3

Bootstrapping grounded word semantics

Luc Steels
Sony Computer Science Laboratory
Paris and VUB AI Laboratory Brussels
and
Frederic Kaplan
Sony Computer Science Laboratory
Paris

3.1 Introduction

The paper reports on experiments with a population of visually grounded robotic agents capable of bootstrapping their own ontology and shared lexicon without prior design nor other forms of human intervention. The agents do so while playing a particular language game called the guessing game. We show that synonymy and ambiguity arise as emergent properties in the lexicon, due to the situated grounded character of the agent–environment interaction, but that there are also tendencies to dampen them so as to make the language more coherent and thus more optimal from the viewpoints of communicative success, cognitive complexity, and learnability.

How do words get their meanings? An answer to this question requires a theory of the origins of meanings, a theory of how forms get recruited for expressing meanings, and a theory of how associations between forms and meanings may propagate in a population. Each theory must characterize properties of a cognitive agent's architecture: components a cognitive agent needs to have, and details of how the different components coordinate their activities. More specifically, the theories should detail what kind of associative memory the agents must have for storing and acquiring form–meaning relations, what type of mechanisms they might use to categorize the environment through sensory inputs, how they might acquire a repertoire of perceptually grounded categories (an ontology), and what behaviors the agents must be capable of so as to communicate successfully through language.

To allow validation, theories of agent architecture should be formally specified and hence testable through computer simulations or even better through experiments with robotic agents interacting with real world environments through a sensory apparatus. When shared lexicons and ontologies emerge from these experiments, the architecture is at least functionally adequate. Ideally other emergent properties also seen in the evolution of natural languages should be observed, such as the damping of synonymy, or an expansion of the language when the environment introduces new challenges and thus the creation of new meanings. Experiments with physical robots ensure that real world constraints are not ignored or overlooked.

In the last five years, substantial progress has been reported on these objectives (see the overview in Steels, 1997b). There has been a first wave of research in the early 1990's strongly inspired by artificial life concepts (Maclennan, 1991; Werner and Dyer, 1991; Hurford, 1989; Noble, 1998; Yanco Stein, 1993). This early research has often used a genetic approach and assumed that the set of meanings is fixed and given a priori by the designer. The primary emphasis was on understanding the emergence and evolution of animal communication rather than human natural language. There has been a second wave of research in the mid 1990's featuring more systematic investigations of different possible architectures (Oliphant, 1996) and a deeper study of the complex adaptive system properties of linguistic populations and their evolving lexicons (Steels and Kaplan, 1998; Arita and Koyama, 1998). The issue of meaning creation in co-evolution with lexicon formation has also been studied (Steels, 1997b; Hutchins and Hazelhurst, 1995) and more sophisticated experiments have been reported to ground lexicon formation on real robots (Steels and Vogt, 1997). In this second wave of research, the cultural approach dominates. Lexicons are no longer transmitted genetically but by learning. There is a growing interest to model the complex phenomena seen in human lexicon formation. In several experiments the set of possible meanings is open, expanding and contracting in relation to the demands of the task and the environment. The population of agents is also open so that issues of lexicon acquisition by virgin agents and preservation of a system across generations can be studied.

This paper builds further on these various research results focusing more specifically on two issues:

[1] *The Gavagai problem.* In *Word and Object*, Quine raises the question how a linguist might acquire a language of a foreign tribe (Quine, 1960:29). He points out that if a native says *Gavagai*, while pointing to a rabbit scurrying by, it is in no way possible to uniquely determine

its meaning. *Gavagai* could mean 'rabbit', 'animal', 'white', as well as hundreds of other things. So, how can one ever acquire the meaning of a word? As pointed out by Eve Clark (1993), children have a very similar problem and it is therefore not surprising that overextensions or underextensions are observed in the first words. In computer simulations so far, most researchers have assumed that agents have direct access to each other's meanings, so that the problem of lexicon construction and acquisition becomes one of learning associations between words and meanings, with direct feedback on whether the right association has been learned. But in more realistic circumstances, humans as well as autonomous agents only get feedback on the communicative success of an interaction, not on what meanings were used. Even communicative success may not be completely clear although this additional source of uncertainty has not been explored further in our experiments. The problem of lexicon construction and acquisition must therefore be reformulated: The agents must acquire word–meaning and meaning–object relations which are compatible with the word–object co-occurrences they overtly observe but without observing word–meaning relations directly. This problem is obviously much harder.

[2] *The grounding problem.* When agents are embodied and situated, and when they have to build up their ontology autonomously from scratch, the problem of lexicon acquisition becomes even harder. The perception of an agent depends on the viewpoint from which he observes the scene and categorization therefore becomes dependent on this viewpoint. For example, something which is to the left for one agent may be to the right for another one and vice-versa. Even the colors of a surface (more precisely wavelength reflection) are not perceptually constant when seen from slightly different angles. This perceptual incoherence makes it more difficult for agents to coordinate their words and meanings. Nevertheless, the success of a language depends to a great extent on whether the agents manage to abstract away from contingencies of viewpoints and situations.

In the past few years, we have designed an agent architecture addressing these two issues and have tested this architecture on physically embodied robotic agents. Although we have studied multi-word expressions and the emergence of syntax within the same experimental context (Steels, 1998), this paper only discusses single word utterances so that we can focus completely on semiotic dynamics. The first section of this paper briefly summarizes the experimental set up and the proposed

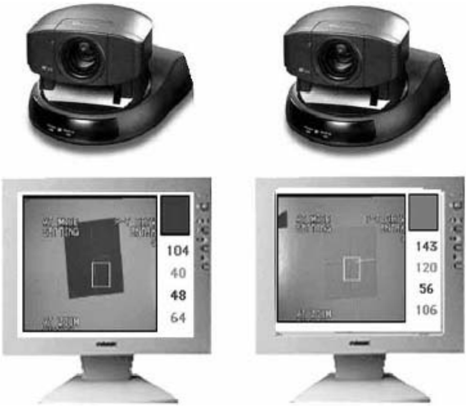

figure 3.1. Two Talking Head cameras and associated monitors showing what each camera perceives.

architecture. The rest of the paper focuses on the macroscopic properties of the lexicons and ontologies that emerge from our robotic experiments.

3.2 The Talking Heads experiment

The robotic setup used for the experiments in this paper consists of a set of 'Talking Heads' connected through the Internet. Each Talking Head features a Sony EVI-D31 camera with controllable pan/tilt motors for horizontal and vertical movement (figure 3.1), a computer for cognitive processing (perception, categorization, lexicon lookup, etc.), a screen on which the internal states of the agent currently loaded in the body are shown, a TV-monitor showing the scene as seen through the camera, and devices for audio input and output. Agents can load themselves in a physical Talking Head and teleport themselves to another Head by traveling through the Internet. By design, an agent can only interact with another one when it is physically instantiated in a body located in a shared physical environment. The experimental infrastructure also features a commentator which reports and comments on dialogs, displays measures of the ontologies and languages of the agents and game statistics, such as average communicative success, lexical coherence, average ontology and lexicon size, etc.

For the experiments reported in this paper, the shared environment consists of a magnetic white board on which various shapes are pasted: colored triangles, circles, rectangles, etc. Although this may seem a strong restriction, we have learned that the environment should be simple enough to be able to follow and experimentally investigate the complex dynamics taking place in the agent population.

The guessing game

The interaction between the agents consists of a language game, called the guessing game. The guessing game is played between two visually grounded agents. One agent plays the role of *speaker* and the other one then plays the role of *hearer*. Agents take turns playing games so all of them develop the capacity to be speaker or hearer. Agents are capable of segmenting the image perceived through the camera into objects and of collecting various sensory data about each object, such as the color (decomposed in RGB channels), average gray-scale or position. The set of objects and their data constitute a *context*. The speaker chooses one object from this context, further called the *topic*. The other objects form the *background*. The speaker then gives a linguistic hint to the hearer.

The linguistic hint is an utterance that identifies the topic with respect to the objects in the background. For example, if the context contains [1] a red square, [2] a blue triangle, and [3] a green circle, then the speaker may say something like *the red one* to communicate that [1] is the topic. If the context contains also a red triangle, he has to be more precise and say something like *the red square*. Of course, the Talking Heads do not say *the red square* but use their own language and concepts which are never going to be the same as those used in English. For example, they may say *malewina* to mean [UPPER EXTREME-LEFT LOW-REDNESS].

Based on the linguistic hint, the hearer tries to guess what topic the speaker has chosen, and he communicates his choice to the speaker by pointing to the object. A robot points by transmitting in which direction he is looking in his own agent-centered coordinates. The other robot is calibrated in the beginning of the experiment to be able to convert these coordinates into his own agent-centered coordinates. The game succeeds if the topic guessed by the hearer is equal to the topic chosen by the speaker. The game fails if the guess was wrong or if the speaker or the hearer failed at some earlier point in the game. In case of a failure, the speaker gives an extra-linguistic hint by pointing to the topic he had in mind, and both agents try to repair their internal structures to be more successful in future games.

The architecture of the agents has two components: a conceptualization module responsible for categorizing reality or for applying categories to find back the referent in the perceptual image, and a verbalization module responsible for verbalising a conceptualization or for interpreting a form to reconstruct its meaning. Agents start with no prior designer-supplied ontology nor lexicon. A shared ontology and lexicon must emerge from scratch in a self-organized process. The agents therefore not only play the game but also expand or adapt their ontology or lexicon to be more successful in future games.

The conceptualization module

Meanings are categories that distinguish the topic from the other objects in the context. The categories are organized in discrimination trees (figure 3.2) where each node contains a discriminator able to filter the set of objects into a subset that satisfies a category and another one that satisfies its opposition. For example, there might be a discriminator based on the horizontal position (HPOS) of the center of an object (scaled between 0.0 and 1.0) sorting the objects in the context in a bin for the category 'left' when HPOS $<$ 0.5, (further labeled as [HPOS-0.0,0.5]) and one for 'right' when HPOS $>$ 0.5 (labeled as [HPOS-0.5,1.0]). Further subcategories are created by restricting the region of each category. For example, the category 'very left' (or [HPOS-0.0,0.25]) applies when an object's HPOS value is in the region [0.0,0.25]. For the experiments in this paper, the agents have only channels for horizontal position (HPOS), vertical position (VPOS), color (RGB indicated as RED, GREEN, BLUE), and grayscale (GRAY). The system is open to exploit any channel with additional raw data, such as audio, or results from more complex image processing.

A distinctive category set is found by filtering the objects in the context from the top in each discrimination tree until there is a bin which only contains the topic. This means that only the topic falls within the category associated with that bin, and so this category uniquely filters out the topic from all the other objects in the scene. Often more than one solution is possible, but all solutions are passed on to the lexicon module.

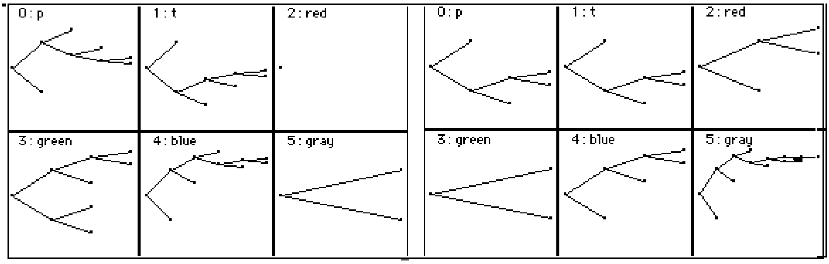

figure 3.2. The discrimination trees of two agents.

The discrimination trees of each agent are formed using a growth and pruning dynamics coupled to the environment, which creates an ecology of distinctions. Discrimination trees grow randomly by the addition of new categorizers splitting the region of existing categories. Categorizers compete in each guessing game. The use and success of a categorizer is monitored and categorizers that are irrelevant for the environments encountered by the agent are pruned. More details about the discrimination game can be found in Steels (1997a).

Verbalization module

The lexicon of each agents consists of a two-way association between forms (which are individual words) and meanings (which are single categories). Each association has a score. Words are random combinations of syllables, although any set of distinct word symbols could be used. When a speaker needs to verbalize a category, he looks up all possible words associated with that category, orders them and picks the one with the best score for transmission to the hearer. When a hearer needs to interpret a word, he looks up all possible meanings, tests which meanings are applicable in the present context, i.e. which ones yield a possible single referent, and uses the remaining meaning with the highest score as the winner. The topic guessed by the hearer is the referent of this meaning.

Based on feedback on the outcome of the guessing game, the speaker and the hearer update the scores. When the game has succeeded, they increase the score of the winning association and decrease the competitors, thus implementing lateral inhibition. When the game has failed, they each decrease the score of the association they used. Occasionally new associations are stored. A speaker creates a new word when he does not have a word yet for a meaning he wants to express. A hearer may encounter a new word he has never heard before and then store a new association between this word and the best guess of the possible meaning. This guess is based on first guessing the topic using the extra-linguistic hint provided by the speaker, and on performing categorization using his own discrimination trees as developed thus far. These lexicon bootstrapping mechanisms have been explained and validated extensively in earlier papers (Steels and Kaplan, 1998) and are basically the same as those reported by Oliphant (1996).

The conceptualization module proposes several solutions to the verbalization module which prefers those that have already been lexicalized. Agents monitor success of categories in the total game and use this to target growth and pruning. The language therefore strongly influences the ontologies agents retain. The two modules are structurally coupled and thus get coordinated without a central coordinator.

Examples

Here is the simplest possible case of a language game. The speaker, **a1**, has picked a triangular object at the bottom of the scene as the topic. There is only one other rectangular object in the scene, nearer to the top. Consequently, the category [VPOS-0.0,0.5]$_{\mathbf{a1}}$, which is valid when the vertical position VPOS < 0.5, is applicable because it is valid for the triangle but not for the rectangle. Assuming that **a1** has an association in his lexicon relating [VPOS-0.0,0.5]$_{\mathbf{a1}}$ with the word *lu*, then **a1** will retrieve this association and transmits the word *lu* to the hearer, which is agent **a2**.

Now suppose that **a2** has stored in his lexicon an association between “lu” and [RED-0.0,0.5]$_{\mathbf{a2}}$. He therefore hypothesises that [RED-0.0,0.5]$_{\mathbf{a2}}$ must be the meaning of *lu*. When he applies this category to the present scene, in other words when he filters out the objects whose value for the redness channel (RED) do not fall in the region $[0.0, 0.5]$, he obtains only one remaining object, the triangle. Hence **a2** concludes that this must be the topic and points to it. The speaker recognises that the hearer has pointed to the right object and so the game succeeds.

The complete dialog is reported by the commentator as follows:

```
Game 125.
 a1 is the speaker. a2 is the hearer.
 a1 segments the context into 2 objects
 a1 categorizes the topic as [VPOS-0.0,0.5]
 a1 says:  ‘lu’
 a2 interprets ‘lu’ as [RED-0.0,0.5]
 a2 points to the topic
 a1 says: ‘OK’
```

This game illustrates a situation where the speaker and the hearer picks out the same referent even though they use a different meaning. The speaker uses vertical position and the hearer the degree of redness in RGB space.

Here is a second example, The speaker is again **a1** and he uses the same category and the same word *lu*. But the hearer, **a3**, interprets *lu* in terms of horizontal position [HPOS-0.0,0.5]$_{\mathbf{a3}}$ (left of the scene). Because there is more than one object satisfying this category in the scene the agents look at, the hearer is confused. The speaker then points to the topic and the hearer acquires a new association between *lu* and [VPOS-0.0,0.5]$_{\mathbf{a3}}$, which starts to compete with the one he already had. The commentator reports this kind of interaction as follows:

```
Game 137.
 a1 is the speaker. a3 is the hearer.
 a1 segments the context into 2 objects
 a1 categorizes the topic as [VPOS-0.0,0.5]
 a1 says:  ‘lu’
 a3 interprets ‘lu’ as [HPOS-0.0,0.5]
 There is more than one such object
 a3 says: ‘lu?’
 a1 points to the topic
 a3 categorizes the topic as [VPOS-0.0,0.5]
 a3 stores ‘lu’ as [VPOS-0.0,0.5]
```

Table 3.1 shows part of a vocabulary of a single agent after 3,000 language games. It shows also the score (Sc.). We see in this table that for some meanings (such as [RED-0.0,0.125]) a single form *wovota* has firmly established itself. For other meanings, like [GRAY-0.25,0.5], a word was known at some point but is now no longer in use. For other meanings, like [VPOS-0.0,0.5], two words are still competing: *gorepe* and *zuga*. There are words, like *zafe*, which have two possible meanings [VPOS-0.0,0.25] and [GREEN-0.5,1.0].

Table 3.1. *Agent vocabulary*

Form	*Meaning*	*Sc.*	*Form*	*Meaning*	*Sc.*
wovota	[RED-0.0,0.125]	1.0	sogavo	[GREEN-0.5,1.0]	0.0
tu	[GRAY-0.25,0.5]	0.0	naxesi	[GREEN-0.5,1.0]	0.0
gorepe	[VPOS-0.0,0.5]	0.3	ko	[GREEN-0.5,1.0]	0.0
zuga	[VPOS-0.0,0.5]	0.1	ve	[GREEN-0.5,1.0]	0.0
lora	[VPOS-0.25,0.5]	0.1	migine	[GREEN-0.5,1.0]	0.0
wovota	[VPOS-0.25,0.5]	0.2	zota	[GREEN-0.5,1.0]	0.9
di	[VPOS-0.25,0.5]	0.0	zafe	[GREEN-0.5,1.0]	0.1
zafe	[VPOS-0.0,0.25]	0.2	zulebo	[HPOS-0.0,1.0]	0.0
wowore	[VPOS-0.0,0.25]	0.9	xi	[HPOS-0.0,1.0]	0.0
mifo	[HPOS-0.0,1.0]	1.0			

3.3 Tendencies in natural language

Clearly, to have success in the game the speaker and the hearer must share a list of words, and the meanings of these words must pick out the same referent in the same context. However agents can only coordinate their language based on overt behavior. This leads to various forms of incoherence. An incoherence remains until the environment produces situations that cause further disentanglement, as in the example above where a speaker uses a word which is interpreted by the hearer as referring to more than one object instead of just one.

There are clear tendencies in natural languages towards coherence and indeed a coherent language is ‘better’. First of all coherence gives a higher chance of success in multiple contexts. For example, if every agent preferentially associates the same meaning with the same word, there is a higher chance that the same word will designate the same referent, even in a context that has not been seen before. Second, coherence diminishes cognitive complexity. For example, if all agents preferentially use the same word for the same meaning, there will be fewer words and therefore less words need to be stored. If all words preferentially have the same meaning, there is less cognitive effort needed in disambiguation. Third, coherence helps in language acquisition by future generations. If there are fewer words and they tend to have the same meanings, a language learner has an easier time to acquire them.

Natural languages are clearly not totally coherent even in the same language community, and languages developed autonomously by physically embodied agents will not be fully coherent either.

1. Different agents may prefer a different word for the same meaning. These words are said to be *synonyms* of each other. An example is “pavement” versus “sidewalk”. The situation arises because an agent may construct a new word not knowing that one is already in existence. Synonymy is often an intermediate stage for new meanings whose lexicalization has not stabilized yet. Natural languages show a clear tendency for the elimination of synonyms. Accidental synonyms tend to specialize, incorporating different shades of meaning from the context or reflecting socio-linguistic and dialectal differences of speaker and hearer.
2. The same word may have different preferred meanings in the population. These words thus become *ambiguous*. This situation may arise completely accidentally, as in the case of *bank* which can mean river bank and financial institution. These words are then called *homonyms*. The situation may also arise whenever there is more than one possible meaning compatible with the same situation. An agent on hearing an unknown word may therefore incorrectly guess its meaning. Ambiguity also arises because most words are *polysemous*: The original source meaning has become extended by metaphor and metonymy to cover a family of meanings (Victorri and Fuchs, 1996). Real ambiguity tends to survive in natural languages only when the contexts of each meaning are sufficiently different, otherwise the hearer would be unable to derive the correct meaning.
3. The same meaning may denote different referents for different agents *in the same context*. This is the case when the application of a category is strongly situated, for example ‘left’ for the speaker may be ‘right’ for the hearer. Deictic terms like *this* and *that* are even clearer examples from natural language. In natural languages, this *multi-referentiality* is counter-acted by verbalizing more information about the context or by avoiding words with multi-referential meanings when they may cause confusion.
4. It is possible and very common with a richer categorial repertoire, that a particular referent in a particular context can be conceptualized in more than one way. For example, an object may be to the left of all the others, *and* much higher positioned than all the others. In the same situation different agents may therefore use different meanings. Agents only get feedback about whether they guessed the object

the speaker had in mind, not whether they used the same meaning as the speaker. This *indeterminacy* of categories is a cause of ambiguity. A speaker may mean ‘left’ by *bovubo*, but a hearer may have inferred that it meant ‘upper’.

So, although circumstances cause agents to introduce incoherence in the language system, there are at the same time opposing tendencies, attempting to restore coherence. Synonyms tend to disappear and ambiguity is avoided. In the remainder of this paper, we want to show that the dynamics of the guessing game, particularly when it is played by situated embodied robotic agents, leads unavoidably to incoherence, but that there are tendencies towards coherence as well. Both tendencies are emergent properties of the dynamics. There is no central controlling agency that weeds out synonyms or eliminates ambiguity, rather they get pushed out as a side effect of the collective dynamics of the game. Before we can see whether all this is indeed the case we need a set of analysis tools.

3.4 Analysis tools

Semiotic landscapes

We propose the notion of a *semiotic landscape* (which we also call RMF-landscape) to analyse grounded semiotic dynamics. The semiotic landscape is a graph, in which the nodes in the landscape are formed by referents, meanings and forms, and there are links if the items associated with two nodes indeed co-occur (figure 3.3). The relations are labeled RM for referent to meaning, MR for meaning to referent, RF for referent to form, FR for form to referent, and FM for form to meaning and MF for meaning to form. For real world environments, the set of possible referents is infinite, so the semiotic landscape is infinite. However, for purposes of analysis, we can restrict the possible environments and thus the possible referents artificially and then study the semiotic dynamics very precisely. This is what we will do in the remainder of the paper.

In the case of a perfectly coherent communication system, the semiotic landscape consists of unconnected triangles. Each referent has a unique meaning, each meaning has a unique form, and each form a unique referent. Otherwise more complex networks appear. The RMF-landscape in figure 3.3 contains an example where the agents use two

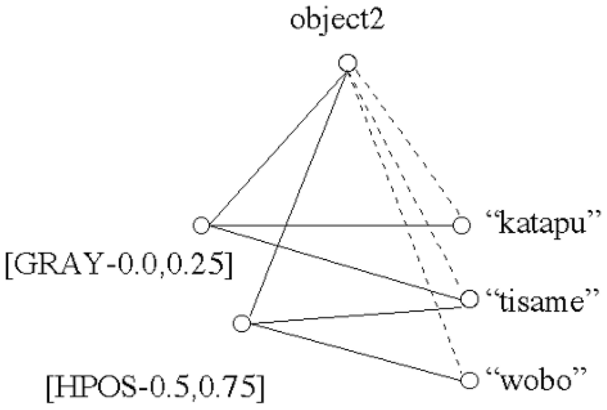

figure 3.3. A semiotic landscape represents the co-occurrences between referents, meanings and forms.

possible meanings for denoting **object2** namely [GRAY-0.0,0.25] (very light) and [HPOS-0.5,0.75] (lower upper), and the words *katapu* and *tisame* for [GRAY-0.0,0.25] and *wobo* and *tisame*, for [HPOS-0.5,0.75]. Each meaning has therefore two synonyms and *tisame* is ambiguous; it can mean both [GRAY-0.0,0.25] and [HPOS-0.5,0.75]. Three words are used to refer to *object2*. This kind of situation is typical in certain stages of our experiments and complexity rapidly increases when the same meaning is also used to denote other referents (which is obviously very common and indeed desirable).

As mentioned earlier, incoherence is not necessarily decreasing the communicative success of the language. The RMF-landscape in figure 3.3 still leads to total success in communication whenever both meanings are equally adequate for picking out the referent. Even if a speaker uses *tisame* to mean [GRAY-0.0,0.25] and the hearer understands *tisame* to mean [HPOS-0.5,0.75], they still have communicative success. The goal of the language game is to find the referent. It does not matter whether the meanings are the same. The agents cannot even know which meaning the other one uses because they have no access to each other's internal states.

Measuring coherence

The degree of coherence of a language can be measured by observing the actual linguistic behavior of the agents while they play language games, more specifically, by collecting data on the frequency of co-occurrence of items such as the possible forms used with a certain referent or all the possible meanings used with a certain form. Frequency of co-occurrence will be represented in competition diagrams, such as the RF-diagram in figure 3.8, which plots the evolution of the frequency of use of the Referent–Form relations for a given referent in a series of games. Similar diagrams can be made for the FR, FM, MF, RM and MR relations.

One co-occurrence relation for a particular item will be most frequent, and this is taken as the dominating relation along that dimension. The average frequency of the dominating relations along a particular dimension is an indication of how coherent the community's language system is along that dimension. For example, suppose we want to know the coherence along the meaning–form dimension, in other words whether there are many synonyms in the language or not. For a given series of games, we calculate for each meaning that was indeed used somewhere in the series, the frequency of the most common form for that meaning. Then we take the average of these frequencies and this represents the MF-coherence. If all meanings had only one form the MF-coherence is equal to 1.0. If two forms were used for the same meaning with equal frequency, it will be 0.5. When plotting the MF-coherence, we can therefore track tendencies towards an increase or decrease of synonyms.

3.5 Global evolution of coherence

We are now ready to study the semiotic dynamics of the guessing game, as played by situated embodied agents interacting in a shared physical environment. We typically start with a limited set of objects (for example four) that allow agents to play four different games, each object being in turn the topic. Then we progressively add new objects to the environment (by pasting new objects on the white board or moving them around) and study the impact on the lexicon and ontologies of the agents. Figure 3.4 shows the result of such an experiment involving 5 agents. It shows the progressive increase in environmental complexity and the average success in the game. We see clearly that the agents manage to bootstrap a successful lexicon from scratch. Success then

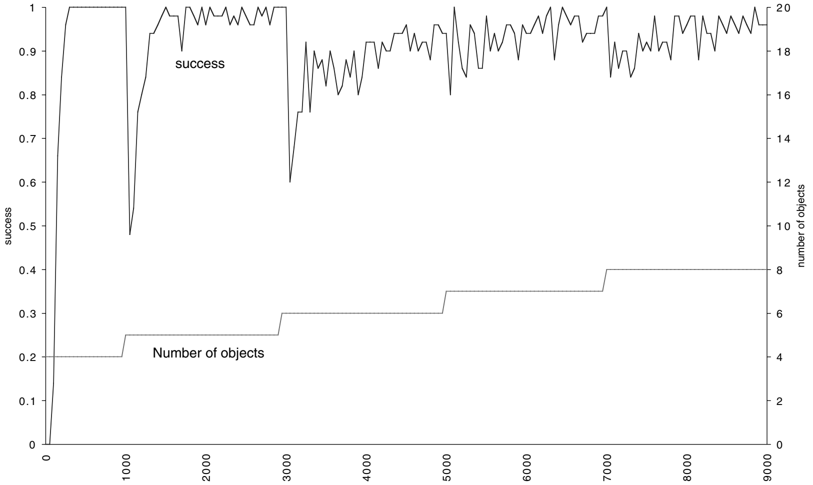

figure 3.4. The graph shows average success and increase in environmental complexity, for a group of 5 embodied agents engaged in the guessing game.

drops every time the environment increases in complexity but regains as the agents invent new words or create new meanings. Progressively it is less and less difficult to cope with expansions of the environment because words are less ambiguous and the repertoire is covering more and more meanings.

Figure 3.5 shows the evolution of the RF and FR complexity measures discussed earlier, for the same series of games as in figure 3.4. In an initial phase, the first 1000 games, the FR and RF relations fixate (even though agents do not necessarily use the same meaning to establish the relation). This is expected because agents get immediate feedback on this relation. The same word is used for the same referent and the same referent for the same word. As the complexity of the environment increases, this breaks down because many different words can be used for the same referent. Each of these words has another possibly appropriate meaning. This graph therefore shows that the lexicon becomes more general.

Figure 3.6 shows the evolution of the RM/MR and FM/MF complexity for the same series of games. We see very clearly that both the meaning–form and the form–meaning coherence increases, particularly after the initial period when ambiguities have been cleared from the language. The MR and RM co-occurrence is an indication of the degree of multi-referentiality and indeterminacy of categories. The same mean-

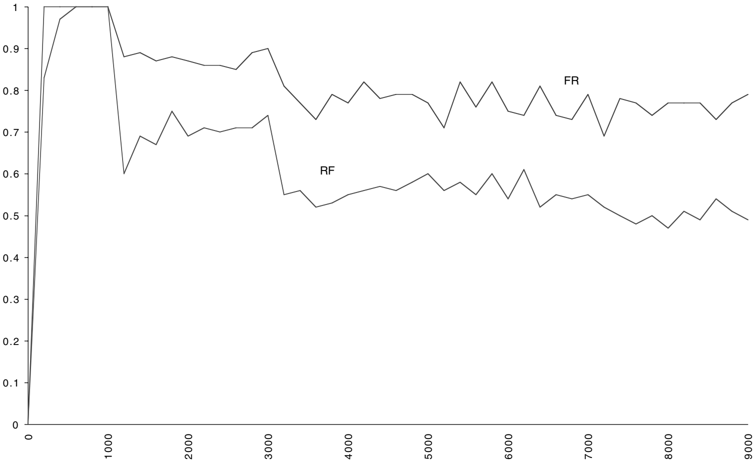

figure 3.5. The graph shows the coherence of the co-occurrence relation between referents and forms (RF), and forms and referents (FR).

ing co-occurs with different referents and the same referent is designated by different meanings.

3.6 Damping synonymy and ambiguity

We now inspect in more detail the kind of semiotic dynamics that is observed in the very beginning, when the agents do not have a lexicon nor ontology yet. Figure 3.7 shows a series of 5000 games played by a group of 20 agents. The first tendency that we see clearly is the damping of synonymy, by a winner-take-all process deciding on which form to use for a particular referent. This is shown in the RF-diagram displayed in figure 3.8, for the series of games displayed in figure 3.7, which shows that one word *va* comes to dominate for expressing this one referent. This damping is expected because the agents get explicit feedback about the RF relation and there is lateral inhibition as well as a positive feedback loop between use and success.

When we inspect the different meanings of *va*, through the FM-diagram (figure 3.9), we clearly see that even after 3000 games ambiguity stays in the language. Three stable meanings for *va* have emerged: [RED-0,0.125], [BLUE-0.3125,0.375], and [VPOS-0.25,0.5]. They are all equally good for distinguishing the topic *va* designates, and there are no situations yet that would have allowed disambiguation.

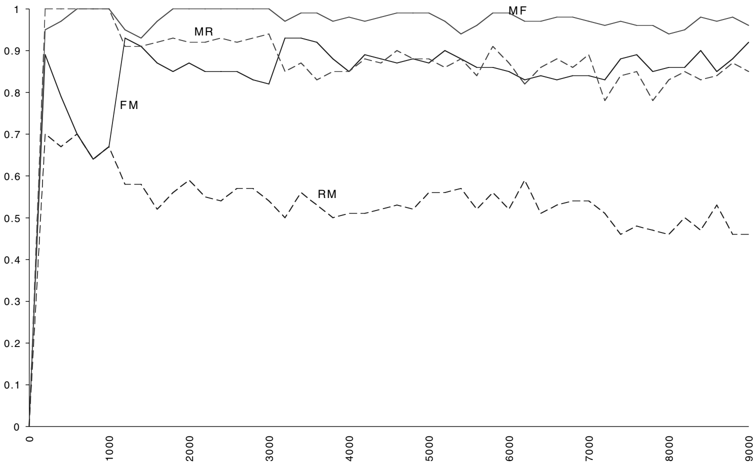

figure 3.6. The graph shows the coherence of the co-occurrence relation between referents and meanings (RM), meanings and referents (MR), meanings and forms (MF) and forms and meanings (FM) for the same series as in figure 3.6.

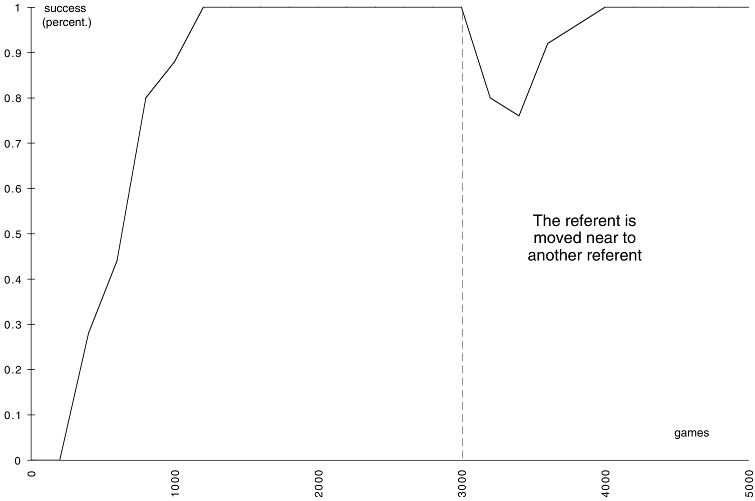

figure 3.7. This graph shows the average success per 200 games in a series of 5000 games played by 20 agents. The agents evolve towards total success in their communication after about 1000 games. A change in the environment induced after 3000 games gives a decrease in average success which rebounds quickly.

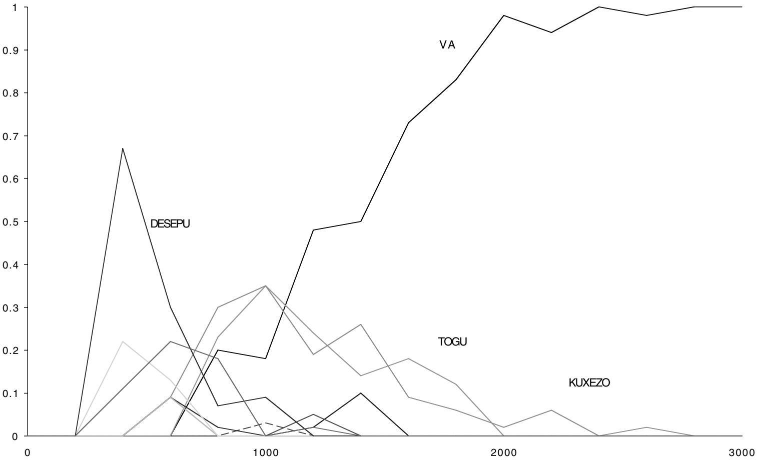

figure 3.8. This RF-diagram shows the frequency of all forms used for the same referent in 3000 language games.

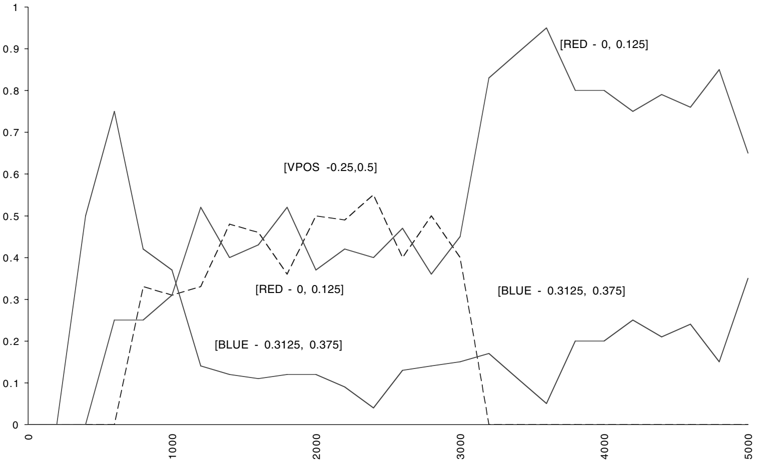

figure 3.9. This FM-diagram shows the frequency of each form–meaning co-occurrence for "va" in a series of 5000 games. A disambiguating situation occurs in game 3000 causing the loss of one meaning of *va*.

In game 3000, the environment produces a scene in which a category which was distinctive for the object designated by *va* is no longer distinctive. More precisely, we, as experimenters, have moved the object very close to another object so that the position is no longer distinctive.

Figure 3.7 shows that success drops (meaning there have been some failures in the game), but that it rebounds quickly. Failures occur because *va* no longer picks out the right object for those who believe that it means [VPOS-0.25,0.5], so they have to shift to an alternative meaning for *va*, compatible with the new situation. The FM-diagram in figure 3.9 shows that the positional meaning of *va* (namely [VPOS-0.25,0.5]) has disappeared. The other meanings, based on color, are still possible because they are not affected when the object designated by *va* moved its position.

3.7 Conclusions

The approach we have used for lexicon formation is different from the more traditional Quinean approach. Quine assumes that agents learn the meaning of words by making progressive inductive abstraction from the situations in which they observe a particular word-object relation. The common properties of referents constitute the meaning of the word and these common properties are learned by seeing many examples and retaining their similarities. Such an approach also underlies neural network approaches to word acquisition as discussed for example by Hutchins and Hazelhurst (1995), or Regier (1995). In contrast, we have adopted a Wittgensteinian approach, where agents invent words and meanings as part of a language game, formulate different hypotheses about the meanings of words used by others and test these meanings in their own language production. The evolution towards lexicon coherence in the population (where one word has one dominant meaning and one meaning has one dominant word) is a collective phenomenon triggered by responses of the system to new situations in which the multiple meanings of a word are no longer compatible with each other.

In the literature, intelligent generalization, specialization, or elimination operators are often ascribed to the individuals acquiring language (Clark, 1993). This is not the case in the present paper. Agents have very minimal forms of intelligence. We have observed that an agent sometimes starts to prefer a more general category for a word (or a more specific one), but this is not due to an explicit generalization operation in the agent, it takes place as a side effect of the repair action undertaken after a failing game. Similarly, the phenomenon of blocking (Copestake and Briscoe, 1995), which means that the availability of a specialized word (like *pork*) blocks the use of a more general one (like *pig*), arises as a side effect of the collective dynamics but cannot be ascribed to structure inside the agent.

A large amount of work needs to be done to further study the complex dynamics of lexicon formation. We need better tools to track the semiotic dynamics and study the interaction between lexicon and meaning formation. Nevertheless, we believe that the results discussed in this paper constitute a major step forward, because we have shown for the first time the evolution of an open-ended set of meanings and words by a group of autonomous distributed agents in interaction with physical environments through their sensory apparatus. The most valuable result of our experiments is that we are able to demonstrate a scale up. New meanings arise when the environment becomes more complex and the lexicon keeps adapting and expanding to sustain successful communication.

Acknowledgement

This research was conducted at the Sony Computer Science Laboratory. We are strongly indebted to Angus McIntyre for creating the Babel tool that is the technical backbone of our experiments and for technical support in the low level vision aspects.

References

Arita, T. and Y. Koyama (1998). Evolution of linguisitic diversity in a simple communication system. In Christopher Adami, *et al.* (1998) *Proceedings of Alife VI.* MIT Press, Cambridge MA. 9–17.

Cangelosi, A. and D. Parisi (1996). The Emergence of a "Language" in an Evolving Population of Neural Networks. *Conference of the Cognitive Science Society*, San Diego.

Clark, E. (1993). *The lexicon in acquisition.* Cambridge University Press, Cambridge.

Copestake, A. A. and E. J Briscoe (1995). Regular polysemy and semi-productive sense extension. *Journal of Semantics.* 12, 15–67.

Hurford, J. (1989) Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua*, 77, 187–222.

Hutchins, E. and B. Hazelhurst (1995). How to invent a lexicon. The development of shared symbols in interaction. In Gilbert, N. and R. Conte (eds.) *Artificial societies: The computer simulation of social life.* UCL Press, London.

Langton, C. (ed.) (1995). *Artificial Life. An overview.* MIT Press, Cambridge MA.

MacLennan, B. (1991). Synthetic Ethology: An approach to the study of communication. In Langton, C., *et al.* (1991) *Artificial Life II.* Addison-

Wesley Pub. Cy, Redwood City CA. pp. 603–631.

Noble, J. (1998). Evolved signals: expensive hype vs. conspirational whispers. In Christopher Adami, *et al.* (1998) *Proceedings of Alife VI.* MIT Press, Cambridge MA. pp. 358–367.

Oliphant, M. (1996). The dilemma of Saussurean communication. *Biosystems*, 37(1-2), 31–38.

Quine, W. (1960). *Word and Object.* MIT Press, Cambridge MA.

Regier, T. (1995). A model of the human capacity for categorizing spatial relations. *Cognitive Linguistics*, 6-1, 63–88.

Steels, L. (1996). Emergent adaptive lexicons. In Maes, P., M. Mataric, J-A. Meyer, J. Pollack, and S. Wilson, (eds.) (1996) *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior.* MIT Press, Cambridge, MA.

Steels, L. (1997a). 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.

Steels, L. (1997). The synthetic modeling of language origins. *Evolution of Communication*, 1(1), 1–35.

Steels, L. (1998). The origins of syntax in visually grounded robotic agents. *Artificial Intelligence* 103, 1–24.

Steels, L. and F. Kaplan (1998). Spontaneous lexicon change, In *Proceedings of COLING-ACL*, Montreal, August 1998, 1243–1249.

Steels, L. and P. Vogt (1997). Grounding adaptive language games in robotic agents. In Harvey, I. *et al.* (eds.) *Proceedings of ECAL 97*, Brighton UK, July 1997. MIT Press, Cambridge MA.

Victorri, B. and C. Fuchs. (1996). *La polysemie. Construction dynamique du sens.* Hermes, Paris.

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., 659–687.

Yanco, H. and L. Stein (1993). An adaptive communication protocol for cooperating mobile robots. In: Meyer, J-A, H. L. Roitblat, and S. Wilson (1993) *From Animals to Animats 2. Proceedings of the Second International Conference on Simulation of Adaptive Behavior.* MIT Press, Cambridge MA., pp. 478–485.