

NLP-Approaches to Reference Resolution

Michael Strube
European Media Laboratory GmbH

The last decade has seen major progress in the area of reference resolution, in particular with respect to coverage and robustness of the approaches and their evaluation. However, the work published was mostly restricted to the ‘easy’ cases, i.e. to pronouns, for which simple algorithms exist, and to those definite NPs which share some (sub)string with their antecedent. It was shown that these cases constitute only a fraction of all referring expressions.

Most current NLP applications which employ reference resolution (e.g. IR and IE systems, text summarization systems, etc.) do not require sophisticated reference resolution components. However, the next generation of NLP applications (e.g. robust dialogue systems) will definitely need them.

The tutorial consists of three parts:

1. A survey of work on pronoun and definite NP resolution will be given presenting both linguistic and corpus-based approaches and heuristics. It will be shown that most of these approaches share major underlying assumptions. The most important factors for pronoun and definite NP resolution will be identified.
2. The tutorial will continue with an overview of the remaining linguistic phenomena: bridging expressions, demonstrative pronouns, metonymies, etc. Several corpus studies have shown that, depending on the domain, these cases constitute a substantial fraction of all referring expressions. Also, spoken and multimodal dialogue shows some particularities which prevent the straightforward application of algorithms developed for reference resolution in written texts. A survey of existing approaches and systems will be given. A major problem in dealing with the difficult cases is the lack of annotated corpora. The tutorial will present annotation tools, suggest guidelines for annotating data and achieving high inter-annotator reliability so that corpus-based approaches may be employed in the future even for the difficult cases.
3. The tutorial will conclude with a proposal on how to proceed. A list of phenomena with increasing complexity will be presented which can be used as a guideline for the development of next generation NLP systems.