# Layer Structures and Conceptual Hierarchies in Semantic Representations for NLP

Hermann Helbig, Ingo Glöckner, and Rainer Osswald

Intelligent Information and Communication Systems
FernUniversität in Hagen, Germany

**Abstract.** Knowledge representation systems aiming at full natural language understanding need to cover a wide range of semantic phenomena including lexical ambiguities, coreference, modalities, counterfactuals, and generic sentences. In order to achieve this goal, we argue for a multidimensional view on the representation of natural language semantics. The proposed approach, which has been successfully applied to various NLP tasks including text retrieval and question answering, tries to keep the balance between expressiveness and manageability by introducing separate semantic layers for capturing dimensions such as facticity, degree of generalization, and determination of reference. Layer specifications are also used to express the distinction between categorical and situational knowledge and the encapsulation of knowledge needed e.g. for a proper modeling of propositional attitudes. The paper describes the role of these classificational means for natural language understanding, knowledge representation, and reasoning, and exemplifies their use in NLP applications.

## 1 Introduction

Although envisaged since the early days of automated natural language processing (NLP), there are currently only a few implemented systems that aim at a full semantic analysis of unrestricted natural language. One of the reasons for this situation may be the diversification in formal semantics with its highly elaborate but specialized theories focusing on specific semantic phenomena such as presuppositions, generics, pluralities or modalities (see e.g. [1]), whereas building language understanding systems calls for a uniform and concise formalism covering all aspects of natural language semantics up to a certain degree of granularity. A second reason may be the current dominance of shallow NLP even in areas where traditionally deep semantic analysis has been taken as a *sine qua non*. For instance, many approaches to open-domain textual question answering rest mainly on text retrieval techniques with no or little analysis of the underlying documents. However, this approach will fail if the relevant information is mentioned only once in the text collection, has no lexical overlap with the question and uses other syntactic constructions, or is distributed over several sentences linked by anaphora.

In this paper, we argue for a concept-centered representation of natural language semantics that employs a moderate amount of reification and represents semantic aspects like plurality or facticity by ontological features whenever appropriate. The proposed approach is explicated by a slightly simplified version of the *MultiNet* knowledge representation formalism, which is described in detail in [2]. MultiNet has been designed