Ontology Driven Resolution of Bridging References *

Philipp Cimiano

Institute for Computational Linguistics,
University of Stuttgart
cimianpp@ims.uni-stuttgart.de

1 Introduction

This paper presents a formal DRT-based approach to bridging reference resolution in line with Van der Sandt’s analysis of presuppositions as anaphora ([11]). It is novel because it makes use of an ontology as declarative representation formalism of world knowledge and as backbone of the inference mechanism driving the bridging reference resolution. Within this paper, bridging will be understood in line with Asher et al. ([1]) as the phenomenon of a linguistic expression introduced in a text referring to a certain antecedent and both being related in a way which is not explicitly stated. Here follow two examples of bridging out of biochemical texts, which is the domain the author concentrated on 1:

(1) (Rhombotin 2) Binds to the basic helix-loop protein TAL-1.
   <This interaction> is critical for the regulation of red blood cell development.

(2) (LBP) Binds to the lipid A moiety of bacterial lipopolysaccharides (LPS). [...] <The LBP/LPS complex> seems to interact with the CD14 receptor.

In both cases the referring definite description is related to the antecedent in a way which is not explicitly stated. The relations are respectively Identity and Result ([7]).

*Thanks to U. Reyle for comments on the first versions of this paper.
1In the examples given in this paper, the referring expression will be enclosed in brackets, while the antecedent will appear underlined.
2 The Ontology Driven Approach

An ontology is a specification of a conceptualization, which has to be understood as an abstract representation of the domain we want to model for a certain purpose ([3]). Formally, an ontology will be defined as follows:

**Definition 1 (Ontology)**

An ontology $O$ is a triple $(C,T,D)$ where $C$ is a set of predicates representing concepts, $T$ is a set of first order logic formulas representing taxonomic relations between the concepts in $C$ and $D$ is a set of first order logic formulas defining the concepts in $C$ in terms of necessary conditions.

In order to reason with such an ontology, a notion of implication between DRSs has been defined by making use of the DRS calculus presented in [6]:

**Definition 2 (Implication between DRSs)**

A DRS $K$ implies a DRS $K'$ with regard to a DRS $K_1 \implies K_2$ ($K \implies [K_1 \implies K_2]$) iff $K \oplus K_1 \implies K_2 \oplus K'$ follows logically from $K \oplus K_1 \implies K_2$ as defined by the Generalized Modus Ponens (GMP) in [6] and it is not the case that $K'$ follows from $K$ alone ($K \not\implies K'$).

In order to exploit the taxonomic relations $T$ defined within the ontology, a notion of specialization between DRSs has been introduced. In this sense a DRS $K_1$ will be a specialization of a DRS $K_2$ if $K_1$ represents an ontological concept which is more special than the one represented by $K_2$:

**Definition 3 (Specialization)**

A DRS $K_1$ is a specialization of a DRS $K_2$ with regard to an ontology $O=(C,T,A)$ ($K_1 <_O K_2$) iff $\exists$ DRS $K_1 \implies K_2$ and $[K_1 \implies K_2]_{PL_1} \in T$ and $K_1 \implies [K_1 \implies K_2]_{PL_1} \in D$ and $K \implies [K_1 \implies K_2]$ $K_2$, where $[K]_{PL_1}$ is the translation of the DRS $K$ to predicate logic as defined in [5].

In the following, $<_O$ will denote the reflexive and transitive closure of $<_O$. Now it can be defined what it means for a DRS $K'$ to follow logically from $K$ with regard to an ontology $O$. Intuitively, a DRS $K'$ will follow from $K$ with regard to the ontology if there is a DRS $K''$ such that $K$ is more special than $K''$ and $K'$ follows from $K''$ by a conceptual definition in $D$:

**Definition 4 (Implication with regard to an ontology)**

A DRS $K$ implies a DRS $K'$ with regard to an ontology $O=(C,T,A)$ ($K \implies _O K'$) iff $\exists$ DRS $K_1 \implies K_2$ such that $[K_1 \implies K_2]_{PL_1} \in D$ and $K \implies [K_1 \implies K_2]$ $K'$ or $\exists$ DRS $K''$ such that $K \leq_O K''$ and $K'' \implies _O K'$.
Though the inference mechanism is a very simple one, an obvious advantage is that reasoning complexity is in the worst case $O(|T|^2)$, i.e. quadratic in the number of taxonomic relations of the ontology. Furthermore, bridging reference resolution can be made determinate without introducing costs for accommodation as in [4] or using an additional proof system as in [9], but simply by choosing the shortest path between two concepts with regard to the tree representing the ontology.

Following Van der Sandt and Bos et al., unresolved anaphoric expressions will be represented by $\alpha$-marked DRSs and merged with the discourse processed so far before being resolved. Resolution can take place in three ways (in this preference order): linking, bridging and accommodation ([2]). Due to space limitations, in this paper only a rather informal definition of the bridging-operation is given (the interested reader is referred to [8] for a formal definition as well as a detailed resolution of the examples presented):

**Definition 5 (Bridging)**

An $\alpha$-marked DRS $K_\alpha$ will be interpreted as a bridging reference to an accessible DRS $K_1$ ($K_\alpha \leq K_1$) if $K'_1$ is an ontological generalization of $K_1$ ($K_1 \leq^O K'_1$) and furthermore $K_\alpha$ is suitable to $K'_1$, where suitability is defined as in [8], or if $K'_1$ follows logically from $K_1$ with regard to the ontology ($K_1 \Rightarrow^O K'_1$), $K_\alpha$ is a specialization of $K'_\alpha$ ($K_{\alpha} \leq^O K'_{\alpha}$) and furthermore $K'_\alpha$ is suitable to $K'_1$.

The first part of the disjunction captures the intuition that certain expressions can be referred to later in the discourse in a (ontologically) more general way, such as in example 1. The second part accounts for example 2, where the referring expression can be linked to an antecedent inferred from an accessible DRS.

### 3 Evaluation and Conclusion

The approach described has been implemented in Prolog within the GenIE system ([10]). A small ontology of 129 concepts has been developed for the domain of molecular biology. Furthermore, a training and a test corpus each consisting of 250 short texts describing the function of a protein have been annotated by hand with DRSs representing ontological concepts. The results of the bridging reference resolution approach on the test corpus measured against the annotation of three different subjects were a recall of 54.54% and a precision of 90.98% ([8]). These results look very promising and show that given a suitable ontology the approach is actually scalable.
References


