Evolving grounded communication for robots

Luc Steels

SONY Computer Science Laboratory, Paris, France; and University of Brussels (VUB AI Laboratory), Belgium

Corresponding author: Luc Steels (e-mail: steels@arti.vub.ac.be)

The computational and robotic synthesis of language evolution is emerging as a new exciting field of research. The objective is to come up with precise operational models of how communities of agents, equipped with a cognitive apparatus, a sensori-motor system, and a body, can arrive at shared grounded communication systems. Such systems may have similar characteristics to animal communication or human language. Apart from its technological interest in building novel applications in the domain of human–robot or robot–robot interaction, this research is of interest to the many disciplines concerned with the origins and evolution of language and communication.

Artificial Life (AL) developed in the early 1990s as a field investigating principles of living systems by building artificial systems that model some of their key aspects. For example, researchers have built operational models of pattern formation in biological systems (as observed on shells or in nest structures of insect societies) that generate similar structures to those found in nature [1]. These models thus provide a causal insight in how these phenomena arise. The notion of a complex adaptive system, in which many independent elements dynamically coordinate their activity through self-organization, has emerged as a key concept. Several other system concepts from biology, such as evolution by natural selection, structural coupling, order arising out of chaos, and network dynamics, could be operationalized and shown to be relevant for understanding a wide range of natural phenomena.

The interest of this work is not only theoretical. Genetic evolution, as captured in AL models known as genetic algorithms, have made it possible to solve a large set of engineering problems by using the search power implicit in selectionist processes [2]. Also in robotics, the AL approach, which insists on a bottom-up emergence of complexity and the role of collective dynamics in behaviour, has had an important impact and is largely responsible for the impressive jump forwards in autonomous robots accomplished during the past decade [3]. More recently, AL models have begun to impact on research in biomolecular information processing [4].

Language as a complex adaptive system

Within the same methodological framework and pool of ideas, a few researchers began to build computer simulations and robotic experiments in which artificial communication systems emerge, invented and learned by artificial agents. The objective is again twofold.

The first goal is scientific: to make precise models of how certain key properties of language-like communication systems might in principle originate, and how such communication systems might continue to evolve and remain adapted to the needs of their users. Quite naturally, language is viewed as a living system, that self-organizes and evolves through the collective dynamics of agents engaged in situated verbal interactions. The insights and apparatus for studying complex adaptive systems, as emerging from research into complex systems and AL, thus becomes relevant for the study of language. A new field, evolutionary linguistics, was born with a specific focus (understand the origins of language and meaning), a specific hypothesis (language is a complex adaptive system), and a specific methodology (construct artificial systems as a way to develop and test theories). An overview of earlier work in this field can be found in [5] and more recent work in [6]. Several recent collections of papers [7–9] provide additional source material.

The relationship to linguistics is indirect. Linguistics is an empirical science which studies human natural languages as they exist. Evolutionary linguistics is a theoretical endeavour, focusing on the evolution of grounded communication systems in general. The artificial communication systems generated in experiments may have similar properties to human languages, to various degrees, but need not be identical – in fact they will never be. The artificial languages generated in robotic experiments need not make use of sound or might express meanings that are irrelevant or even inaccessible to human beings but pertinent to robots. They might involve grammatical structures that are not used in any known human language. Whether the results of these investigations are relevant to contemporary linguistics is an open question, but it cannot be denied that fascinating questions are being raised.

The relationship with computational linguistics and AI is also indirect, even though a lot of the technology and implementation techniques are very similar. The goal of evolutionary linguistics is not to build systems that can parse or produce English, but to understand the generic forces and mechanisms that give rise to these capabilities. The goal is not to build systems that acquire an existing natural language such as English, using neural networks or symbolic machine learning techniques operating over large datasets, but to understand how natural language-like lexicons or grammars might arise in the first place and by what mechanisms they can continue evolving.

Relevance to robotic applications

This research has a secondary, more practical goal, namely to forge a new technology for communication between humans and robots, or among robots. Recently, tremendous advances in robotics and artificial intelligence have given us fully operational autonomous robots, even humanoid robots walking on two legs, with stereo vision, surround audition, real-time adapted dynamical trajectory planning, vision-based navigation, and many other features [10]. A new breed of 'pet' robots, such as the Sony dog-like AIBO robot [11] (Fig. 1), has been commercialized, whose primary aim is not to perform practical tasks but to interact with people and entertain them. These robots create a semi-artificial world, similar to the world of opera or puppet theatre, in which various forms of pretend play are spontaneously initiated by humans.

There is also a growing variety of autonomous robots that are inspired by living systems [13]. These robots are intended for inspection of sewage pipes, monitoring of pollution through underwater measurements, space exploration, bio-medical interventions, or nano-engineering. There is an increasing number of applications which requires that collections of such robots coordinate their efforts through grounded communication systems.

Finally there are various kinds of software robots or agents, operating in the fast changing world of electronic communication, such as the Internet. Colonies of software agents and human communities aided by wireless communication devices are beginning to show the properties of complex adaptive systems. They behave like 'smart mobs' that spontaneously self-organize and develop communication protocols adapted to their needs [14].

There is a general consensus that pre-programmed communication is inadequate for these new generations of robots and software agents. First, the environment in which they find themselves is the real world and hence it is open-ended. It is not possible to foresee all things that might happen or all task situations that may be encountered, nor what communication is going to be needed. Second, humans are known to negotiate shared conventions as part of dialogues [15]. Robots interacting with humans therefore need the ability to cope with new meanings, new variations in speech from unknown users, new words or shifts in the meaning of existing words, new grammatical constructions, and new interaction patterns [16]. Third, some applications, particularly those oriented towards entertainment, require that the robots can self-develop because that makes them much more exciting. This implies that robots should keep discovering new behaviours and new modes of interaction, including new ways to communicate.

Evolutionary language games

Game theory

Game theory has proven its usefulness in evolutionary biology and economics and plays a profound role in many artificial life discussions [17]. It is therefore not surprising that it has been adopted as a framework for studying the origins and evolution of communication systems in populations of agents, not only for performing large-scale computational and robotic experiments [18] but also for developing mathematical theories [19].

A language game model consists of a population of agents. Each agent is embodied or its embodiment in the physical world is simulated. It has a cognitive apparatus including learning mechanisms relevant to the aspects of language and meaning one wants to study. For example, an agent might consist of an articulatory system, simulating the human vocal apparatus, an auditory perception component, and an associative memory, relating speech percepts to vocal motor control programs. This would be relevant for studying the emergence of human-like speech sounds [20].

The agents typically interact with the environment and each other through a sensori-motor system, which can take on various degrees of complexity. In the case of robotic implementations, the environment is the real world itself, possibly including different actors and dynamic events.

Agents have or must develop scripts for playing language games and they take turns playing the role of speaker and hearer so that they build up competence both for interpreting and for producing language utterances. They never have access to each other's mental states. All interaction goes through the world. Of particular importance are mechanisms for handling breakdowns in the game: introducing or learning a new sound, learning the meaning of a word never heard before, extending by analogy a new word for a meaning that was never expressed before, introduce or acquire a new grammatical construct, and so on.

The Talking Heads experiment

Figure 2 shows the set-up of a large-scale grounded language game experiment, with which a population of thousands of agents has played close to half a million guessing games [18]. The objective of this 'Talking Heads experiment' was to show how such a population would be able to generate and self-organize a shared lexicon as well as the perceptually grounded categorizations of the world expressed by this lexicon, all without human intervention or prior specification.

The environment consists of an open-ended set of geometric figures pasted on a white board (Fig. 2). One figure is chosen randomly by the 'speaker' as topic of the game and the 'hearer' has to guess the intended topic based on words supplied by the speaker. The robots use their pan-tilt cameras for visual sensing and for pointing. At the start of a game, the speaker moves his camera towards a specific area of the white board, thus indicating roughly to which area the hearer should pay attention. After interpreting the words, the hearer points the camera to the object he guessed. When the hearer points to the wrong topic or signals failure in understanding, the speaker points to the topic he originally intended to indicate.

Two processes are required to play these lexical language games. First, the speaking agent must conceptualize the context in such a way that he finds a category or set of categories which distinguishes the topic from the other objects. Thus if all objects have the same colour, but the topic is much bigger in size, this would be an appropriate distinction. If the agent has no adequate distinction yet, he should create a new more refined distinction. Prototype-based categorization [21], neural networks [22], or some symbolic decision tree learning algorithm [23], have all been shown to be effective for generating new categories and refining them in an accumulative fashion.

Second, the speaker must be able to express the selected categories through words and the hearer must be able to parse them. This can be accomplished with an associative memory relating possible meanings (i.e. visual categories) and forms (words). The speaker chooses words for maximum communicative success. Hence agents keep a score for each form–meaning pair in their lexicons and choose the one with the maximum success in the past. If a word–meaning pair fails in the game, its score is decreased. If it succeeds, the score is increased and that of competing word forms (i.e. words associated with the same meaning) decreased. When the speaker does not have a word yet for a particular category, he randomly generates a new one, and the hearer can learn this world by guessing the meaning from the context.

This lateral-inhibition dynamic progressively results in the self-organization of a lexicon in a group (see Fig. 3). It plots the change in frequency of different words to express a single meaning in one run of the Talking Heads experiment [17]. There is first a struggle in which different words compete, until the population settles on a single dominant word. The winner-take-all effect is due to a positive feedback loop between use and success: The more agents prefer a particular word, the more they use this word and hence the more success the word has. This further increases its score in the different agents and progressively causes all agents to adopt it. By tightly coupling the processes that create or learn categorial distinctions and the processes that lexicalize them, the conceptual repertoires of the agents become coordinated without having been programmed in, and without central control or telepathy (see Fig. 4) [22].

Horizontal and vertical transmission

The Talking Heads experiment is an example of a language game model where linguistic structures arise in the interaction between agents and propagate horizontally in the population, like viruses. The structures that emerge depend on the properties of the communication medium, the environments in which the agents operate, the embodiment of the agents, their cognitive capacities, and the types of interactions they have. Other experiments have explored the same framework for the emergence of a shared repertoire of speech sounds [19] or grammatical constructions [24]. The same type of self-organizing complex adaptive dynamics occurs in a wide variety of natural systems and has been studied intensely in complex systems science [1] and embodied cognitive science [25]. Despite several intriguing simulations and robotic experiments, most of the work is still before us. More work is needed on experiments with dynamical environments in which the robots can perform actions as part of the language games. Many aspects of grammar, such as the spontaneous formation of new levels mapping form to meaning and including the origins of new syntactic and semantic categories, have hardly been touched upon. Most semantic domains grammaticalized in the world's natural languages have not yet been studied.

Another class of language game models relies on vertical transmission, either genetically [26] or culturally [27], or in a combination of the two [28]. In this case, there is a population of agents which inherits or learns its linguistic behaviour from a previous generation, and it is in the transmission process that novelty (including increased complexity) arises. For example, an agent who learns the grammar from another agent based on a series of sentences that the teacher produces, may streamline and introduce regularities which are not necessarily present in the data. When the learner becomes a teacher for the next generation, these regularities will show up in the language data he produces and they thus become part of the language. (Examples of this approach are discussed in another article in this issue [29].)

The future of evolving communications research

Evolving communication systems for new generations of robots brings up a wide range of fascinating open issues. Many of them are far from resolved, even though substantial progress has been made in recent years. More specifically, we can identify the following major research challenges:

(1) Pragmatic feedback is crucial for bootstrapping grounded communication. It requires attention sharing, face identification and tracking, gestural recognition, shared task awareness, script execution and recognition, emotion recognition and synthesis, etc. Many of these capabilities have already been demonstrated on robots [30] but their smooth integration remains an enormous challenge. We need to understand better the operation of these various capabilities and also the nature of pre-linguistic, implicit communication. We also need better models how joint attention and other prerequisites for grounded communication may emerge.

(2) The experiments discussed in this article all assume that agents are able to play language games, but how do the games themselves emerge? How can a pattern of joint behaviour become ritualized? From a dynamical system point of view, coordinated interaction requires that behaviours of individual agents get entrained, for example through mechanisms like structural coupling [31]. We need to understand better the dynamical systems properties of such coupled systems and see how they can be embedded in individual agents.

(3) Research in cognitive semantics [32] has yielded intriguing descriptions of the kinds of conceptualizations humans employ for language. But it is still largely mysterious how such conceptualizations could arise and how they could be embodied in artificial agents. How can agents develop notions of event-structure, tense-aspect-mood, determination, co-reference in discourse, information structure, etc. How can these conceptual frameworks become shared across agents without being given *a priori*?

(4) Grammaticalization processes [33] play a dominant role in the emergence of new grammatical constructions in natural languages. But so far we are lacking precise models of the cognitive mechanisms and collective dynamics that are involved [34], consequently no convincing simulations of grammaticalization phenomena have been shown yet. One of the major difficulties is that grammaticalization is strongly grounded in many aspects of human culture and embodiment and these aspects are very difficult to incorporate in artificial systems.

(5). We need to understand how populations of agents can self-organize a shared repertoire of discrete building blocks grounded in a continuous physical medium and how a combinatorial system can arise from these building blocks. This is the problem of the origins of phonemic coding and syntax. Although intriguing simulations (not reported in this paper) could be shown for the origins of individual speech sounds [19], major work remains to be done to explain phonemic coding and other universal tendencies in the speech sounds of the world [35].

(6) The question how an embodied agent could invent and learn the linkage between complex conceptualizations and complex syntactic structures is still largely an open problem. Intriguing results have been obtained recently using statistical learning techniques [36] but many more experiments need to be done. We need more powerful grammatical formalisms, as developed for example in embodied construction grammar [37], and a further exploration of constructivist learning techniques [24].

Natural languages are extraordinarily complex communication and representation systems and it is therefore not surprising that they cannot be built by hand. Symbolic machine learning and neural network techniques [38] have shown to be quite effective for capturing aspects of language, but research in evolutionary linguistics provides the first timid steps towards a radically new approach, in which language-like communication systems autonomously evolve in embodied agents through grounded language games [39]. The main difference is that agents participate actively in the invention and propagation of a dynamically evolving language. Much remains to be done but the path ahead of us is full of excitement.

References

1 Camazine, S. *et.al.* (2001) *Self-Organization in Biological Systems*, Princeton University Press
2 Koza John R. (1992) *Genetic Programming: on the Programming of Computers by Means of Natural Selection*, MIT Press
3 Steels, L. and Brooks, R.A., eds (1995) *The Artificial Life Route to Artificial Intelligence: Building Embodied Situated Agents*, Erlbaum
4 Hotani, H. *et al.* (1992) Liposomes, microtubules, and artificial cells. *Nanobiology* 1, 61–74
5 Steels, L. (1997) The synthetic modeling of language origins. *Evol. Commun.* 1, 1–35
6 Kirby, S. (2002) Natural language from artificial life. *Artif. Life* 8, 185–215
7 Briscoe, T., ed. (2002) *Linguistic Evolution Through Language Acquisition: Formal and Computational Models*, Cambridge University Press
8 Cangelosi, A. and Parisi, D. eds (2001) *Simulating the Evolution of Language*, Springer-Verlag
9 Wray, A., ed. (2002) *The Transition to Language*, Oxford University Press
10 Kuroki, Y. *et al.* (2001) A small biped entertainment robot. In *Proc. IEEE–RAS Int. Conf. Humanoid Robots* (Waseda University, Tokyo), pp. 181–186
11 Fujita, M. and Kitano, H. (1998) Development of an autonomous quadruped robot for robot entertainment. *Autonomous Robots* 5, 7–20
12 Steels, L. and Kaplan, F. (2001) AIBO's first words. The social learning of language and meaning. *Evol. Commun.* 4, 3–32
13 Pfeifer, R. and Scheier, C. (2001) *Understanding Intelligence*, MIT Press
14 Rheingold, H. (2002) *Smart Mobs*, Perseus Publishing
15 Clark, H.H. and Brennan, S.A. (1991) Grounding in communication. In *Perspectives on Socially Shared Cognition* (Resnick, L.B., *et al.*, eds), APA Books
16 Steels, L. (2001) Social Learning and Verbal Communication with Humanoid Robots. In *Proc. IEEE–RAS Int. Conf. Humanoid Robots* (Waseda University, Tokyo), pp. 335–342
17 Langton, C. (ed.) (1995) *Artificial Life: An Overview*, MIT Press
18 Steels, L. *et al.* (2001) Crucial factors in the origins of word-meaning. In *The Transition to Language* (Wray, A., ed.), pp. 252–271, Oxford University Press
19 Nowak, M.A. *et al.* (2000) The evolution of syntactic communication. *Nature* 404, 495–498
20 de Boer, B. (2000) Self organization in vowel systems. *J. Phonetics* 28, 441–465
21 Vogt, P. (2001) Bootstrapping grounded symbols by minimal autonomous robots. *Evol. Commun.* 4, 87–116
22 Belpaeme, T. (2001) Simulating the formation of colour categories. In *Proceedings of IJCAI-01* (Nebel, B., ed.), pp. 393–398, Morgan Kauffman Pub.
23 Steels, L. (1996) Perceptually grounded meaning creation. In *Proceedings of the International Conference on Multi-Agent Systems*, pp. 338–344, AAAI Press
24 Steels, L. (2003) Simulating the evolution of a grammar for case. In *Proceedings of the Fourth Evolution of Language Conference* (in press)
25 Thelen, E. and Smith, L.B. (1999) *A Dynamic Systems Approach to the Development of Cognition and Action*, MIT Press
26 Berwick, R.C. (1998). Language evolution and the minimalist program: the origins of syntax. In *Approaches to the Evolution of Language – Social and Cognitive Bases*. (Hurford, J. R. *et al.*), Cambridge University Press
27 Kirby, S. (2002) Learning, bottlenecks and the evolution of recursive syntax. In *Linguistic Evolution Through Language Acquisition: Formal and Computational Models* (Briscoe, T., ed.), Cambridge University Press
28 Briscoe, T. (2002) Grammatical acquisition and linguistic selection. In *Linguistic Evolution Through Language Acquisition: Formal and Computational Models* (Briscoe, T., ed.), Cambridge University Press
29 Christiansen, M. and Kirby, S. (2003) Language evolution: consensus and controversies. *Trends Cogn. Sci.* 7, 000–000 (DOI: 10.1016/S1364-6613(02)02106-X)
30 Breazeal, C. and Scassellati, B. (2002) Robots that imitate humans. *Trends Cogn. Sci.* 6, 481–487
31 Ikegami, T. and Taiji, M. (1999) Imitation and cooperation in coupled dynamical recognizers. In *Advances in Artificial Life* (Floreano, D., *et.al.*), pp. 545–554, Springer-Verlag
32 Talmy, L. (2000) *Toward a Cognitive Semantics: Concept Structuring Systems (Language, Speech, and Communication)*, MIT Press
33 Traugott, E. and Heine, B. (1991) *Approaches to Grammaticalization* (Vols. 1 and 2), John Benjamins Publishing Company
34 Heine (1997) *The Cognitive Foundations Of Grammar*, Oxford University Press
35 Oudeyer, P-Y. (2002). Phonemic coding might be a result of sensory-motor coupling dynamics. In *Proceedings of SAB02* (Hallam B., *et.al.*, eds), p. 406–416, MIT Press
36 Roy, D. (2001) Learning visually grounded words and syntax of natural spoken language. *Evol. Commun.* 4, 33–56
37 Bergen, B. and Chang, N. (2002) *Embodied Construction Grammar: TR-02-004*, International Computer Science Institute
38 Broeder, P. and Murre, J. (2000) *Models of Language Acquisition. Inductive and Deductive Approaches*, Oxford University Press
39 Steels, L. (2001) Language games for autonomous robots. *IEEE Intelligent Systems*, Sep/Oct issue, pp. 16–22

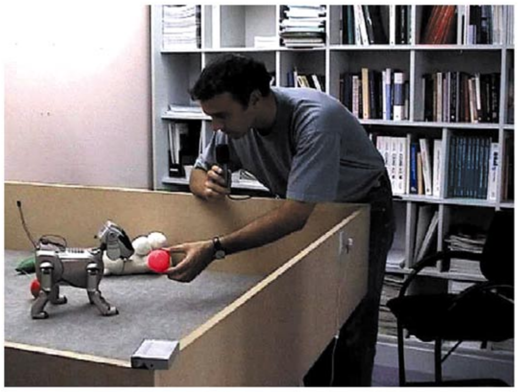

Fig. 1. The AIBO™ pet robot has been used in experiments to see how language-like communication in robots might be bootstrapped by interaction with a human mediator. Reproduced with permission from Ref. [12].

Fig. 2. The 'Talking Heads' experiment featured two pan-tilted cameras oriented towards a white board on which geometric figures were pasted. Agents used these cameras to play lexical games drawing attention to a chosen figure, which is the 'topic' of that game.

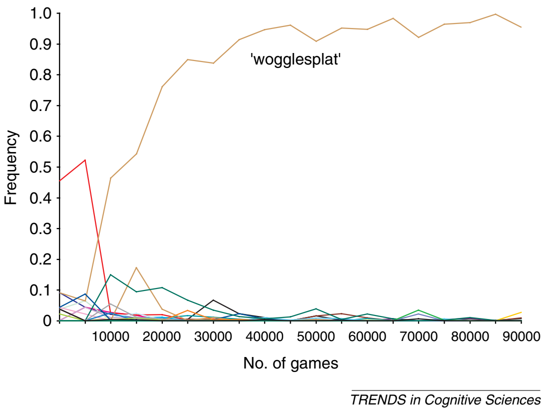

Fig. 3. Results from a Talking Heads experiment. The frequency with which a particular word is used for the same meaning is plotted as a function of the number of games played. Different words are represented by different coloured lines. There is a winner-take-all effect owing to the positive feedback between use of a word and success in the game. Redrawn with permission from Ref. [17].

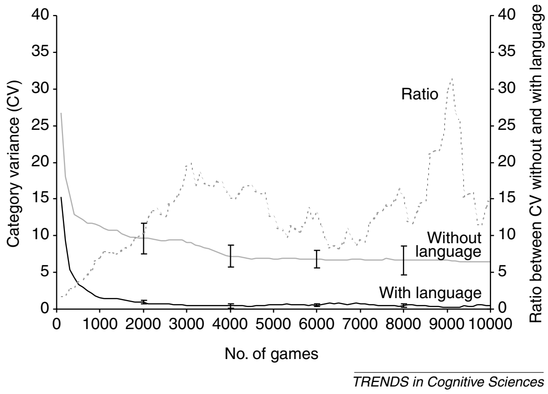

Fig. 4. The variance in categorical repertoires (y-axis) for a series of language games (x-axis) played by a population of agents, both with (bottom plot) and without (middle plot) language. The ratio between the two (top plot) shows clearly how categorical similarity drastically increases when category formation is coupled to language.