

# **Emergence of Linguistic Features: Independent Component Analysis of Contexts**

Timo Honkela<sup>1</sup>, Aapo Hyvärinen<sup>2</sup> and Jaakko Väyrynen<sup>1</sup>

<sup>1</sup>Neural Networks Research Center, Helsinki University of Technology

<sup>2</sup>HIIT Basic Research Unit, Department of Computer Science, University of Helsinki  
timo.honkela@hut.fi

Independent component analysis (ICA) (Hyvärinen et al., 2001) applied on word context data gives distinct features which reflect syntactic and semantic categories. The analysis gives features or categories that are both explicit and can easily be interpreted by humans. This result can be obtained without any human supervision or tagged corpora that would have some predetermined morphological, syntactic or semantic information. The results include both an emergence of clear distinctive categories or features and a distributed representation. This is based on the fact that a word may belong to several categories simultaneously in a graded manner. Within cognitive linguistics we wish that our model can provide additional understanding on potential cognitive mechanisms in natural language learning and understanding. Our approach is based on the assumption that much of the linguistic knowledge is emergent in nature and based on specific learning mechanisms.

In the full paper, we will provide a brief introduction to independent component analysis, review the earlier published results (Honkela and Hyvärinen, 2004), provide the very latest results, compare our model with some other approaches based on unsupervised learning paradigm and discuss in general the potential of the ICA model within the theme of modelling language, cognition and action.

## **References**

Hyvärinen, A., Karhunen, J. and Oja, E. Independent Component Analysis. John Wiley & Sons, 2001.

Honkela, T. and Hyvärinen, A. Linguistic Feature Extraction using Independent Component Analysis. Proceedings of IJCNN'04, 2004 (in print).