Semantic Tagging and NLP Applications

Thierry Declerck & Judith Klein
DFKI GmbH
(German Research Center for Artificial Intelligence)
{declerck;klein}@dfki.uni-sb.de

There are hardly any annotation schemes including semantic information, with the exception of Princeton WordNet (which will be extended by EuroWordNet for European languages). But some projects already addressed this topic, like FraCaS (Framework for Computational Semantics), or are starting to do this, like DIET (Diagnostic and Evaluation Tools for NL Applications, an extension of the TSNLP framework, see Lehmann et al., Coling 96)[1].

What makes semantic tagging appealing is, among others, the (justified) hope that it will contribute to the improvement of the performances and the robustness of NLP systems. Besides this aspect, evaluation will also benefit from semantically tagged test corpora.

In this working session, we focus on both the question how semantic tagging can support the development of NLP applications and, the other way round, how NLP systems can support semantic tagging. Among the different NLP projects making a (limited) use of semantic annotations, we are aiming at common annotation methodologies beyond particular approaches.

As an example we describe a scenario which has been adopted within the context of a NLP project concerned with appointment scheduling (COSMA, see Busemann et al., ANLP 97), where information extraction techniques combined with a shallow-parsing strategy (see Neumann, ANLP 97) have been used in order to process just the relevant fragments of input texts.

To support the development of the system and to delimit the linguistic coverage of the NLP application, a small corpus has been semantically hand-tagged, where the semantic annotations have been added to the mainly syntactic annotation scheme of the TSNLP framework. Thus the evaluation tool of TSNLP has been extended by a certain class of semantic information (non-ambiguous temporal expressions). Furthermore the FST automata developed for the purpose of message extraction have been designed along the lines of this annotation scheme. And the output of the FST automata has been defined in such a way that they can be used for an automatic (rule-based) semantic annotation of new text input (the annotation being limited to the temporal expression).

Other NLP applications could reuse such a simple annotation in order to determine, for example, selectional restrictions or text classifications. Addressing the reusability of annotation schemes for particular domains, one will have to consider if they can be just added to existing morpho-syntactic annotation schemes, as we described in the example above, or if the annotation work should be started from scratch, which could be necessary for more complex applications. A recently developed annotation scheme (see Skut et al., ANLP 97) is proposing an architecture with multiple levels of linguistic representation, for argument structure, grammatical function and syntactic category. We will investigate how semantic information can be integrated in such a framework and if the bidirectional interface between semantic tagging and NLP system, described above, can be adopted to this architecture.

[1] We just mention here some projects funded by the CEC.