CASA MILA: Cross-cultural and social aspects of multimodal interactions in language acquisition

Paul Vogt
Tilburg center for Creative Computing, Tilburg University
P.O. Box 90153, NL-5000 LE Tilburg
http://www.paul-vogt.nl, science@paul-vogt.nl

Abstract

This poster introduces the recently started CASA MILA project, which aims to study cross-cultural and social aspects of multimodal interactions aimed to establish joint attention and their impact on language development with infants and artificial agents. One of the objectives of the study is to collect data on the usage frequencies of three different types of joint attention and feed these into a simulation of the Talking Heads experiment in order to test mechanisms that would underlie the learning of word-meaning mappings. The poster will present some empirical findings from the pilot project that is currently been carried out in Mozambique.

1. Introduction

One of the biggest problems humans face when learning a language is identifying the meaning of words. The extent of this problem has been famously illustrated by (Quine, 1960), who sketched the situation of an anthropologist studying a – to him – unknown language. When a native speaker exclaims 'Gavagai!' at the moment a rabbit scurries by, the anthropologist notes that 'gavagai' means *rabbit*, but how can he be sure? Gavagai, Quine argued, could mean an infinite number of things, such as *undetached rabbit parts*, *dinner*, *animal with large ears* or even a completely unrelated event such as *it's going to rain*.

Humans, especially children, are notoriously good at solving this problem. Various biases, constraints and (social) mechanisms have been proposed trying to explain how humans acquire word-meaning mappings. Examples include the whole object bias, shape bias, taxonomic bias, mutual exclusivity, principle of contrast, Theory of Mind and joint attention; for an overview see, e.g., (Bloom, 2000). All these biases, constraints and mechanisms serve to reduce the uncertainty of a word's meaning.

The recently started CASA MILA project aims to study the effect of joint attention on language acquisition in different cultural societies and simulated robots. In particular, the objective is to investigate how the usage-frequencies of various multimodal interactions (e.g., pointing gestures, gaze following, etc.) between infants and caregivers affect the speed of word learning. The multimodal interactions, which the project focuses on, are used to establish one of the three forms of joint attention proposed in (Carpenter et al., 1998):

Sharing / checking attention is the first form that emerges around the age of 9-10 months in a child's development and occurs when a caregiver follows a child's attention to an object, while both are aware of sharing attention (the child looks back and forth from object to the caregiver).

Following attention emerges second and happens around 10.5 months when a child's attention is drawn to an object by the caregiver (e.g., through eye-gaze following).

Directing attention emerges thirdly at the average age of 12.6 months when a child directs the attention of the caregiver to an object.

Various studies have suggested that the onset of these types of attention, as well as, the frequency of joint attentional usage have an effect on the early vocabulary development of infants (Carpenter et al., 1998, Mundy et al., 2007).

One approach to study the effects of joint attentional usage on language development is by using developmental robotics, realised either in physical systems or in simulations. A recent study that used a simulation of Steels' Talking Heads experiment (Steels et al., 2002) has shown that differences in the use of the three joint attentional mechanisms can lead to strong differences in vocabulary development, assuming a statistical language learning mechanism (Kwisthout et al., 2008). In this model the word-meaning mappings are acquired through *cross-situational learning* (Siskind, 1996), which is a statistical learning mechanism based on the co-variance in the occurrences of words and meanings across situations (or learning contexts). In the model it is

assumed that the three joint attentional mechanisms have different ways to reduce the learning context size. It was shown that, relatively speaking, checking attention yielded the largest reduction, following attention the second largest reduction and directing attention the smallest. Since cross-situational learning works faster when the learning mechanisms are smallest (Smith et al., 2006), the same ordering was found for the speed of vocabulary development (Kwisthout et al., 2008).

In this study, however, the frequencies by which the different joint attentional mechanisms were used was all or nothing. This is very unrealistic, because humans tend to use these mechanisms in various frequencies that differ individually (Mundy et al., 2007), and possibly cross-culturally as well (Keller et al., 2005). In order to computationally verify the validity of the underlying language learning mechanisms and the influence that joint attention can have on language development, it is desirable to predict the speed of vocabulary development using empirically obtained data on joint attentional usage and compare the outcome with relating development with human children (Vogt and de Boer, 2009).

The CASA MILA project aims to collect such empirical data in three cultures: one urban and one rural Changana speaking culture from Mozambique, and a Dutch speaking culture. The study will involve an observational study in which infants are videotaped in a natural setting in their native environment. The purpose is to collect the frequency distributions with which multimodal interactions occur with caregivers, siblings and others that lead to the three joint attentional forms at various stages during their development between the ages of 9 to 24 months. It is anticipated that there are differences between the three cultures regarding the frequencies with which the different forms of joint attention are used. The question remains whether such differences are also found in the speed of vocabulary development. By closely monitoring their language development, it will be possible to correlate the differences in joint attentional use with the development of joint attention.

The empirically obtained frequency distributions will then be used as input to an adaptation of the computational model used in (Kwisthout et al., 2008) that simulates the acquisition and evolution of language to investigate the effects that different distributions have on language development. Such simulations are helpful to investigate whether the learning mechanism used in the computer model predicts a similar development as the empirical findings. If this is the case, then the investigated learning mechanism is a likely candidate for the mechanism used by humans. If not, the implemented learning mechanism probably needs revision.

Acknowledgements

The CASA MILA project is funded by the Netherlands Organisation for Scientific Research (NWO) through a VIDI grant awarded to the author.

References

Bloom, P. (2000). *How Children Learn the Meanings of Words*. The MIT Press, Cambridge, MA. and London, UK.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., and Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63(4).

Keller, H., Voelker, S., and Yovsi, R. (2005). Conceptions of parenting in different cultural communities: The case of West African Nso and Northern German women. *Social Development*, 14:158–180.

Kwisthout, J., Vogt, P., Haselager, P., and Dijkstra, T. (2008). Joint attention and language evolution. *Connection Science*, 20:155–171.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., and Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child development*, 78:938–954.

Quine, W. V. O. (1960). *Word and object*. Cambridge University Press.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61:39–91.

Smith, K., Smith, A., Blythe, R., and Vogt, P. (2006). Cross-situational learning: a mathematical approach. In Vogt, P., Sugita, Y., Tuci, E., and Nehaniv, C., (Eds.), *Symbol grounding and beyond*. Springer.

Steels, L., Kaplan, F., McIntyre, A., and Van Looveren, J. (2002). Crucial factors in the origins of word-meaning. In Wray, A., (Ed.), *The Transition to Language*, Oxford, UK. Oxford University Press.

Vogt, P. and de Boer, B. (2009). Language evolution: Computer models for empirical data. *Adaptive Behavior*. to appear.