Leveraging Different Meronym Discovery Methods for Bridging Resolution in French

Emmanuel Lassalle and Pascal Denis
Alpage, INRIA & Univ. Paris Diderot
Sorbonne Paris Cité, F-75205 Paris, France
emmanuel.lassalle@ens-lyon.org, pascal.denis@inria.fr

Abstract. This paper presents a statistical system for resolving bridging descriptions in French, a language for which current lexical resources have a very low coverage. The system is similar to that developed for English by [22], but it was enriched to integrate meronymic information extracted automatically from both web queries and raw text using syntactic patterns. Through various experiments on the DEDE corpus [8], we show that although still mediocre the performance of our system compare favorably to those obtained by [22] for English. In addition, our evaluation indicates that the different meronym extraction methods have a cumulative effect but that the text pattern-based extraction method is more robust and leads to higher accuracy than the Web-based approach.

Keywords: bridging anaphora resolution, relation extraction, syntactic patterns.

1 Introduction

Bridging descriptions are a special kind of anaphora whose interpretation requires not only identifying an antecedent, but also inferring a specific relation linking it to the anaphor [6]. In this paper, we focus on merological bridging anaphora (that is, cases wherein the inferred relation is a part-whole relation). An illustrative English example is given in (1):

(1) The car will not move. The engine is broken.

The resolution of bridging anaphora currently represents one of the most challenging tasks in discourse processing. It is considerably much harder than standard coreferential anaphora resolution for which shallow predictors (like distance, string matching, or morphosyntactic agreement) have been shown to be rather effective [15,25,18]. Part of the challenge is due to an important information bottle-neck. Lexical resources like WordNet are still too poor and uneven in coverage to provide a realistic solution [19]. In turn, more recent approaches to bridging resolution have turned to web-based extraction methods to fill this “knowledge gap” [4,22]. To date, the most complete and best-performing approach combines focus and lexical distance predictors using machine learning techniques [22].
While there has been a lot of research on resolving bridging anaphora in English [19][28][4][22], much fewer work exist for other languages (see [27] for an attempt on German). In this paper, we develop a system that performs bridging resolution in French; to our knowledge, this is first such system for this language. Note that languages other than English, the knowledge gap is even wider, since lexical resources are typically scarcer. Our system directly builds upon the system developed for English by [22], and enriches it in several ways. First, we refine the search engine queries to include a wider range of meronym-holonym extraction patterns (some of them specific to French). Second, and more substantially, we augment the system with an original iterative pattern-based relation extraction method from raw text.

Our system is evaluated on DEDE [8], a corpus of French definite descriptions annotated for bridging. Initial experiments indicate that our system achieves performance that compares favorably to those of [22] for English under the same simple evaluation protocols. We found that the two meronym discovery approaches have a cumulative effect for resolution accuracy, and further comparison between the two methods suggests extraction from raw text is less noisy. But we also show that, when tested in a more realistic setting, the system still has rather low accuracy (at least for French). Our analysis reveals that the current features are still not discriminative enough, which in turns suggests the need for acquiring additional relational data.

The rest of this paper is organized as follows. Section 2 briefly presents the corpus we used. In section 3, we first describe a free lexical database for French, as well as our two methods for extracting meronyms. Section 4 details various experiments wherein we use these different resources for bridging resolution. A detailed comparison of the two extraction methods is given in section 5. Finally, section 6 discusses related work and section 7 offers some conclusions and directions for future work.

2 DEDE: A French Corpus for Bridging Descriptions

For this study, we use DEDE [8], a French corpus annotated with definite descriptions [8], built from Le Monde articles in 1987. The corpus is not only the first of its kind in French with a high agreement rate between annotators, but it also introduces a fine-grained classification of definite descriptions.

The corpus contains 4,910 descriptions classified in 5 main categories (autonomous, coreferential, associative, situational and non-referential) themselves further divided into subcategories, so that the corpus provides an appropriate resource for evaluating a complete system for processing definite descriptions. As we are solely interested in bridging resolution for now, we only detail the “associative” class (that is bridging).

The “associative” class contains 530 anaphors of 4 different types (see [8] for details): 322 MERO (meronymy e.g. a tree/the trunk), 87 CIRC (modifier-modified

1 http://www.cnrtl.fr/corpus/dede/