# Chapter 11

# Emergence of linguistic and cognitive representations

**Timo Honkela, Krista Lagus, Tiina Lindh-Knuutila, Matti Pöllä, Juha Raitio, Sami Virpioja, Jaakko J. Väyrynen and Paul Wagner**

## 11.1 Introduction

Computational Cognitive Systems group conducts research on artificial systems that combine perception, action, reasoning, learning and communication. This area of research draws upon biological, cognitive and social system approaches to understanding cognition. Cognitive systems research is multidisciplinary and benefits from sharing and leveraging expertise and resources between disciplines. For example, statistical machine learning, pattern recognition and signal processing are central tools within computational cognitive systems research. Our research focuses on modeling and applying methods of unsupervised and reinforcement learning.

The general aim is to provide a methodological framework for theories of conceptual development, symbol grounding, communication among autonomous agents, agent modeling, and constructive learning. We have also worked in close collaboration with other groups in our laboratory, for instance, related to multimodal interfaces.

In the following chapters of this report, we describe the main results gained during 2006-07 in the four main thematic areas of the computational cognitive systems research group: *emergence of linguistic and cognitive representations*, *learning social interactions between agents*, *learning to translate* and *knowledge translation and innovation using adaptive informatics*. Each of these areas is described in a section of its own except for emergence of linguistic and cognitive representations that is a shared research area with Multimodal Interfaces research group and the results are mainly described in the chapter on Natural Language Processing. In the following, this area is discussed briefly. In general, the group has benefited strongly from the closeness and the collaboration with other research groups in the center. A notable example of such collaboration is the development of a *speech-to-speech machine translation system protytype*, developed in collaboration with the Multimodal Interfaces research group.

## 11.2 Research on emergence

The research on emergent linguistic and cognitive representations enables computers to deal with semantics: to process data having certain access to its meaning within multimodal contexts. Our focus is on the analysis and generation of conceptual structures and complex meanings.

The emergence of representations can be considered to consist of the following interrelated tasks: the discovery of elements of representation (e.g. words, morphemes, phonemes), their meaning relations (syntax and semantics), and structures or "rules" of their use in natural utterances (syntax and pragmatics).

An example of work on the first topic is the study of unsupervised machine learning techniques for finding the optimal segmentation of words into sub-word units called morphs, with the intent of finding realizations of morphemes (see Section 10.1 for details). Another example of the use of independent component analysis in discovering meaningful syntactic and semantic features for words (see Section 10.1 for details).

We have studied the emergence of linguistic representations through the analysis of words in contexts using the Independent Component Analysis (ICA). The ICA learns features automatically in an unsupervised manner. Several features for a word may exist, and the ICA gives the explicit values of each feature for each word. In our experiments, we have shown that the features coincide with known syntactic and semantic categories. More detailed description of this research is given in the section on Natural Language Processing in this report.

## 11.3 Events and projects

An important part of our activity has been the active role in organizing international scientific events. The main activity in 2006-2007 was the organization of the Scandinavian Conference on Artificial Intelligence [1].

In addition to the research areas discussed above, we continue to use the Self-Organizing Map (SOM) where applicable. The most important piece of research based on the SOM during this period was an analysis of conducted for the Academy of Finland. As a continuation to an earlier manually conducted qualitative analysis, the Academy of Finland, one of the country's largest funding agencies, commissioned a study to investigate whether text mining based on the Self-Organizing Map could be used to support assessment of the applications. A collection of 3224 applications was analyzed [2]. A collection of 1331 term candidates was extracted automatically. The 3224 application documents were encoded as term distribution patterns. The SOM algorithm organized the documents into a map in which similar applications are close to each other and in which thematic areas emerged.

Many parts of the research in this new group's agenda have been started during the reported period. Therefore, so far the results have mainly been reported in conferences and as technical reports. However, in early 2008 we were pleased to receive news about four accepted journal papers. Some of the related results are described already in this report, based on publications that have appeared during 2006 and 2007.

Some of the research activities by the group take place in projects funded by Tekes and EU Commission. The most notable examples are the projects Kulta (using adaptive informatics methods to model and simulate changing needs of consumers) and MeIEQ (developing quality labeling methods for medical web resources), discussed in some detail in the section on Knowledge translation and innovation using adaptive informatics.

## References

[1] T. Honkela, T. Raiko, J. Kortela and H. Valpola (eds.) (2006). *Proceedings of SCAI'06, the Ninth Scandinavian Conference on Artificial Intelligence*. Finnish Artificial Intelligence Society, 239 p. Espoo.

[2] T. Honkela, M. Klami (2007). Text mining of applications submitted to the Academy of Finland [in Finnish], unpublished report. Helsinki University of Technology, Espoo.