Related Work sec:Related Work

Anaphora resolution systems emerged into two different strategies. First of all, there are rule-based techniques which focus more on theoretical considerations. The second strategy uses machine learning and is based on annotated data. The following chapter will briefly present both and discuss their advantages and disadvantages, followed by exemplary implementations and approaches. Especially the handling of gender information is of interest. Since anaphora resolution is a subtask of coreference resolution, coreference resolution systems will be considered as well.

Rule-Based Approaches Rule-based techniques rely on manual understanding and implementation of syntactic and semantic principles in natural language kennedy1996anaphora,mitkov1994integrated,ingria1989computational. Clues that could be helpful for antecedent identification are manually implemented as rules. To identify relevant clues, prior knowledge about linguistic principles (such as binding principles) is necessary. Since rules might be domain-specific, the implementation would most likely be worse on other domains. Refinements for different domains would make the development even more complex and time-consuming. Nevertheless, rule-based techniques are much more transparent in contrast to machine learning. In the last section, a comparing evaluation of both techniques will be presented.

The Naive Hobbs algorithm The Naive Hobbs algorithm described by hobbs1978resolving relies on parsed syntax trees containing the grammatical structure. Put simply, the tree containing the anaphora is searched left-to-right with breadth-first search and the algorithm stops when a matching noun phrase is found. Noun phrases mismatching in gender or number are neglected. Information on gender could only be selected through surface clues. The algorithm also limits the list of possible antecedents, as for instance the antecedent cannot occur in the same non-dividable noun phrase. As long as no matching antecedent is found, the preceding sentence will be searched successively.

High Precision Pronoun Resolution baldwinCogNIAC Another rule-based approach was presented by baldwin1997cogniac with CogNIAC, a high precision pronoun resolution system. It only resolves pronouns when high confidence rules (shown in Table table:cogniacRules) are satisfied in order to avoid decisions under ambiguity and to ensure, that only very likely antecedents are attached (high precision). This might lead to a neglect of less probable but still correct antecedents and could therefore lower the recall score.

table[h] CogNIAC core rules tabular— l —p8cm — Rule Description

Anaphora Resolution with Limited Knowledge anaphoraLimitedKnowledgeSection

A domain independent approach by mitkov1998robust tried to eliminate the disadvantages of previous rule-based systems. Mitkov renounced complex syntax and semantic analysis in order to keep the algorithm as less domain specific as possible. Only a part-of-speech tagger and a simple noun phrase identifitcation module were applied. The algorithm was informally described by Mitkov in three steps: enumerate

E xamine the current sentence and the two preceding sentences (if available). Look for noun phrases only to the left of the anaphora

S elect from the noun phrases identified only those which agree in gender and number with the pronominal anaphora and group them as a set of potential candidates

A pply the antecedent indicators to each potential candidate and assign scores; the candidate with the highest aggregate score is proposed as antecedent