

# **Ontology-driven disambiguation of syntactic and semantic ambiguities in GenIE**

**Philipp Cimiano Lavin<sup>1</sup>, Uwe Reyle<sup>1</sup>, Isabel Rojas<sup>2</sup>, Jasmin Saric<sup>2</sup>, Heike Zinsmeister<sup>1</sup>**

(1) Institut für Maschinelle Sprachverarbeitung, IMS, University of Stuttgart, Germany

(2) European Media Laboratory (EML), Heidelberg, Germany

The overall objective of GenIE is to develop a system that automatically extracts information about biochemical pathways (information about genomes and proteins, as well as their sequences, structures and functions) from free text sources (research papers, (Medline-) abstracts, comments found in existing databases) and uses this information to fill a biochemical database.

GenIE's extraction mechanism is based on a new model of NL-interpretation. This model allows syntactic and semantic representations to be underspecified but nevertheless interact at any stage of the extraction process. Furthermore it incorporates reasoning on underspecified forms. This allows morphological, syntactic, semantic and ontological information to mutually constrain each other and hence to eliminate ambiguities and to correctly identify the intended interpretation.

Ambiguity is ubiquitous and one of the major problems in NLP. It occurs at the level of individual words, in certain syntactic constructions, meaning assignment and also contextually in determining referents for anaphoric expressions and implicit arguments as well as the logical relationships between sentences. Each type of ambiguity requires its own representation formalism and has its own disambiguation criteria. We explain this by discussing typical examples of these ambiguity problems and show how they resolved in GenIE.