How Language Bootstraps Cognition

Luc Steels

VUB AI Lab (Brussels) and
Sony Computer Science Laboratory (Paris)
steels@arti.vub.ac.be

Abstract. Where do perceptually grounded categories come from? Some researchers claim they are innate, others claim they are learned. This paper presents a third 'ecological' approach. I propose a system by which discrimination networks, capable of categorial distinctions, spontaneously grow, rather independently of specific examples. I show that linguistic interaction can drive a spiraling increase in the ontological complexity of an agent. The ontology formation mechanism can be coupled to adaptive language games by which a shared lexicon spontaneously self-organises. Software simulations and experiments with robotic agents serve as examples.

1 Introduction

The past few years, a number of researchers have become interested in building operational theories for how language and the conceptualisation of reality that underlies language can bootstrap itself, both in a developing single individual (ontogenesis) and in the human species (glossogenesis); (Hurford, 1989), (Steels, 1979b), (Kirby, 1999). An operational theory is one that is sufficiently worked out and formalised that it can be tested by computer simulations and robotic experiments. Whether such a theory is also valid for humans remains an open question, but at least the theory is operationally adequate and therefore suggests interesting hypotheses for psychologists, linguists, and neurophysiologists on how things might be.

The ultimate test of an operational theory consists in creating a set of artificial agents with a specific architecture and grounded through a sensori-motor apparatus in the world, and in demonstrating that these agents are autonomously capable to bootstrap their own language and their own ontologies without any human intervention. When the agents are in contact with human language behavior, they should be influenced up to a point where they adopt an existing language and its underlying ontology. This is a formidable challenge because it requires not only a system capable to engage in linguistic interaction about scenes perceived through its senses, but also in mechanisms capable to acquire perceptually grounded concepts, a lexicon, and a grammar, and participate in their collective construction.

2 The Talking Heads Experiment

We have been working in this direction and have reached already some important milestones, particularly in the area of lexical semantics (Steels, 1996a; 1996b; 1997a). These results have been integrated in the Talking Heads experiment[1], a large-scale public experiment in which visually grounded and situated robotic agents autonomously develop a shared lexicon and ontology from scratch (Steels and Kaplan, 1999b). The agents use vision to make contact with the world. There are different physical installations (Paris, Tokyo, etc.) connected through the Internet, and agents can teleport from one robot body to another thus experiencing different realities. The agents can also interact with humans so that their evolving languages are influenced by human communication.

[1] `http://talking-heads.csl.sony.fr`

Inspired by a Wittgensteinian view, the interaction between the agents consists of a language game, called the guessing game. 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. 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 conceptualisation module responsible for categorising reality or for applying categories to find back the referent in the perceptual image, and a verbalisation module responsible for verbalising a conceptualisation 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-organised 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 languages emerging from these experiments can be called proto-languages (in the sense of Bickerton (1995)). This means that they consist of single or multi-word phrases without any grammar. At the moment research is progressing to study the emergence of grammar, which is attempted by (1) introducing a productive way to generate more complex semantic structures, and (2) introducing the ability to recognise and produce hierarchical structures (Steels, 1998).

3 Underlying hypotheses

Our first main strategic goal has been to solve the problem at all without any theoretical preconceptions, in other words to find an agent architecture and modes of interaction that would exhibit the desired bootstrapping of cognitive and linguistic capabilities. We have tried many 'classical' solutions on the way, such as induction of perceptually grounded concepts through neural networks, language learning by abstraction from examples, designer-supplied perceptually grounded categories (as would be the case if they were innate), incorporation of an innate universal grammar, etc. These solutions could not be made to work for various reasons and had to be rejected in favor of rather radical alternatives.

The main insights so far can be summarised as follows:

1. Selectionism. We do not adopt an instructionist approach, in which language and meaning are assumed to be learned by progressive abstraction from a series of examples, neither a nativist approach, where language and meaning is assumed to be largely innate. Instead we provide agents with mechanisms for growing in a partly random fashion new internal structures (for perception, categorisation, lexicalisation, syntactic structuring) which internally compete to participate in a particular game. Feedback then propagates about the success of these structures in the game, and this feedback is used to determine which structures will be retained and which ones are pruned. This selectionist process assumes massive parallelism both in the operation of the cognitive system and in the search towards viable solutions. There is a structural coupling between the different cognitive layers so that they become coordinated without a central coordinator.

2. Self-organisation. Language should be viewed as a complex adaptive system. Self-organisation (as opposed to innateness) plays a key role in achieving coherence and general cognitive and sensori-motor constraints as well as constraints on the physics of our natural world help to explain the universal tendencies observed in human languages. Linguistic self-organisation is achieved when a positive feedback loop is established between use and success in use, so that some conventions flourish and others are overtaken. Various selectionist pressures help to decide which structures survive: efficiency in memory use and processing, learnability, and historical spread. The complex dynamics viewpoint introduces new ways to study language contact (Steels and McIntyre, 1999) as well as language evolution (Steels and Kaplan, 1998) and is completely complementary to the study of the competence and operation of an individual speaker/hearer.

3. Situatedness. Meaning arises in the individual through strongly situated and embodied interactions, as opposed to being given in a clear a priori. It is subjective in the sense of not necessarily shared even by members of the same language community. Individual meaning is structurally coupled to the conceptualisations of other agents through language. Language and meaning co-evolve and their joint simultaneous bootstrapping strongly influence both.

4 Conclusions

Research on operational theories for the origins of language turn out to be a gold mine for discovering and testing new ideas related to the acquisition and collective construction of meaning, lexicon and grammar. Although we are still in the early phases of this kind of research, already some intriguing novel hypotheses have come forward, which are offered to the cognitive science community for critical examination and empirical testing.

References

Bickerton, D. (1995). *Language and human behavior.* UCL Press, London.

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

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

Steels, L. (1996). Self-organizing vocabularies. In: Langton, C. (ed.) (1996) *Proceedings of Alife V*, Nara Japan. The MIT Press, Cambridge.

Steels, L. (1996b). Perceptually grounded meaning creation. In: Tokoro, M. (ed.) (1996) *Proceedings of ICMAS-96*, Kyoto 1996. AAAI Press, Menlo Park CA.

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. (1997b). 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 Kaplan, F. (1998). Spontaneous Lexicon Change, In: *Proceedings of COLING-ACL*, Montreal, Morgan Kaufmann, Publishers. Los Altos CA, pp. 1243-1249.

Steels, L. and F. Kaplan (1999). Situated Grounded Word Semantics. *Proceedings of IJCAI-99*, Stockholm.

Steels, L. and A. McIntyre (1999). Spatially Distributed Naming Games. *Journal of Complex Systems*, Hermes Pub., Paris.