Language as a complex adaptive system

Luc Steels
VUB AI Lab – Brussels & Sony Computer Science Laboratory - Paris

1. Introduction

The paper surveys recent work on modeling the origins of communication systems in groups of autonomous distributed agents. It is shown that five principles gleaned from biology are crucial: reinforcement learning, self-organization, selectionism, co-evolution through structural coupling, and level formation.

It is by now well accepted that we can discover new computational and problem solving paradigms by studying natural systems, particularly complex dynamical systems. This strategy was still a dream 15 years ago (Steels 1988) but has borne rich fruits. Novel computational problem solving paradigms now exist based on analogies with spin glasses, genetic evolution, immune system dynamics, collective insect behavior, DNA, biochemical reactions, and last but not least neural networks. In each case, a set of primitive elements and their behavior is defined, an interaction is set up with the environment, and the collective behavior is studied that emerges from interactions between the individual elements and the environment. A specific problem is solved by mapping it into the initial states of a complex system and defining the dynamics, such that a solution corresponds to one of its attractors. For example, the traveling salesman problem can be mapped onto the initial state of a spin-glass system, and after the spin-glass dynamics has operated, a final solution can be read from the resulting state. The same problem can also be mapped into DNA strings (Adleman 1994) or the initial states of other dynamical systems.

About 5 years ago, a number of researchers started to adopt this same strategy with respect to language (see Steels (1997) for review of earlier work). The basic idea is that a community of language users (further called agents) can be viewed as a complex adaptive system which collectively solves the problem of developing a shared communication system. To do so, the community must reach an agreement on a repertoire of forms (a sound system in the case of spoken language), a repertoire of meanings (the conceptualizations of reality), and a repertoire of form-meaning pairs (lexicon and grammar). Communication is not a general computational problem of course (although neither is the traveling salesman) but nevertheless a problem of great interest.

First of all there is a strong interest from a scientific point of view. Finding the key to how communication systems of the complexity of human natural languages emerge may help solve the problem of how human language itself may have originated and evolved. This longstanding, fascinating question is receiving increasing attention lately (Hurford et al. 1998, Jackendoff 1999), but only clear scientific models that explain HOW language evolved (as opposed to enumerating conditions why language evolved) can be expected to steer us away from the many speculations that made the field suspect for a long time. By clear scientific models I mean that the cognitive structures and interaction behaviors of each agent is specified and that it is shown how they collectively lead to a language.

Second, there is an interest because of possible applications. On one hand, autonomous artificial agents which need to coordinate their activity in open-ended environments could make use of these mechanisms to develop and continuously adapt their communication systems (Steels 1998a). On the other hand, understanding how language develops and evolves is probably our only hope to ever get to technological artifacts that exhibit human-level language understanding and production. Human languages are constantly changing and differ significantly from one speaker to the next and from one context to the next. So, we need language technologies which exhibit the same adaptivity as humans.

The rest of the paper reviews some of the experiments conducted so far. They always have the same form: (1) They involve a population of (artificial) agents, possibly robots. (2) The agents engage in interactions situated in a specific environment. Such an interaction is called a game. (3) Each agent has a sensori-motor apparatus, a cognitive architecture, and a script determining how it interacts with others. (4) There is an environment (possibly the real world) which consists of situations that are ideally open-ended. The situation in modeling the evolution of communication systems is different from that of using spin glasses or other natural dynamical systems. Spin glasses are much simpler systems and have been thoroughly studied in the natural sciences, whereas the way in which humans acquire, interpret, and produce language remains to a great extent a mystery.

2. Imitiation games for the emergence of sound systems

The work of De Boer (1997) is one of the best examples of how a repertoire of forms may become agreed upon in a distributed group of agents. This work focuses exclusively on the emergence of vowels. Clear universal tendencies exist for vowel systems (Schwartz et al. 1997) and it was already shown that they are due to functional and sensori-motor constraints (Lindblom et al. 1984). The question being addressed in the new experiments is how agents can come to SHARE a system of vowels without having been given a preprogrammed set and without any central supervision.

In the robotic simulations, the sensori-motor apparatus of the agents consists of an acoustic analyzer on one hand, which extracts the first formants from the signal, and an articulatory synthesizer on the other. The agents play an imitation game. One agent produces a random sound from its repertoire. The other agent (the imitator) recognizes it in terms of its own repertoire and then reproduces the sound. Then the first agent attempts to recognize the sound of the imitator again and if it is similar to its own, the game is a success, otherwise it is a failure. This setup therefore adopts the motor theory of perception, whereby the recognition of a sound amounts to the retrieval of a motor program that can reproduce it.

To achieve this task, the agents in the De Boer experiment use two cognitive structures: The vowels are mapped as points into a space formed by the first, second, and third formants (see Figure 1), and a nearest-neighbor algorithm is used to identify an incoming sound with the sounds already stored as prototypes. These prototypes have an associated motor program that can be used to reproduce the sound. When an imitation game succeeds, the score of the prototype goes up, which means that the certainty that it is in the repertoire increases. There are two types of failure. Either the incoming sound is nowhere near any of the sounds already in the repertoire. In that case it is added to the prototype space and the agent tries to find its corresponding motor program by a hill-climbing process, producing and listening to itself. Alternatively, the incoming sound is near an existing sound but the reproduction is rejected by the producing agent. This means that the imitator does not make sufficiently fine-grained distinctions. Consequently, the failure can be repaired by adding this incoming sound as a new prototype to the repertoire and associating it with a motor program learned again by hill climbing. In order to get new sounds into the repertoire, agents occasionally ‘invent’ a new sound by a random choice of the articulatory parameters and store its acoustic image in the prototype space. Sounds which have consistently low scores are thrown out, and two sounds that are very close together in the prototype space are merged.

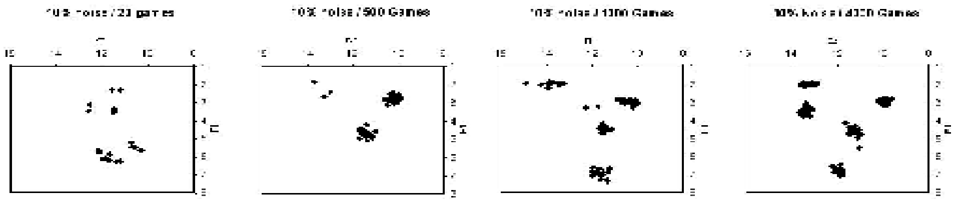

Figure 1: Example of the evolution of a vowel system. Vowels are represented in formant space (first and second formant). We see that coherence as well as increased complexity emerge progressively.

Quite remarkably, the following phenomena are perceived when a consecutive series of games is played by a population of agents: (1) A repertoire of shared sounds emerges through self-organization (see again Figure 1). (2) The repertoire keeps expanding as long as there is pressure to do so. (3) Most interestingly, the kinds of vowel system that emerge have the same characteristics as those of natural vowel systems. The experiment therefore not only shows that the problem can be solved in a distributed fashion, but also that it captures some essential properties of natural systems.

Three principles have been used: Reinforcement learning (Sutton & Barto 1998), based on feedback after each game. It is used to reinforce a vowel in the repertoire of an individual agent or dismiss it. Reinforcement learning in itself does not explain, however, how the group arrives at a shared solution. There is a second principle at work here: self-organization. Self-organization (in the sense of Nicolis & Prigogine 1989) arises when there is a positive feedback loop in an open nonlinear system. Concretely, there is a positive feedback between use and success. Sounds that are (culturally) successful propagate. The more a sound is used the more success it has and it will be used even more. Self-organization explains that the group reaches coherence, but not why these specific vowels occur and not others. For this we need a third principle, namely selectionism. The scores of vowels that can be successfully distinguished and reproduced, given a specific sensori-motor apparatus, have a tendency to increase and they hence survive in the population. Novel sounds or deviations from existing sounds (which automatically get produced due to the unavoidable stochasticity) create variation, and sensori-motor constraints select those that can be reproduced and recognized. The closer we can model human natural sensori-motor behavior, the more realistic the vowel systems become.

3. Discrimination games for evolving meaning repertoires

Another series of experiments has demonstrated how a meaning repertoire may emerge in a population. Once again, a population of agents is defined which play a consecutive series of games. The games are typically discrimination games. An agent perceives some part of reality, for example through a camera and low-level segmentation and feature detection, and selects one of the objects (more precisely segments in the visual image) as the topic. The agent then tries to distinguish the topic from the other objects in the context, for example by finding a category (or a logical combination of categories) that is valid for the topic but not for the other objects in the context. Thus, suppose the scene contains a red triangle to the left of the image, a green square to the right, and a red square above it. Suppose that the red triangle has been chosen as a topic, then its possible distinctive features are: red, triangle, object to the left, or a conjunctive combination of these.

In the experiments reported in Steels (1998b), the agents start from a visual image captured by a camera. The cognitive structure being used consists of discrimination trees for every sensori-motor channel available to the agent (Figure 2). The tree divides a sensory channel up into finer and finer sub-regions. New divisions are generated by randomly selecting a channel and making a further subdivision, somewhat like leaves growing on a tree in a random fashion. Data about each segment in the scene falls into a sub-region on each of the channels and a distinctive sub-region (and hence a category) is identified if the data of the topic falls into it but not the data of any other object in the scene. Each category has a score reflecting how much it has been used successfully. When categories have no success they are eventually pruned. Experiments with robotic agents confronted with real world environments have shown that a stable repertoire of categories builds up. The categorial repertoires are not necessarily identical in each agent and they continue to expand or contract when the environment changes.

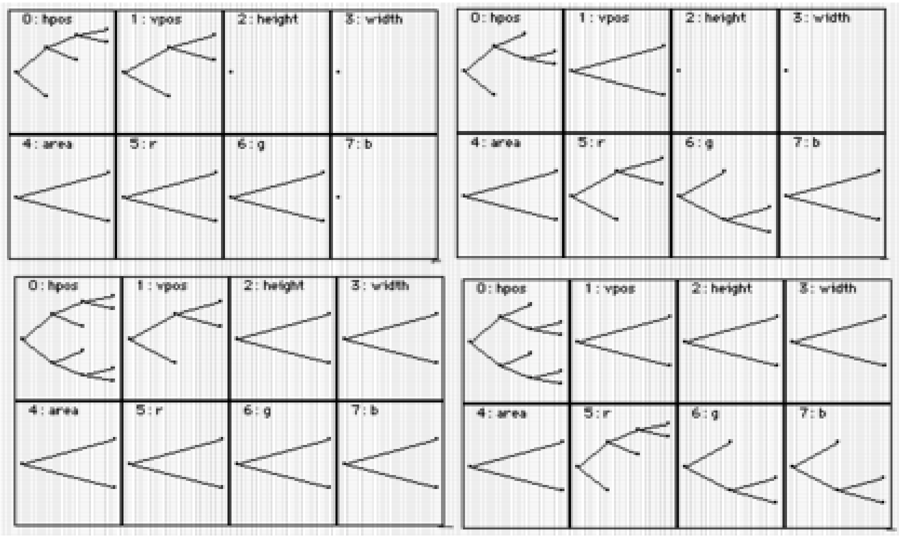

Figure 2: The discrimination trees developed by two physically embodied agents (left) and (right). The top of the figure shows the trees after playing 100 games and bottom after 200 games.

Other similar experiments use other kinds of categorial mechanisms (for example, prototypes or adaptive subspaces; De Jong & Vogt 1998). What is important, therefore, is not the specific learning mechanism, but rather the idea of organizing the repertoire formation process in terms of a series of games. We see that some of the same mechanisms have been put to work as for emergent sound systems: Reinforcement learning, because the environment gives feedback on which categories are distinctive and this is used to maintain or eliminate them, and selectionism, because the spontaneously generated categorial distinctions undergo selection pressure from the environment and the discrimination task itself. There is no self-organization because each agent individually constructs its own repertoire, which will only be similar to the extent that the agents are in the same environment and use the same sensori-motor apparatus.

4. Naming games for form-meaning repertoires

Other work has focused on how a shared set of form-meaning relations could collectively be built up by a population of agents (see e.g. Hurford et al. 1998, Steels 1996). Once again, there is a population of agents. They play naming games. The agents have two components in their cognitive architecture: a mechanism for categorization, such as the discrimination trees discussed earlier, and a lexicon which consists of a two-way associative memory storing form-meaning pairs. The agents do not have a general overview nor can they inspect each other's lexicons.

When the speaker has a meaning to express, she looks up in her lexicon what the preferred word is. The hearer uses her own lexicon to retrieve the most expected meaning. The game succeeds when the meaning retrieved by the hearer is compatible with that of the speaker. There is lateral inhibition: The scores of the associations that were used go up and those of competing associations go down. When the game fails, scores of participating associations go down. In some simulations, speaker and hearer have access to each other's meanings, which violates the 'no-telepathy' assumption. In other experiments, this is not the case and speaker and hearer can only indirectly learn whether the same meaning was used. This introduces additional difficulties such as word-sense ambiguity (one word with multiple meanings) in addition to synonymy (one meaning for multiple words).

Figure 3 (from Steels 1999) gives a typical example of experimental results in which an increasing population of distributed agents collectively creates a shared lexicon by playing naming games (and discrimination games, as discussed in the previous subsection). In this case, the agents are robots perceiving geometric scenes in the form of colored figures pasted on the white board in front of them. We clearly see a winner-take-all situation in which one word dominates after a struggle against alternatives. We also see that word-sense ambiguity gets damped. This is more difficult because often more than one meaning is compatible with the same situation and agents have to wait until a situation arises that disambiguates the word-meaning relation.

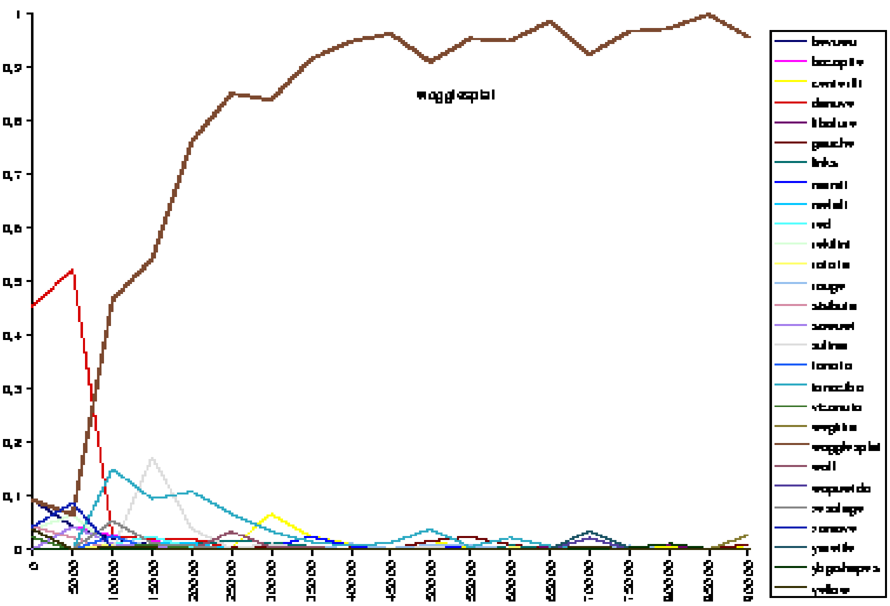

Figure 3: A meaning-form diagram which graphs, for a specific meaning, all the possible forms and their scores. A winner-takes-all situation is clearly observed.

Analyzing this experiment, we see the same three principles as in earlier experiments. There is again a form of reinforcement learning, which is now based on success or failure in the language game as a whole. The game succeeds if the hearer identifies the object that the speaker had in mind. There is self-organization due to the positive feedback loop between use and success: When a word-meaning pair is successful, the score goes up. Agents prefer the word that has the highest score for producing and interpreting, hence words that have more success will be used and this leads again towards greater success. Selectionism also plays a role here because the sensori-motor apparatus of the agent and the situations they encounter act as selectionist forces. For example, a word-meaning pair for which the meaning cannot reliably be recognized by the hearer will have less chances of survival.

A fourth principle, also widely used in biology, plays an additional role now, namely structural coupling (Maturana and Varela) leading to co-evolution between meaning repertoires and lexicons. Specifically, a meaning formation repertoire (such as the one discussed in the previous subsection) is coupled to the lexicon formation repertoire in two ways. New categories may be generated at any time to be successful in discrimination. This obviously influences what word-meaning pairs may arise. But conversely, success in the language game influences the score of the categories used, so the categorization process becomes tuned to be better adapted to the language in the environment of the agents. This way, the ontologies of the different agents become similar even though there is no telepathy and it is not innately given.

5. Experiments in the origins of grammar

Several researchers, most notably Batali (1998), Kirby (1999), and Steels (1998b, 2000), have been conducting experiments to explain how languages with the grammatical complexity of human natural languages may emerge. This requires a scale-up along all dimensions (form, meaning, and form-meaning association), and it is therefore not surprising that many open questions remain. I briefly discuss experiments conducted by Batali (1998) as a representative example.

The experiment once again starts by setting up a population of agents. They have a cognitive architecture which consists of a repertoire of meaning structures and a grammar able to relate (structured) meanings with expressions that have a syntactic structure. The computational and learning mechanisms used by Batali are based on recurrent neural networks, but other types of learning such as memory-based learning or grammar induction could equally well be used. The agents play a language game in which the speaker conceives of a meaning, uses its grammar to translate that to a form, and then transmits that to the hearer. The hearer parses the form and interprets its possible meaning. In this experiment, the (unrealistic) assumption is made that speaker and hearer share meaning independently of language, but other experiments (such as Steels 1998b) do not make this assumption and agents only get indirect feedback as to whether the meaning they guessed was the right one. If the game fails (the meanings are not equal), the networks of the hearer adapt to be more successful in future games. Batali has shown that syntactic structures do emerge when a consecutive series of games is played. The syntactic structures are surely not of the same complexity as those found in human languages and the grammar does not exhibit the sometimes very systematic regularity found in natural languages, but nevertheless compositionality is clearly visible: New sentences could be constructed by the combination of parts built up from earlier sentences.

The Batali experiment (and other experiments in emergent grammar, such as the ones reported in Steels 2000) use the same principles as discussed in earlier experiments: reinforcement learning by the individual to tune in to the conventions present in the population, self-organization based on a positive feedback loop between use and success to get coherence, selectionism constrained by the environments, the sensori-motor apparatus, and the cognitive architecture of the agents, and co-evolution between syntax and semantics through structural coupling. But the study of grammar also introduces a new phenomenon, namely level formation, resulting in hierarchical structures and compositionality. Levels emerge because partial structures can be reused and thus form stable islands within larger structures. Level formation is also found in many biological systems. For example, when a symbiotic relation develops between organisms, this may evolve into a dependent relationship leading to a new, higher-level organism.

6. Conclusions

Although we are only at the beginning of the evolutionary approach to linguistics briefly sketched in this paper, it is already quite clear that some general principles are emerging to understand how a group of distributed agents might autonomously generate communication systems of the complexity of human natural language. These principles are: reinforcement learning, self-organization, selection, co-evolution through structural coupling, and level formation. It is not surprising that all these principles have been inspired by biology. The view that emerges from this research is that language can best be seen as a living system that is continuously evolving and adapting in a cultural process based on the distributed activity of its users. Consequently the computational investigations into genetic evolution, ant path formation, neural networks, and other biological systems are an important source of insight.

This view is in stark opposition to the Chomskyan approach to linguistics, which suggests that language is a largely innate static abstraction uniformly present in the population. The paradigm shift implicit in the work reported here is as profound and important as the shift in biology from typological thinking to the population thinking that started the Darwinian revolution.

Acknowledgement

This work was supported by the Sony Computer Science Laboratory and by a GOA grant to the VUB AI Lab. I thank the members of Sony CSL, particularly Angus McIntyre, Frederic Kaplan, and Pierre-Yves Oudeyer, as well as members of the VUB AI Lab, particularly Edwin de Jong, Bart de Boer, Joris Van Looveren, and Paul Vogt for many discussions and participation in the experiments.

References

Adleman, L.M. 1994. Molecular computation of solutions to combinatorial problems. *Science* 226.1021-1024.

Batali, J. 1998. Computational simulations of the emergence of grammar. *Approaches to the evolution of language*, ed. by J. Hurford et al., 405-426. Cambridge: Cambridge University Press.

De Boer, B. 1997. Self-organisation in vowel systems through imitation. *Proceedings of the fourth European Conference on Artificial Life*, ed. by P. Husbands & I. Harvey. Cambridge, MA.

De Jong, E. & P. Vogt. 1998. How should a robot discriminate between objects? A comparison between two methods. *Proceedings of the Fifth International Conference of The Society for Adaptive Behavior*. Cambridge, MA: MIT Press.

Hurford, J., C. Knight & M. Studdert-Kennedy. (eds.) 1998. *Approaches to the evolution of language: Social and cognitive bases.* Cambridge: Cambridge University Press.

Jackendoff, R. 1999. Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences* 3.272-279.

Kirby, S. 1999. *Function, selection and innateness: The emergence of language universals.* Oxford: Oxford University Press.

Lindblom, B., P. MacNeilage & M. Studdert-Kennedy. 1984. Self-organizing processes and the explanation of language universals. *Explanations for language universals*, ed. by B. Butterworth et al., 181-203. Berlin: de Gruyter.

Nicolis, G. & I. Prigogine. 1989. *Exploring complexity*. New York: Freeman.

Schwartz, J.-L., L.-J. Boe, N. Vallee & C. Abry. 1997. Major trends in vowel system inventories. *Journal of Phonetics* 25.233-253.

Steels, L. 1988. Artificial intelligence and complex dynamics. *Concepts and techniques of knowledge-based systems*, ed. by M. Tokoro & A. Yonezawa. Amsterdam: North-Holland.

Steels, L. 1996. Self-organising vocabularies. *Proceedings of Alife V*, ed. by C. Langton. Cambridge, MA: MIT Press.

Steels, L. 1997. The synthetic modeling of language origins. *Evolution of Communication* 1.1-35.

Steels, L. 1998a. The origins of ontologies and communication conventions in multi-agent systems. *Journal of Autonomous Agents and Multi-Agent Systems* 1.169-194.

Steels, L. 1998b. The origins of syntax in visually grounded agents. *Artificial Intelligence* 103.1-24.

Steels, L. 2000. The emergence of grammar in communicating autonomous robotic agents. *Proceedings of ECAI 2000*, ed. by W. Horn. Amsterdam: IOS Publishing.

Steels, L. & F. Kaplan. 1999. Collective learning and semiotic dynamics. *Advances in Artificial Life (ECAL 99): Lecture notes in computer science*, ed. by D. Floreano, J. Nicoud & F. Mondada, 679-688.

Sutton, R. & A. Barto. 1998. *Reinforcement learning*. Cambridge, MA: MIT Press.