Self-Organisation of a Lexicon in Embodied Agents

Joris Van Looveren (joris@arti.vub.ac.be)
Artificial Intelligence Laboratory, Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussel, Belgium

Abstract

The Talking Heads experiment is a distributed, embodied version of earlier experiments on lexicon organisation. A population of agents capable of producing and interpreting one-word utterances are allowed to interact in an environment (a whiteboard which they observe using pan-tilt cameras) containing objects that form the topics of the interactions between the agents. The paper describes observations made from data gathered during a four-month run involving three Talking Heads sites permanently connected with a central server through the Internet, and compares them with observations made in simulation experiments.

Introduction

The core function of a lexicon is to provide a mapping between words and their referents. Referents can be concrete objects, properties of an object or a set of objects, abstract concepts, etc. The relationship between a word and its referent can be quite intricate: figure 1 shows the so-called semiotic square [Ogden and Richards, 1935] which clarifies the relation between a word and a concrete object to which it refers. Clearly, this relationship is not direct. A language user perceives the world through its body, and is limited by the sensory modalities it has; and even then, depending on the intentions of the speaker, the actual meaning it lexicalises may stress a different aspect of the perception.

An important question is then how these complex relationships can emerge or be learned. Most of the experiments on lexicon formation that have been done up to now were done in simulation; there the restrictions of a body are less apparent and pervasive as the experimenter can explicitly design the representation of objects and meanings, and thus also what the relationship between words and referents will look like. This paper will describe an experiment in which a mechanism for lexicon creation and learning (which has been tested in simulations before) has been implemented on robots that live in a real-world environment, using the sensors present in the robotic body.

The Synthetic Approach

Language is a social phenomenon: it exists by virtue of a community of individuals that are each capable of speaking and understanding it. This fact suggests that an important assumption that is often made in language studies, the assumption of an isolated "ideal speaker," is incomplete: a full theory of the origins of language should take into account the population aspect, and the dynamics that interactions between agents create.

Language is also not a centrally controlled phenomenon; in a language community there is no single entity or individual that decides how the members of the community should speak. Every individual can have an influence on the other individuals' language, e.g. by coining a new word, which may or may not spread in the community depending on whether the other individuals decide to adopt the innovation.

An alternative way of constructing theories about language that takes into account the collective and distributed aspects of language and that permits one to examine in detail the influence of the dynamics of the population on a language, is by modeling a community of language-capable individuals and analysing the model and its results. Three different types of models will be discussed below.

Mathematical models

Formal, mathematical models make it possible to prove properties and derive laws about the particular phenomenon that one researches, such as upper and lower bounds for certain properties. This is an attractive property, because it yields results that can both be used to evaluate other models, and that can be tested against real data to evaluate the model itself. One should be careful though to avoid making oversimplifications and unrealistic assumptions in order not to make the model too complex.

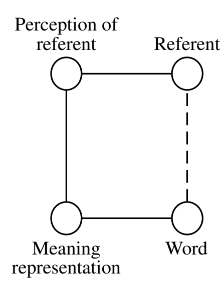

Figure 1: The semiotic square

An example of a mathematical model in the domain of language evolution is the model by Nowak, Komarova and Niyogi [Nowak et al., 2001], in which they present a framework for studying the dynamics of language learning, and a study of the competition between different possible universal grammars.

Simulation Models

Simulation models implement a community of language users in a computer program. The community inhabits a simulated environment; each individual in the community is capable of perceiving the environment, and can produce and interpret utterances.

Most models are of this type, because simulations have many advantages. They are relatively easy to construct, can easily be tuned to experiment with slightly different setups to see what the impact is on the results, and the degree of complexity of the model can easily be varied.

Nevertheless, a danger of simulations, as for mathematical models, is that here too it is easy to oversimplify and make unrealistic assumptions, even unconsciously, because programming a simulation forces one to take many decisions about how to implement the details that are not explicitly specified in the theory. Also, the question whether one took into account all relevant aspects of the real world remains open.

Some examples of simulation models in the area of lexicon emergence and acquisition are [Hurford, 1989], who compares different strategies for acquiring a lexicon from examples, and Oliphant's model to study the conditions in which a Saussurean communication system can evolve in a population of agents [Oliphant, 1996]. Steels researches the emergence of a lexicon through self-organisation [Steels, 1996]. Simulation experiments have also been done in the area of grammar, most notably by Batali [Batali, 1998], Kirby [Kirby, 1999], and also by Steels [Steels, 1997]. De Boer has done experiments in the area of phonetics and phonology [de Boer, 1997].

Robotic models

A third way of modeling a theory is by constructing actual robots that implement the theory. These robots live in a real environment, and are subject to the constraints of this real world. In these kinds of experiments, many aspects of the environment (lighting conditions, noise, external interferences, ...) are not any more under the control of the experimenter. This puts stronger demands on the performance of the language model that is built into the robots.

It can be argued that it is possible to model the important aspects of a real environment to an arbitrarily large precision. While this may be true, there are still a number of issues speaking in favour of complementing simulated experiments with real-world robotic experiments. First of all, there is always the question whether one actually took into account all relevant aspects of the real world. Secondly, if the agent implementing the model can accomplish its task even in the much-less-controlled real world, this is a stronger argument for the soundness of the model/theory than any simulation. Thirdly, the more complex a simulation becomes, the more time it takes to both program it and run it. There may be a point where it becomes more practical and economical to resort to a real robotic system instead of a simulated one, even though robotics is also a labour-intensive discipline.

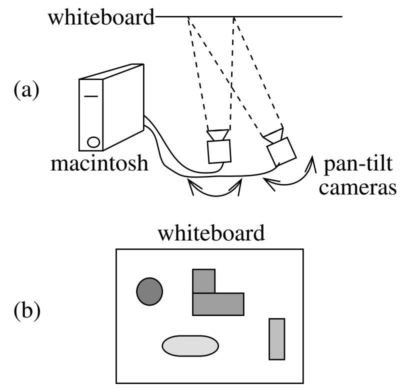

Figure 2: Schema of a Talking Heads installation.

Language experiments in which real robots are used are still relatively rare; an example can be found in [Steels and Vogt, 1997]. Another example, which is described in [Steels and Kaplan, 1999] and is called the “Talking Heads” experiment, is further examined in this paper, and compared with the basic experiment of [Steels, 1996].

Talking Heads

The Experiment

The Talking Heads experiment was a four-month public experiment which was run at two different occasions. The data examined in this paper is from the first experiment, which ran from July 1999 through October 1999. In the experiment, there were 3 permanent sites (in Antwerp, Paris and Brussels) set up as shown in fig. 2 with two pan-tilt cameras looking a whiteboard (a) on which geometrical shapes are pasted (b). Through a public web site, anyone could create new individuals (called *agents*), introduce them into the system, send them to the different sites of the Talking Heads network, and even teach their agents new words.

Although a whiteboard is not exactly an unrestricted environment, many aspects of the experiment are not any more under the (full) control of the experimenter. Lighting conditions change over the course of a day, the scene on the whiteboard may change, people sometimes walk in front of the cameras, calibration is never perfect, etc. Also, since most of the agents were created and controlled by people that were not directly involved in the experiment, there was no control over the movements of the agents in the network.

From this, it is clear that the Talking Heads experiment is much less controlled than a typical simulation experiment. As a first step, we shall examine the system to see if it still exhibits the properties that are observed in the simulated experiments. Additionally, one could study it to see if it exhibits new phenomena not observed in simulation, but this is beyond the scope of this paper.

Interactions

During an interaction, two agents are picked randomly from the population. Each agent gets a camera assigned to it, and processes the image from its camera to retrieve a list of feature-value lists (one for each object discovered). Due to the fact that the cameras are physically in a different location, these images will be slightly different (different zoom, angle, distortion of the shapes, . . .). Also, digitalisation of the images will introduce noise in the colors and dependence on the light conditions.

The data for the objects is then used to play a *guessing game*. In a guessing game, the speaker agent chooses a topic from its object data and tries to find a discriminating description, i.e. a description that describes how the topic is different from the other objects in the environment. When such a description is found, it becomes the meaning that the agent will try to lexicalise, and the best word for it is looked up in the lexicon. (Every word has a score associated to it that reflects its success in previous guessing games. The best word is the word with the highest score that is associated to the meaning.) The word is then uttered by the speaker.

The hearer agent hears the word and observes the same objects as the speaker (allowing for some stochasticity: sometimes, objects that the speaker finds at the edges of its image will not show on the hearer's image, or vice versa). Its task is then to find the topic using only the speaker's linguistic information. In order to do this, it looks up the word in its lexicon, and retrieves the meanings associated with it. It will then apply the meanings to all objects it "sees." If there are meanings for which exactly one object complies to the meaning, these are candidates for being the topic.

If the interaction fails, both agents need to update their lexicons in order to increase the chances for success in the future. The speaker will point to the topic, while the hearer will use this information to infer the meaning of the speaker's word and possibly add a new association to its lexicon. In case of success, both agents increase the score of their respective meaning-word associations to reflect the fact that the interaction was successful.

Results & Discussion

The focus of this study of Talking Heads data has been to see if there are discrepancies in behaviour between the basic, abstract naming game model and the embodied naming games of the Talking Heads, and if so, which they are. The two basic measures used in the simulation experiments have been examined: communicative success and lexicon coherence.

Communicative Success

Communicative success is the simplest possible measure for either model. It measures the fraction of interactions that was successful, i.e. in which the hearer succeeds in finding the topic based solely on linguistic information.

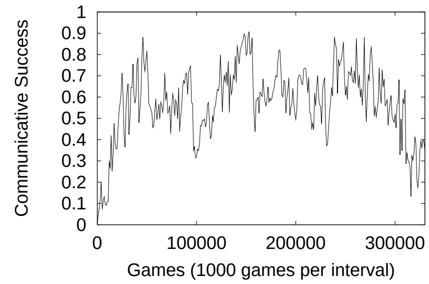

Figure 3: Global communicative success.

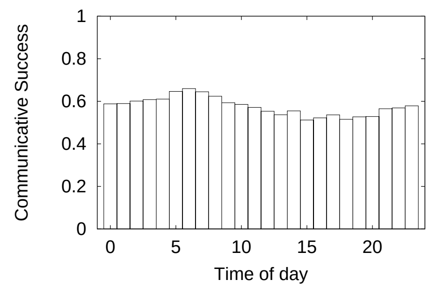

Figure 4: Average percentage of communicative success per hour of the day.

In the basic experiments, communicative success consistently rises quickly to 100%, and stays there. As can be seen in fig. 3, this is not the case for the Talking Heads. On average, communicative success is around 60%, with very large fluctuations. There are several external factors that influence the success of the agent population. In the basic experiment, speakers and hearers are selected randomly from the population, which means that, in a long series of interactions, every agent will interact equally often with every other agent. In the Talking Heads experiment the speaker and hearer are still selected randomly, but the population changes constantly as agents arrive at a site or leave again.

Figure 4 shows the communicative success over the whole experiment, arranged by the hour of the day in which each interaction occurred. The figure shows that the success rate varies with the time of the day; there is about 15% difference between the best hour and the worst hour. The same observation has been made at each individual site, although the average success per site is different (Paris: ±70%; Antwerp: ±55%; Brussels: ±50%).

This is probably due to the fact that during the night, there are less agents arriving at each site, while agents that completed their games transport out to the main server. The population then becomes smaller, so that the probability of an agent present at a certain site to play a

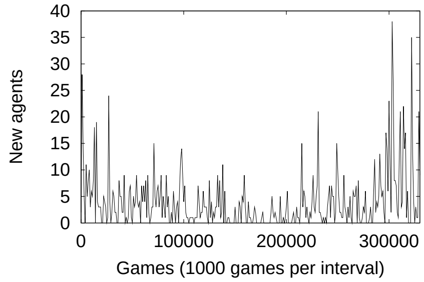

Figure 5: Number of new agents every 1000 games.

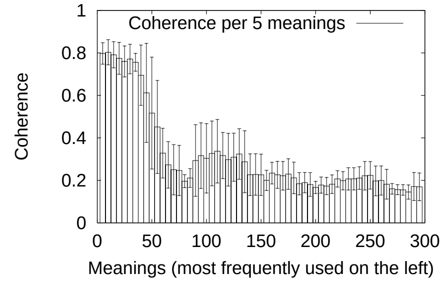

Figure 6: Meaning-word coherence.

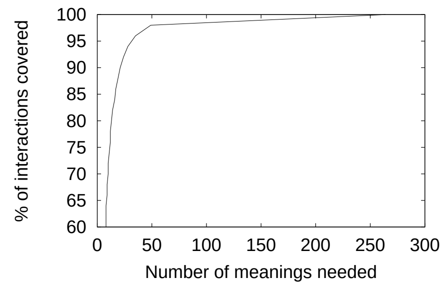

Figure 7: Percentage of interactions covered.

game increases. When no new agents arrive, there is less disturbance, and success rates can rise. During the day, when people start sending out their agents again, success levels drop again.

Another factor influencing the success rates is the influx of new agents. Figure 5 shows the number of new, agents in the system per interval of 1000 games. These new agents have to learn the language from scratch, which causes a number of unsuccessful games until they learn enough of the language.

There are a number of other external events that can cause communicative success to drop, such as changes in the layout of the geometrical shapes on the whiteboard, breakdowns of the equipment or errors in the calibration, etc.

Coherence

In the basic naming game implementation, coherence measures the extent to which all agents in the population use the same words for the same objects. It is measured by counting, for every object, how many agents prefer to use the same word for it. Generally, coherence will not be total, so that some agents will prefer different words than others. In this case, the word that most agents prefer will be considered the one that the population prefers. Averaging this over all meanings gives a measure for the quality of the language that the agents developed.

For the Talking Heads, the notion of coherence has to be extended, because there is no direct relationship any more between objects and words (see fig. 1). In this case, coherence can be calculated not only between words and objects, but also between words and meanings, and meanings and objects. Unfortunately, in the Talking Heads experiment calculation of meaning-object or word-object coherence is not possible, because there is not enough information in the database to reconstruct the referents of the interactions.

When meaning-word coherence is calculated in the standard way, averaging over all meanings, the Talking Heads experiment scores a mere 43.2%. This is not very much compared to the 90–100% found in the basic experiments, suggesting that indirect reference to objects results in much worse performance (which would be corroborated by the low success scores). However, figure 6 shows meaning-word coherence in an alternative way. Every bar in the graph shows the average coherence for 5 meanings, with the error bar showing the standard deviation. The meanings are sorted according to their frequency of use in the interactions; the most used meanings are on the left side. It can be seen clearly that for the most frequently used meanings, coherence is up to the levels achieved in the basic language game experiments. Only the meanings that are less frequently or almost never used, have a very low coherence.

Figure 7 shows how large a fraction of the interactions (y-axis) is covered by how many of the most frequent meanings (x-axis). It can be seen that in 98% of the interactions, one out of the 50 most frequent meanings is used. This confirms that there is a small number of meanings that are used very often. This is not observed in the basic naming game: there is a fixed number of meanings, and for each meaning there will be a preferred word. The extra meaning selection step that is performed by the Talking Heads agents introduces a lot of meanings that are used only once or very few times. Since the agents are not capable of removing unused associations from their lexicons, they remain in the lexicon for the duration of the experiment.

Conclusion

The Talking Heads experiment was a publically accessible, real-world implementation of the naming game, which was previously studied using simulation experiments. People from all over the world were able to create agents, launch them into the network of sites that were available at different places and teach them new words.

The experimental data reveal the influence of the real world. Communicative success was much lower than in the idealised simulation experiments because of many external factors, such as the inflow of new agents in the system, and the movements of agents between different sites. Specifically, a correlation has been found between the time of day and communicative success, which is probably due to reduced agent movements during the night.

Despite the comparatively low communicative success, suggesting that a shared lexicon did not develop at all, the language that developed in the early stages of the experiment remained stable. One reason for this is that in the majority of the language games, only relatively few meanings are used: in 98% of the interactions, one of the 50 most frequent meanings is used. This phenomenon is not seen in the basic simulation experiments, where the set of meanings is fixed, and the lexicon only contains associations with these relevant meanings. The real-world agents, which have to do a meaning selection process next to the word selection process, collect many associations that turn out to be not useful later on; it would probably be advantageous for them to have a mechanism for "forgetting" these unused associations.

For the core set of 50 meanings, coherence is almost 80%, while for the other meanings coherence is between 20 and 35%. Consequently it will be very hard for newly arriving agents to change the core language. They will have to learn it and conform to it in order to be successful in communication, which shows that the language is resilient against changes.

Examination of the behaviour of the agents in the real-world Talking Heads experiment shows many similarities to the behaviour of the agents in the simulation experiments, but there are also many differences. The observed behaviour is influenced by many external factors that are not present in the simulation experiment. The mechanisms that work in a straightforward way in a simulation experiment, such as the establishment of coherence in a lexicon, behave in more subtle ways, and might even break down if the "real" circumstances are too different from the simulated ones. For this reason, it is a good idea to complement simulation experiments with real-world robotic experiments.

Acknowledgements

The research presented in this paper was funded by the Flemish Institute for the Promotion of Innovation by Science and Technology (IWT). It is based on data gathered from a joint experiment done by the VUB AI-lab and Sony CSL Paris, notably prof. dr. Luc Steels who was also the initiator and driving force behind the experiment, Frédéric Kaplan, Angus McIntyre, Tony Belpaeme and myself.

References

[Batali, 1998] Batali, J. (1998). Computational simulations of the emergence of grammar. In *Approaches to the Evolution of Language: Social and Cognitive Bases*, pages 405–426. Cambridge University Press, Cambridge.

[de Boer, 1997] de Boer, B. (1997). Emergent vowel systems in a population of agents. In *Proceedings of ECAL-97, Brighton UK, 1997*. The MIT Press, Cambridge, Ma.

[Hurford, 1989] Hurford, J. R. (1989). Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua*, 77:189–222.

[Kirby, 1999] Kirby, S. (1999). *Function, Selection and Innateness: the Emergence of Language Universals*. Oxford University Press.

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

[Ogden and Richards, 1935] Ogden, C. and Richards, I. (1935). *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. Routledge Pub., New York.

[Oliphant, 1996] Oliphant, M. (1996). The dilemma of saussurean communication. *BioSystems*, 37 (1-2):31–38.

[Steels, 1996] Steels, L. (1996). Emergent adaptive lexicons. In Maes, P., editor, *Proceedings of the Simulation of Adaptive Behaviour Conference*. The MIT Press, Cambridge, Ma.

[Steels, 1997] Steels, L. (1997). The origins of syntax in visually grounded robotic agents. In Pollack, M., editor, *Proceedings of the 15th International Joint Conference on Artificial Intelligence*. Morgan Kauffman.

[Steels and Kaplan, 1999] Steels, L. and Kaplan, F. (1999). Bootstrapping grounded word semantics. In Briscoe, T., editor, *Linguistic evolution through language acquisition: formal and computational models*. Cambridge University Press, Cambridge, UK.

[Steels and Vogt, 1997] Steels, L. and Vogt, P. (1997). Grounding adaptive language games in robotic agents. In *Proceedings of the 4th European Conference on Artificial Life*. The MIT Press, Cambridge, Ma.