall))$.

```
Dialog 78 between a-4 and a-5 about a-2.
    Context: {a-4 a-2 a-5 }
    a-2 = (((weight light))
           ((size tall)))
 a-4: ((weight light)) => ((z u))
 a-5: => nil
     !! a-5: ((weight light)) -> (z u)
     !! a-5: ((size tall)) -> (z u)
(failure-no-words-recipient ((z u)))
```

2.2. Some words could be uncovered

It could be that some words could be uncovered but others could not be. When there is only one word which is unknown, the recipient can deduce the meaning and a new association can be created. When there is too much uncertainty no changes are made. This is shown in the next example. a-11 uncovers the feature $(size\ tall)$ from the word "(d e)" and can infer that the word "(b i)" must be associated with the other feature $(weight\ heavy)$.

```
Dialog 946 between a-7 and a-11 about a-11.
    Context: {a-7 a-10 a-11 a-9 }
    a-11 = (((size tall) (weight heavy)))
 a-7: ((size tall) (weight heavy)) => ((d e) (b i))
 a-11: => nil
     !! a-11: ((weight heavy)) -> (b i)
(failure-partial-cover-recipient
    (((d e) ((size tall)) (14 . 12))) ((b i)))
```

3. No words missing

The next cases are concerned with situations where both the initiator and the recipient have associations to encode or decode the distinctive feature sets.

3.1. Success

Complete success is reached when the distinctive feature sets expected by the recipient includes the one uncovered from the expression used by the initiator. Both the use and the success of the association is incremented by the initiator and the recipient. This is shown in the next example:

```
Dialog 47 between a-2 and a-4 about a-2.
    Context: {a-2 a-6 a-5 a-4 }
    a-2 = (((size tall)))
 a-2: ((size tall)) => ((d o))
 a-4:  => ((size tall))
(success)
```

3.2. Success but too general

It may also be that there is success but that the recipient decoded more possible meanings for the expression. In that case, only the success of the association that was effectively relevant gets incremented. This leads to a progressive disambiguation of associations that were unnecessarily broad. Thus in the next example, a-2 decodes "((d o))" into both $((size\ tall))$ and $((weight\ light))$. Both are applicable but only one is really necessary. "((d o))" is therefore more broad than necessary.

```
Dialog 72 between a-3 and a-2 about a-2.
    Context: {a-2 a-3 a-5 a-6 a-7 }
    a-2  = (((size tall))
            ((weight light)))
 a-3: ((size tall)) => ((d o))
 a-2: => ((size tall))
        ((weight light))
(success-but-too-general-recipient
    (((d o) ((weight light)) (3 . 2)))
    ((d o)))
```

3.3. Mismatch in meaning

It may be that the feature set decoded by the recipient is not one of the feature sets that is distinctive for the object. The success record of the implied association is therefore not incremented. This is shown in the next example.

```
Dialog 505 between a-7 and a-6 about a-6.
    Context: {a-6 a-7 a-5 a-4 }
    a-6  = (((weight light)))
 a-7: ((weight light)) => ((d o))
 a-6: => ((size tall))
(failure-mismatch-meaning-recipient
    (((d o) ((size tall)) (10 . 6)))
    (((weight light))))
```

The same could happen in a multiple word sentence, as shown in the following example, where the word "(d o)" is decoded as $(size\ tall)$ whereas $(weight\ light)$ is expected.

```
Dialog 521 between a-7 and a-6 about a-6.
    Context: {a-7 a-4 a-5 a-2 a-6 }
    a-6 = (((weight light) (size medium))))
 a-7: ((weight light) (size medium)) => ((v i) (d o))
 a-6: => (((size tall)) ((size medium)))
(failure-mismatch-meaning-recipient
    (((d o)((size tall)) (13 . 8))
     ((v i)((size medium)) (2 . 1)))
    (((weight light) (size medium))))
```

4 The formation of a language from scratch

This section and the next ones, discuss in detail some simulation results obtained through an implementation of the mechanisms introduced in the previous section. It starts with a population of 5 agents. There are three possible features: $weight$ with possible values heavy, light, average, $shape$ with possible values square, oval, round, and $size$ with values small, medium, tall. The agents have the following features:

```
a-10:                          a-7:
  (weight heavy)                  (weight light)
  (size small)                    (size medium)
  (shape square)                  (shape round)
a-9:                           a-6:
```

```
  (weight average)                  (weight average)
  (size tall)                       (size tall)
  (shape round)                     (shape square)
a-8:
  (weight heavy)
  (size medium)
  (shape round)
```

After about 4000 language games, a stable language has emerged. The language is stable in the sense that the agents are no longer creating new words or new associations, although there may still be shifts in usage. The following shows the language that the agents preferentially use. For each possible feature, the words in use, and for each word, the agents that use it are shown.

```
((shape square)):
  [(t i) (a-6 a-7 a-8 a-9 a-10)]
((weight heavy)):
  [(n e) (a-6 a-7 a-8 a-9 a-10)]
((size small)):
  [(n e) (a-6 a-8 a-9 a-10)]
  [(t i) (a-7)]
((weight average)):
  [(r e) (a-6 a-7 a-9 a-10)]
  [(d a) (a-8)]
((size tall)):
  [(r e) (a-8 a-9 a-10)]
  [(d a) (a-6 a-7)]
((shape round)):
  [(b e) (a-6 a-7 a-8 a-9 a-10)]
((weight light)):
  [(s o) (a-6 a-7 a-8 a-9 a-10)]
((size medium)):
  [(z u) (a-6 a-7 a-8 a-9 a-10)]
```

Some features, such as $(shape\ oval)$ are not lexicalised because they are irrelevant. Others like $(weight\ average)$ and $(size\ tall)$ are covered by the same word "(r e)" or "(d a)" because they always occur together (in a-9 and a-6) so agents cannot distinguish which one of the features is intended. Some words are ambiguous. For example, a-6 uses “(n e)” for $(weight\ heavy)$ as well as $(size\ small)$. However it does not have a further impact on communicative success. Either the distinction does not matter because either one of these features is enough to distinguish. For example, if a-8 is not part of the context and a-10 needs to be identified, either feature will do. Or, agents use an additional word to express the value for $shape$.

When we look at the lexicons of individual agents, a more complex picture is emerging. Here are for example the associations involving $(size\ small)$ with u and s printed out for each agent.

```
((size small)):
  [(n e) (a-6 (272 . 204)) (a-7 (203 . 88))
         (a-8 (285 . 156)) (a-9 (225 . 152))
         (a-10 (328 . 296))]
  [(t i) (a-6 (138 . 10)) (a-7 (139 . 70))
         (a-8 (115 . 12)) (a-9 (129 . 5))
         (a-10 (136 . 24))]
```

For the feature $(size\ small)$ most agents use the word “(n e)” but they all have an association with “(t i)” as well, whose score is less, except for a-7. “(t i)” is a rest from earlier usage where it meant both $(size\ small)$ and $(shape\ square)$ as required for distinguishing a-10.

A typical conversation at this point is the following:

```
  a-10 = (((weight heavy)))
a-8: ((weight heavy)) => ((n e))
a-7: => ((size small)) ((weight heavy))
```

a-8 uses “(n e)” to encode $(weight\ heavy)$. This is ambiguous for a-7. But $((weight\ heavy))$ is the only way the topic a-10 can be distinguished. The language game succeeds (case 3.2) and the success of the association between “(n e)” and $(weight\ heavy)$ is incremented. Resolution of the ambiguity of “(n e)” will occur when there are more cases like this.

Here is another example:

```
  a-8 = (((shape round) (weight heavy))
         ((weight heavy) (size medium)))
a-10: ((shape round) (weight heavy)) => ((b e) (n e))
```

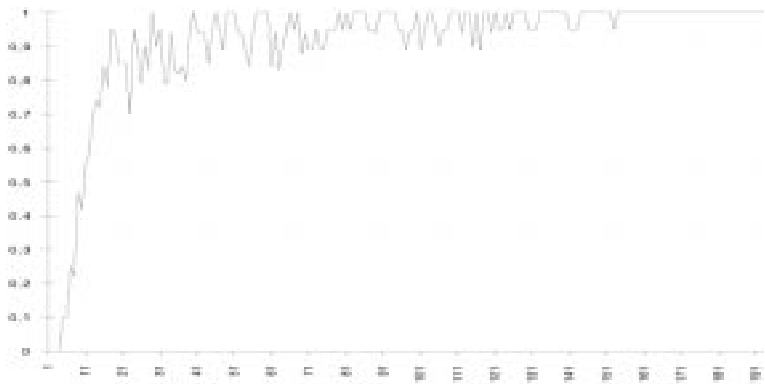

Figure 2: This figure plots the formation of a language from scratch. 4000 language games are shown. The x-axis plots the number of language games (scale 1/20). The y-axis shows the average communicative success.

```
a-9: => ((size medium))
        ((size small))
        ((weight heavy))
        ((shape round))
```

a-9 decodes “(b e)” as $((shape\ round))$ or $((size\ medium))$ and “(n e)” as $((weight\ heavy))$ or $((size\ small))$. In this case, the use of $((weight\ heavy))$ is confirmed because one way the topic can be distinguished is as $((shape\ round)\ (weight\ heavy))$. Another combination $((size\ medium)(weight\ heavy))$ also matches with possible decodings of “((b e)(n e))”.

Fig. 2 shows the evolution of the language from the beginning. The communicative success climbs steadily from 0 to become absolute. The agents have developed sufficient words to distinguish each other based on their features.

5 Adding new agents

A new agent entering the group acquires the already existing language. At the same time, new words may get created because new distinctions become lexicalised. This is shown in the following experiment. The new agent has the following features:

```
a-11:
```

```
  (weight heavy)
  (size medium)
  (shape oval)
```

The prefered language of each agent after about 2000 language games looks as follows:

```
((weight heavy)):
  [(n e) (a-6 a-7 a-8 a-9 a-10 a-11)]
((weight light)):
  [(s o) (a-6 a-7 a-8 a-9 a-10 a-11)]
((shape round)):
  [(b e) (a-6 a-7 a-8 a-9 a-10 a-11)]
((size small)):
  [(n e) (a-6 a-8 a-9 a-10 a-11)]
  [(t i) (a-7)]
((shape square)):
  [(t i) (a-6 a-7 a-8 a-9 a-10 a-11)]
((size medium)):
  [(z u) (a-6 a-7 a-8 a-9 a-10 a-11)]
((shape oval)):
  [(z u) (a-6 a-7 a-8 a-9 a-10 a-11)]
((weight average)):
  [(n u) (a-11)]
  [(r e) (a-6 a-7 a-9 a-10)]
  [(d a) (a-8)]
((size tall)):
  [(d a) (a-6 a-7 a-10 a-11)]
  [(r e) (a-8 a-9)]
```

We see that $(shape\ oval)$ has become lexicalised to make the distinction between a-8 and a-11. Note also that a-11 has introduced a new word "(n u)" for $(weight\ average)$. The other agents understand this word but will not use it. There is now a majority using "(d a)" for $(size\ tall)$. The semantic resolution of "(d a)" and "(r e)" has therefore made some progress.

A typical conversation with this language is the following:

```
  a-6 = (((shape square) (weight average))
```

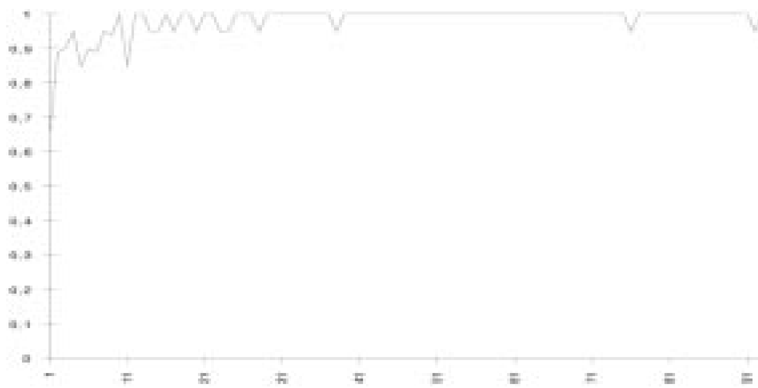

Figure 3: This figure plots 2000 language games illustrating the adaptation of language after a new agent comes in. The x-axis plots the number of language games (scale 1/20). The y-axis shows the communicative success.

```
         ((shape square) (size tall)))
a-10: ((shape square) (size tall)) => ((t i) (d a))
a-11: => ((size tall))((weight average))
         ((shape square))((size small))
```

a-10 selects $((shape\ square)(size\ tall))$ and encodes this as "((t i) (d a))". "(d a)" is decoded by a-11 as $((size\ tall))$ or $((weight\ average))$, "(t i)" as $((size\ small))$ or $((shape\ square))$. Out of this a-11 can reconstruct one possible distinctive feature set compatible with the expected ones. The language game therefore ends in success.

Fig 3. plots how the language adapts after the next agent (a-12) is added. Initially there is a drop in communicative success but after some time period, adjustments and new lexicalisations restore it.

6 Adding new features

When more and more agents are added, the available features become insufficient to distinguish one agent from others. Consequently language games start to fail on that basis. This is seen in fig 4. When the existing set of features fails to make distinctions, new features could be added by an independent meaning creation process. The language must then adapt by lexicalising the features that are

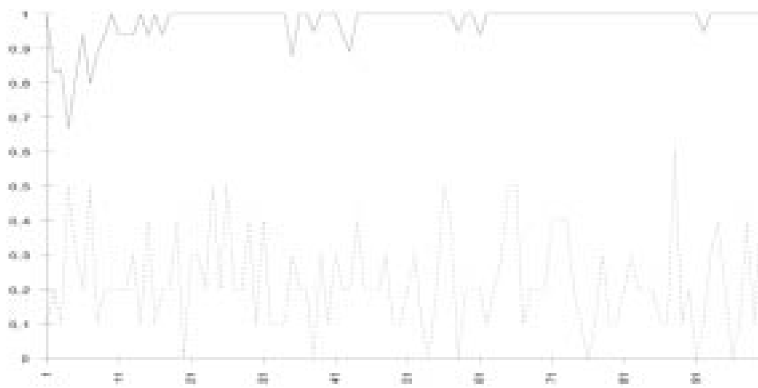

Figure 4: Another agent enters and the language adjusts. 2000 language games are shown. The y-axis shows the communicative success (top graph) and the failure in making distinctions (bottom graph). Many language games fail by lack of available distinctions because too many agents have joined the group that have the same features as those already present.

relevant for expressing the new distinctions. This is shown in the following experiment. A new feature $color$ has been added with possible values blue, red, yellow and green. All agents now receive a value for this new feature. For example, a-14 and a-13 have the following features:

```
a-14:                         a-13:
  (color blue)                  (color red)
  (weight heavy)                (weight average)
  (size medium)                 (size small)
  (shape oval)                  (shape round)
```

After about 2000 language games, these colors have become lexicalised:

```
((color blue)):
  [(r u) (a-6 a-7 a-8 a-9 a-10 a-11 a-12 a-13 a-14)]
((color yellow)):
  [(m e) (a-6 a-7 a-8 a-9 a-10 a-11 a-12 a-13 a-14)]
((color red)):
  [(t e) (a-6 a-7 a-8 a-9 a-10 a-11 a-12 a-13 a-14)]
((color green)):
  [(p i) (a-6 a-7 a-8 a-9 a-10 a-11 a-12 a-13 a-14)]
```

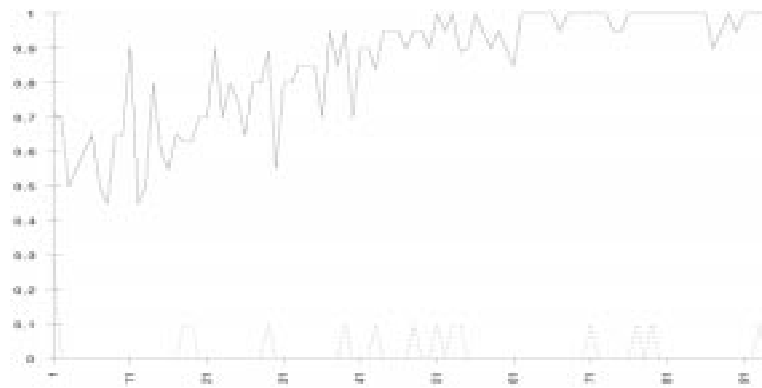

Figure 5: This figure plots the adaptation of language after a new distinction has been introduced. The failure to make distinctions has decreased. 2000 language games are shown.

Fig 5. plots the adaptation of the language. There is a drop in communicative success when the new distinction is introduced but after a certain period, communicative success reaches again the maximum. The success in making distinctions has also improved.

A similar experiment where a new feature with 5 possible values is introduced is shown in fig 6. Again we see a drop in communicative success followed by a rebound.

7 Further Results

The results discussed in the previous paragraphs were originally obtained in 1994. Since then, considerable progress has been made on many related issues. This section gives a brief overview with references to additional papers. Also other research groups have been advancing, as surveyed in [**?**].

The first question concerns scale up: Is it possible to increase the population of agents substantially? Is it possible to have a steady in- and outflux of agents and meanings? Further computational results show that a scale up is indeed possible. It is obvious that the time needed to reach communicative success (and hence coherence in the language) increases with the number of agents in the population, and this imposes a limit on what is eventually feasable. On the other hand, we have been able to show that a population can cope with a growing set of meanings and a

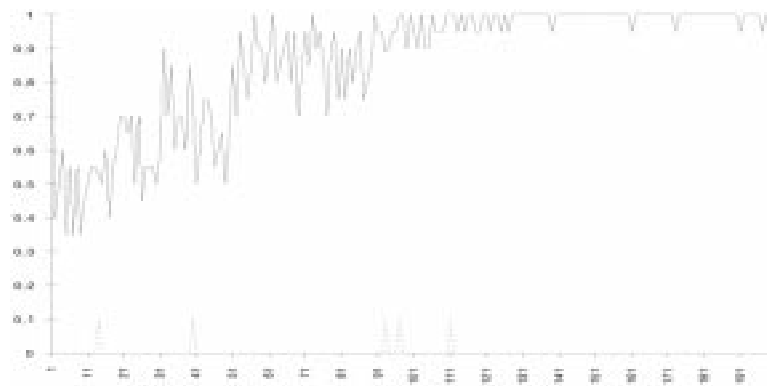

Figure 6: This figure plots the adaptation of language after yet another distinction has been added. 2000 language games are shown.

growing set of new agents entering. So scale-up is reached by gradual extension of the population. The growth rates are obviously bounded but nevertheless positive (see fig 7). More detailed results and discussions can be found in [16].

A second question concerns the origin of the distinctions used for constructing distinctive features. Could they be derived as well by the agents without human intervention? The answer to this question has turned out to be positive as well. It has been shown that a group of agents can develop a repertoire of distinctions recognised through discrimination trees, driven by the need to find distinctive feature sets [15]. This mechanism has been coupled to the lexicon formation process described in this paper [17], showing a co-evolution of a meaning repertoire and a lexicon. The system is still open in the sense that new objects may enter in the environment at all times. These results are illustrated in fig 8.

A third question concerns the embedding of these mechanisms in physical robots. One of the original targets of the research is to understand mechanisms by which agents could autonomously develop visually grounded categories and a language which is grounded in their sensori- motor experience. Again we have been able to obtain positive results. The lexicon formation and ontology creation mechanisms have been ported to mobile robots [**?**] and to two ”talking heads” [18]. In both cases, the agents are challenged to develop a repertoire of distinctions and a lexicon which is adequate for identifying the objects present in the environment. Fig 9 illustrates the physical setup used in these experiments.

A final question relates to the development of a syntax. We have seen already

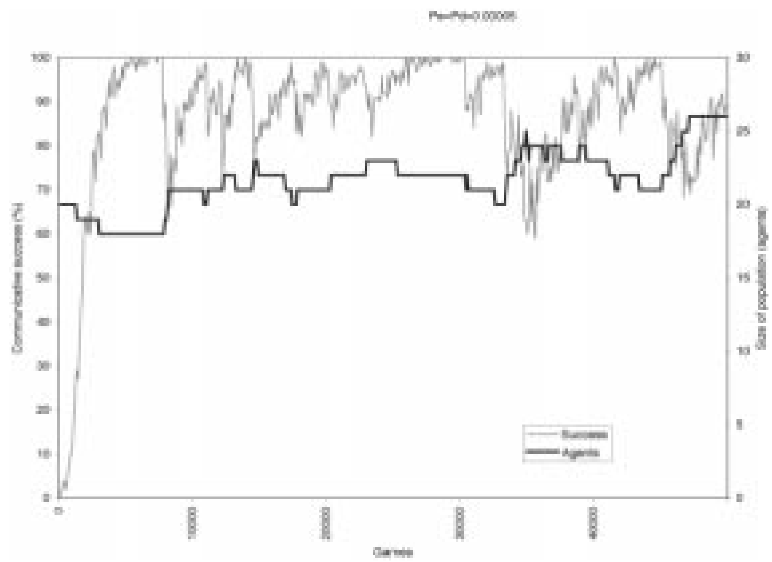

Figure 7: The figure shows the evolution of the communicative success in a fluctuating population of agents. New agents enter with a probability 0.00005 and they depart with the same probability. The agents engage in games in which they identify themselves. The population is able to cope with the flux.

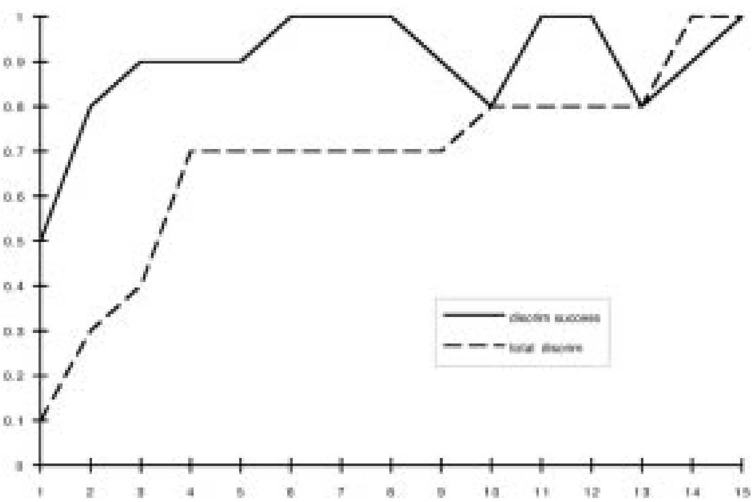

Figure 8: The graph shows the evolution of the discriminatory capacities of a single agent. The total number of objects (10) is fixed. There are 5 data channels. The average success in discrimination games as well as the total success is shown on the y-axis. The number of discriminations is mapped on the x-axis (scale 1/10). All objects can be discriminated after 150 discrimination games.

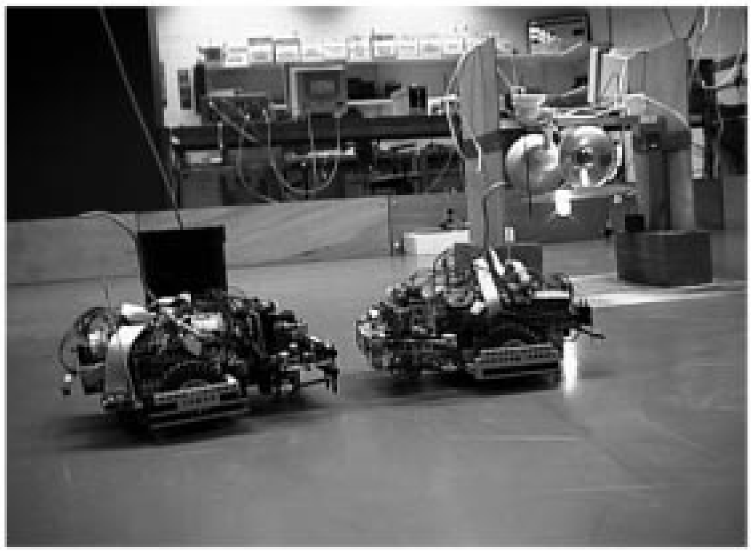

Figure 9: Two robots have approached each other and are now facing each other. The robots are equiped with a dozen low-level sensors. The discrimination trees are based on data coming from these sensors. They will give rise to categories which are then used in the language games as described in this paper. Note the other objects in the environment surrounding the robots, which will be the subject of the conversation.

in the examples in this paper that multiple word sentences arise. Recently, the first positive results were obtained where we could show a preliminary syntax by adding an episodic, schema-based memory that starts from recording the situations arising in multi word sentences, and then re-instantiates their constraints when similar situations arise. More details are contained in [18].

8 Conclusions

The paper has proposed a set of mechanisms by which distributed agents spontaneously and autonomously develop a coherent language to identify each other using distinctive features. The system is open in the sense that new agents and new distinctions may be introduced. The language adapts by lexicalising new features or by resolving ambiguities in the use of certain words. The mechanisms have been tested experimentally through computer implementation.

The spontaneous formation of a language in a group of distributed agents appears to be quite feasable and goes surprisingly fast. Cultural processes such as the ones proposed here, provide an alternative to genetic explanations for the origin [7] or acquisition of language [1]. Self-organization is very common in other areas of biology [2] and it is therefore not surprising that it might play an important role in language formation.

9 Acknowledgement

The research and writing of this paper has been financed by the Belgian Federal government FKFO project on emergent functionality (FKFO contract nr. G.0014.95) and the IUAP 'Construct' Project (nr. 20) of the Belgian government, with additional support from the external researcher program of the Sony Computer Science Laboratory in Tokyo. Josefina Sierra provided valuable comments to the paper.

References

[1] Chomsky, N. (1980) Rules and Representations. Brain and Behavior Science. 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) Evolution of Symbolic Grammar Systems. In: Morán, F, A. Moreno, J. Merelo and P. Chacón (Eds.) (1995) Advances in Artificial Life. Third European Conference on Artificial Life. Granada, Spain. Springer-Verlag, Berlin. p. 812-823.

[4] Langton, C. (ed.) (1995) Artificial Life. An overview. The MIT Press, Cambridge, Ma.

[5] Maynard-Smith, J. and E. Szathmary (1994) The major transitions in evolution. Oxford University Press, Oxford.

[6] 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.

[7] Pinker, S. (1994) The language instinct. Penguin Books, London.

[8] Prigogine, I. and I. Stengers (1984) Order Out of Chaos. Bantam Books, New York.

[9] Steels, L., ed. (1995) The Biology and Technology of Intelligent Autonomous Agents. Springer-Verlag, Berlin.

[10] Steels, L. and R. Brooks, eds. (1995) Building Situated Embodied Agents. The Alife route to AI. Lawrence Erlbaum Assoc., New Haven.

[11] Steels, L. (1996a) Self-organizing vocabularies. In: Langton, C. (ed.) Artificial Life V. Nara, Japan. Addison-Wesley Pub. Co. Redwood City, Ca.

[12] Steels, L. (1996b) A self-organizing spatial vocabulary. Artificial Life Journal, **3**, 1996.

[13] Steels, L. (1996c) Discovering the competitors. Journal of Adaptive Behavior. **4**(2), 1996.

[14] Steels, L. (1996d) The origins of intelligence. In: Proceedings of the 1995 Carlo Erba Meeting on Artificial Life. Fondazione Carlo Erba, Milano.

[15] Steels, L. (1996) Perceptually grounded meaning creation. In: Tokoro, M. (ed.) (1996) Proceedings of ICMAS-96, Kyoto 1996. Calif, AAAI Press.

[16] Steels, L., (1997a) Language Learning and Language Contact. In: Daelemans, W., Van den Bosch, A., and Weijters, A., editors (1997). Workshop Notes of the ECML/MLnet Familiarization Workshop on Empirical Learning of Natural Language Processing Tasks. Prague pp. 11-24.

[17] Steels, L. (1997b) Constructing and Sharing Perceptual Distinctions. In: van Someren, M. and G. Widmer (eds.) (1997) Proceedings of the European Conference on Machine Learning. Springer-Verlag, Berlin.

[18] Steels, L. (1997b) The Origins of Syntax in Visually Grounded Robotic Agents. In: Pollack, M. (1997) Proceedings of the IJCAI-97 Conference. The AAAI Press, Morgan Kaufmann Pub. Los Angeles.

[19] Steels, L. (1997d) The synthetic modeling of language origins. *Evolution of Communication*, 1(1):1-35. Walter Benjamins, Amsterdam.

[20] 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.

[21] Wittgenstein, L. (1974) Philosophical Investigations. Translated by G. Anscombe. Basil Blackwell, Oxford.