# Investigating the basis for conversation between human and robot

**Caroline Lyon and Joe Saunders**
University of Hertfordshire, Hatfield, UK, AL10 9AB
C.M.Lyon, J.1.Saunders@herts.ac.uk

## Abstract

We investigate preliminary stages in enabling robots to talk with humans in a natural manner, and outline experiments. The process is inspired by language acquisition in infants, and by recent empirical evidence of neuronal organisation.

## 1. Introduction

In this paper we describe preliminary stages in enabling robots to communicate with humans, using natural language. This starts with babbling, analogous to the pre-linguistic infant in a protoconversation with its carer, progressing to learning the meaning of utterances through mediated physical interaction (Saunders et al., 2009). This work is inspired by the acquisition of language by human infants, and by recent empirical evidence of neuronal organisation.

Participants in our experiments talk to the robot in natural, unrestricted language, about a blocks world with objects of various shapes. The robot must learn to "understand" this highly redundant natural language, in which the same concept can be expressed in a number of alternative ways (e.g. "push the red box", "give the red box a push"). Its own productions may be more limited: asymmetrical development is typical of human infant language acquisition (de Boisson-Bardies, 1999, p 201-209).

## 2. Natural language, evolutionary baggage and neuronal organisation

Language has emerged by recruiting mechanisms that originally evolved for other purposes. For instance, in English, French, Japanese, Chinese and other languages there are many common homophones, ambiguous words such as

*no/know* *to/two/too* *their/there*

We disambiguate such words by taking them in context, processing short sequences of linguistic elements as coherent units.

The fact that we do not avoid ambiguous words but resolve their meaning by processing short sequences suggests that such serial processing methods are easily accessible. It seems likely that sequential processing is based on exaptations of faculties originally developed for different purposes. As Steels says: "the human language faculty is a dynamic configuration of brain mechanisms, which grows and adapts .... recruiting available cognitive/neural resources for optimally achieving the task of communication" (Steels, 2007).

As well as Wernicke's and Broca's areas in the brain other regions are involved in language processing. See, for instance, Lieberman (Lieberman, 2000), Dominey et al. (Dominey et al., 2003), Pulvermuller (Pulvermuller, 2002) on why serial processing is a key factor in the perception and production of speech.

Thus, there is significant evidence that dual systems are needed for language processing. On the one hand there is *implicit* learning of patterns and procedures, without intentional shared reference. On the other hand there is *explicit* declarative learning, in which there is joint attention between teacher and learner, and reference to objects, actions or relationships.

This dichotomy is also described as a dorsal pathway concerned with sub-lexical processing, object interactions and phonetic decoding, in contrast to a ventral pathway specialising in object identification and whole word recognition. This functional segregation is also characterised as a motor-articulatory system on the one hand and a conceptual system on the other (Hickok and Poeppel, 2004, Saur and Kreher, 2008). As described below, we adopt this dichotomy in a simplistic manner in the implementation of a language learning robot.

## 3. Implementation

Work is currently being undertaken influenced by the constructivist approach of Tomasello (Tomasello, 2003). The aim of this work is to enable the development of language capabilities in a robot through interaction with a teacher, an actual or

simulated human. Initial assumptions are:
- The robot has the intention to communicate
- Communicative ability is learnt through interaction with a teacher
- Perception and production of speech are based on simulated mirror neuron type structures, in which the same elements reflect components of perceived speech and generate synthesized speech
- Memory sites include distinct areas associated with implicit, pattern learning on the one hand and explicit word learning on the other

The process is based on turn taking. Initially the robot (a synthetic agent in preliminary experiments) produces simulated babbling while a teacher produces utterances in ordinary English, both represented as streams of phonemes. The agent's output starts as random syllables, but becomes biased towards the teacher's speech. The starting point is taken as analogous to the stage at which infants start canonical babbling (de Boisson-Bardies, 1999, p. 45-46). Babbling is thought to play a key role in early language development (Oudeyer, 2006, p. 148) (Pulvermuller, 2002, p. 50)

The robot or agent segments the teacher's utterance into short sections in a variety of ways, based on observed mechanisms. These include phonotactic constraints based on distributional evidence, taking the end of an utterance as a significant unit, taking identified words or holophrases as anchor points for further segmentation. Prosodic information plays a key role for humans, and we plan to use it in future. These segments join the robot's store of prelexical components, available for use in its productions. When the robot produces, by chance, syllables that can be concatenated to make a word, the teacher will give a positive reaction, metaphorically a "reward". The new word becomes latched in memory, a candidate for future production by the robot along with other syllables. Thus a lexical store is built up, and words will be produced embedded in a stream of non-words, ready to be given semantic reference.

The acquisition of meaning would in reality take place at the same time as speech segmentation described above occurs. However, we are investigating these processes separately initially in order to understand each strand better. Experiments in a blocks world, where a human interacts with the humanoid Kaspar2, are described in (Saunders et al., 2009). Our robots learn to extract the semantics of a series of shapes associated with perceived speech patterns (as strings of phonemes), visual and proprioceptive perceptions.

An associative approach usually requires a large number of learning episodes so that statistical regularities can be established. An alternative mechanism, used here, is to have learning experiences biased through intentional reference, such as shared gazing, pointing and other types of feedback to reinforce the utterances of the teacher.

In developing the basis for conversation between human and robot we cannot avoid the evolutionary baggage that human language carries, and we need to understand our own neural language processors if we are to implement robotic systems that carry out similar tasks.

## Acknowledgement

Work described in this paper is conducted within the EU Integrated Project ITALK, 2008-2012, funded by the EU Commission under contract number FP7-214668.

## References

de Boisson-Bardies, B. (1999). *How Language Comes to Children*. MIT.

Dominey, P. F., Hoen, M., Blanc, J.-M., and Lelekov-Boissard, T. (2003). Neurological basis of language: Evidence from simulation, aphasia and ERP studies. *Brain and Language*, 86:207–225.

Hickok, G. and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92.

Lieberman, P. (2000). *Human Language and our Reptilian Brain*. Harvard University Press.

Oudeyer, P.-Y. (2006). *Self-organization in the Evolution of Speech*. Oxford University Press.

Pulvermuller, F. (2002). *The Neuroscience of Language*. CUP.

Saunders, J., Lyon, C., Nehaniv, C. L., Dautenhahn, K., and Forster, F. (2009). A Constructivist Approach to Robot Language Learning via Simulated Babbling and Holophrase Extraction. In *IEEE ALife09 Conference*.

Saur, D. and Kreher, B. W. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences*, 105(46).

Steels, L. (2007). The Recruitment Theory of Language. In Lyon, C., Nehaniv, C., and Cangelosi, A., (Eds.), *Emergence of Communication and Language*. Springer.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.