Anaphora, which describes the dependence of an expression on a previously mentioned expression, has been a central object of study in descriptive, theoretical and computational linguistics. It has long been recognized that anaphoric dependencies within sentences are an important source of insight into underlying syntactic structure (Chomsky, 1981; Reinhart 1983; Büring, 2004). Since the development of Discourse Representation Theory (DRT) (Kamp, 1981; Kamp & Reyle, 1993) and File Change Semantics, (Heim, 1982), anaphora across sentences has become a major focus in semantic theories. In computational linguistics, with its more practical orientation, there has been a focus on the problem of anaphora resolution algorithms: methods for determining the antecedent for a given anaphoric expression.

In recent years, computational work on anaphora has benefited greatly from the rapid development in the availability of data and statistical techniques. It has become clear that a wide range of practical applications can benefit from progress in anaphora resolution: machine translation, question-answer systems, text summarization or automatic abstracting, information extraction, language generation, and dialogue systems (Mitkov, 2002). At the same time, it has become clear that theoretical investigations in anaphora provide crucial underpinning to the development of automatic systems (Mitkov, 2003).

The papers collected in this special issue reflect both the practical issues in working towards automated systems, as well as the enduring preoccupation with anaphora’s theoretical basis.

1. In their contribution, Amsili, Denis and Roussarie, “Anaphores abstraites en français : représentation formelle”, discuss the constraints on reference to abstract entities such as events, situations, propositions and facts. They examine the discourse representation of these abstract entities, their resolution mechanisms and finally they look at the set of semantic-pragmatic constraints which are at issue in resolving several anaphoric links.

2. Boudreau and Kittredge´s paper “Résolution des anaphores et détermination des chaînes de coréférences. Differences entre variétés de textes”, is primarily concerned with the automatic identification of reference chains headed by a proper noun our a 3rd personal pronoun. They address the problem of partitioning a text’s referring expressions using a simple algorithm which does not exploit full syntactic parse. The work has been made from 60 texts with approximately 500 words for each of three different domain knowledge used.
3. In “Identifying non-referential it”, Boyd, Gegg-Harrison and Byron, describe a machine learning approach to identifying non-referential occurrences of the pronoun “it” in English. This is important in the development of pronoun resolution algorithms, to avoid them being inappropriately bound since in these cases there is no antecedent to find, and being potential antecedent of other pronouns.

4. Byron, Gegg-Harrison and Lee, in their paper “Resolving zero anaphors and pronouns in Korean”, address the problem of zero anaphors, where an empty expression is treated on a par with an overt pronoun. They rely on a version of Centering Theory, and report on a study performed on the Korean Treebank.

5. The work presented in Gardent and Manuélian’s “Création d’un corpus annoté pour le traitement des descriptions définies”, gives a new methodology for the development for annotated corpora designed to improve processes of anaphora resolution involving definite descriptions. The French corpus used contains 48 360 words and 4 910 definite descriptions. They handle the different usages of definite descriptions by means of an annotation scheme according to different categories (autonomous, co-referential, contextual and non-referential).

6. Gutiérrez-Rexach (“Setting up perspectives for demonstratives in discourse”) looks at demonstratives in Spanish and English, developing a theory in the framework of DRT, according to which demonstratives are quantifiers, subject to several restrictions. While Gutiérrez-Rexach takes a somewhat more theoretical perspective, his work is also informed by a corpus study.

7. In “Semantics-based resolution of fragments and underspecified structures”, McShane, Nirenburg and Beale, take a broad view of anaphora, pointing out that elliptical and fragmentary utterances can be viewed as anaphoric. They present an approach that is being implemented as part of an ambitious system, OntoSem. This system is intended to perform extensive morphological, syntactic and semantic analysis, giving rise to meaning representations that provide the basis for the resolution of fragments and other anaphoric structures. They emphasize their ambition to integrate their approach into a working system; at the same time, they are interested in the implications of their claims for theoretical debates about different forms of anaphora.

Joan Busquets
ERSS/CNRS UMR 510
Projet Signes (INRIA-Futurs), Université Bordeaux-3
busquets@u-bordeaux3.fr

Daniel Hardt
Center for Computational Modelling of Language
Department of Computational linguistics
Copenhagen Business School, Denmark
dh@id.cbs.dk
References


Comité de lecture spécifique

Pascal Amsili, Université Paris VII, France
Nicholas Asher, University of Texas, USA
Amit Bagga, Ask Jeeves Inc. USA
Antonio Branco, Facultade de Ciencias, Lisbon
Francis Corblin, Université Paris IV, France
Dan Cristea, University of Lasi, Romania
Denis Delfitto, University Venora, Italy
Richard Evans, University of Wolverhampton, UK
Robert Frank, John Hopkins University, USA
Claire Gardent, CNRS/LORIA, France
Ruslan Mitkov, University of Wolverhampton, UK
Rafael Muñoz, Universidad Alicante, Spain
Masaki Murata, NICT, Japan
Susanne Salmon-Alt, ATILF, LORIA France
François Trouilleux, Université Clermont II, France
Bonnie Webber, University of Edinburgh, Scotland