Which bridges for bridging definite descriptions?

Claire Gardent, Hélène Manuélian, Eric Kow

Which bridges for bridging definite descriptions?

Claire Gardent, Hélène Manuélian and Eric Kow
LORIA
BP239 Campus Scientifique
F-54506 Vandoeuvre-les-Nancy
FirstName.LastName@loria.fr

Abstract

This paper presents a corpus study of bridging definite descriptions in the french corpus PAROLE. It proposes a typology of bridging relations; describes a system for annotating NPs which allows for a user friendly collection of all relevant information on the bridging definite descriptions occurring in the corpus and discusses the results of the corpus study[1].

1 Introduction

It has long been known (cf. e.g., (Clark, 1977)) that the referent of a definite description can be implicitely related (through world or lexical knowledge) to some previously mentioned entity. In (1a) for instance, the referent of "the ceiling" is linked to the referent of its anchor "the room" through the "part-of" relation: because we know that rooms have ceilings, we interpret "the ceiling" in the second sentence, not just as denoting any ceiling, but rather as denoting the ceiling that is part of the room mentioned in the previous sentence.

(1) a. I looked into the room. The ceiling was very high.
b. I looked into the room. The duck was very yellow.

Note that the implicit relation must be supported by shared knowledge. This is illustrated by example (1b) where no relation between "duck" and the entities previously mentioned can be inferred. As a result, no anchor can be provided for the definite description ("the duck" is not interpreted as "the duck that is part of the room") and the overall discourse is distinctly odd.

For natural language processing systems, so-called "bridging" (or "associative") definite descriptions raise the interesting issue of how to integrate knowledge based reasoning so as to correctly interpret or generate them. In the analysis direction, the problem is to recover the missing relation between antecedent and anaphor so as to provide a full interpretation of the input text (assuming the antecedent has somehow been identified). In the generation process, the difficulty is to determine whether the relation that holds between the referent of the definite description and the referent of its antecedent is one that can be ommitted i.e., one that is supported by human knowledge.

In both cases, knowing the spectrum of relations that are possible between a definite description and its antecedent is essential. Without a finite set of relations to start with, a tractable treatment of the inferences involved is unlikely.

For the development of computational systems, it is also necessary to know where the implicit bridging relation "comes from": is it a lexical relation (e.g., meronymy, hyponymy, synonymy) whose encoding is part of tools such as WordNet? Is it given by world knowledge? etc.

1. Héléne Manuelian gratefully acknowledges the financial support provided for part of this research by both INRIA and the Lorraine Region (CPER Ingenierie logicielle, Pole "Ingéniérie des langues, du document et de l'information scientifique, technique et culturelle").

In this paper, we focus on identifying the set of relations that can hold between a bridging definite description and its antecedent. Starting from the literature, we propose a typology for these relations which we then validate on a medium size corpus (9500 definite descriptions of which roughly 400 are bridging definite descriptions). Further, we investigate WordNet (Miller, 1995) and FrameNet (Baker et al., 1998) and quantify the number of cases where these resources actually contain the relation used between the ontological types of the antecedent/anaphor pair. We conclude with some indications of which types of lexical resources are needed in order to interpret and/or generate bridging definite descriptions as well as pointers for further research.

2 A typology of bridging relations

In this section, we start by summarising the bridging relations identified in the literature. We then explains why it cannot be used directly as a basis for annotation and go on to propose a taxonomy of bridging relations which we believe, can reliably be used for large scale annotation. We use the following terminology. The TARGET is the referent of a bridging definite description which is related by some implicit BRIDGING relation to the referent of some previously mentioned entity, the ANCHOR.

2.1 Bridging relations in the literature

As e.g., (Clark, 1977; Strand, 1997; Kleiber, 1997) have shown the semantic nature of this bridging relation can vary. The following examples summarise the various bridging possibilities identified in the literature.

(2) a. Investment/Two thirds **Set/Subset**
b. Class/Student **Set/Element**
c. Murder/Murderer **Event/Argument**
d. Club/President **Individual/Function**
e. Bicyle/Price **Individual/Attribute**
f. Room/Ceiling **Whole/Part**
g. Cake/Slice **Whole/Piece**
h. Suitcase/Leather **Individual/Stuff**
i. Forest/Tree **Collection/Member**
j. France/French waters **Place/Area**
k. Opera/Duet **Whole/Temp.Subpart**
l. Village/Taxi drivers **Location/Object**
m. Today/The news **Time/Object**

When either the anchor or the target is a set, the related object can stand in a **subset-** or an **element-of** relation.

When one of the related individuals is an event, the other individual can describe an **argument** of that event (Clark distinguishes here between obligatory and optional arguments).

When both anchors and targets are objects, one of the related object can either fulfill a certain **function** in the other or specify one of its **attributes**. The related objects can also stands in one of the several meronymic relations described by (Winston et al., 1987). A **part** is a structurally or functionally motivated component of a structured whole. Clark further distinguishes between necessary (room-/ceiling), probable (room/window) and inducible (room/chandelier) parts. **Pieces** differ from parts in that pieces are homoreneous (all pieces are similar to each other and to their whole). A **collection** differs from a set in that the group is based on a spatial or social connection rather than on physical similarity. The **stuff** is the matter of which an individual is made and an **area** is a spatial subpart of a whole which cannot be separated from it.

Finally, an associative link can specify the **spatial** or **temporal** location of the related object.

2.2 Two requirements for a typology of bridging relations

While the set of relations identified in the literature is a useful starting point for defining a typology of bridging relations, it is insufficient to support the development of computational models of bridging definite descriptions for at least two reasons.

First, the proposed typology must provide an accurate semantics for bridging relations. As it stands, and as a first annotation pass quickly showed, the set of relations identified in the previous section does not fulfill this criteria. Consider for instance, the following anchor/target pairs found in our corpus:

(3) a. Operation/Convalescence **Following**
b. Athletism/National federation **For**

c. Question/Answer **to**

d. Investigation/Witness reports **Based on**

In these examples, the relation holding between anchor and target is determined by the lexical semantics of either target (examples 3a-c) or anchor (examples 3d). Thus, a convalescence *follows* an operation, a federation groups together associations acting *for* a common activity, an answer is a reaction *to* a question (or a request) and an investigation is *based on* witness reports.

This is in contrast with the standard view of bridging relations which tends to assume a closed set involving in essence three main types of relations: inclusion (whole-part, set membership, time and space anchor and locative anaphora), possession (subcategorisation, possessor-thing and functional associative anaphora) and thematic (event-argument). As we shall see, the typology we propose take these facts into account and encompasses both the "standard" bridging relations and the more unconventional "lexical" ones.

A second requirement that must be satisfied in order to develop a computational model of bridging definite descriptions, is that the possible sources of bridging relations be identified. Where does a bridging relation "come from"? Is it a lexical relation (meronymy)? Or is it given by a lexical definition, a thematic grid (Event/Argument cases) or by more extensive script- or world-knowledge? If (and as we shall see, this is indeed the case) all of these sources are involved, it is important to know in what proportion each of these sources is involved so as to assess (i) the level of processing difficulty involved in verifying the existence of, or in inferring a bridging relation and (ii) the proportion of cases which can reasonably be expected to be solved by current state-of-the-art methods and tools.

2.3 Proposal

The classification scheme we propose for bridging definite descriptions relates each class to a semantic relation, a source for the bridging relation and, additionally, some constraints on the basic ontological types involved (E stands for the set of eventualities i.e. states and events, I for the set of individuals, Loc and T for disjoint subsets of I denoting locations and time intervals respectively). This classification is summarised in table 1. We now discuss each class in more detail showing in particular, how a specific relation is identified and how its semantics is established.

Set membership. This class covers cases such as (4) where the target is either a member or a subset of a set (a group of similar individuals).

(4) a. Seminars/The last seminar

b. The CGT and the FO/The FO

c. The army/The third

The semantics of this bridging relation is set membership and inclusion. The anchor (or the target) must be a set of individuals and the target (or the anchor) must be an individual.

Thematic. As illustrated by (2c), the target can be related to the anchor via a thematic relation (a murderer is the *agent* of a murder). More generally, a thematic bridge links an event to an individual via a thematic relation defined by the thematic grid of the event. As a result, the property denoted by the noun caracterising the individual (P_i) must be subsumed by the conjunction of the property denoted by the verb or noun caracterising the event (P_e) and the thematic relation holding between this event and the individual($\theta(e,i)$). That is, $P_i \sqsubseteq P_e(e) \wedge \theta(e,i)$.

Definitional. In this case, the implicit bridging relation holding between anchor and target is given by the dictionary definition of either the target or the anchor. For instance, a "convalescence" (in 2, the target) can be defined to be the period "following" an "operation" (the anchor) or a disease so that in this case, the bridging relation between anchor and target is one of temporal succession.

In a definitional bridge, the definition usually imposes a selectional restriction which must be satisfied by the related object (anchor or target). The property declared (in the text) to hold of the related object must thus be subsumed by the property requested to hold of the related object by the definition.

The definitional category covers several of the cases identified in the literature each of them can

Class	Sem. Reln	Types of and Target and Anchor	Source
Set membership	$\in, \subset$	$\langle I, \mathcal{P}(I)\rangle$	Hyponymy
Thematic	Thematic roles agent, patient etc.	$\langle I, E\rangle$ (or $\langle E, I\rangle$)	Event Thematic Grid
Definitional			
Indiv./Attribute	As given by defn.	$\langle I, \textit{Features}\rangle$	Lexicographic defn
Associate/Indiv.	As given by defn.	$\langle I, I\rangle, \langle I, E\rangle$	Lexicographic defn
Meronymic relns	part of	$\langle I, I\rangle, \langle E, E\rangle$ $\langle I, \mathcal{P}(I)\rangle$	Meronymy
Co-participants	As given by defns of target and anchor	$\langle I, I\rangle, \langle I, E\rangle$ $\langle E, E\rangle$	Lex. Defns
Non lexical			
Circumstantial	spatial or temporal	$\langle I, Loc\rangle, \langle I, T\rangle$	Discourse structure
WKL	As given by WKL	Anything	World knowledge

FIG. 1 – *A typology for bridging relations*

be differentiated from the others by various ontological and semantic criteria.

Thus the **meronymic relations** (whole/Part, Whole/Piece, Individual/Stuff, Collection/Member, Place/area, Event/Subevent cf. (Winston et al., 1987)) are relations which can be expressed using the "part of" expressions. For two objects X and Y to be in a meronymic relation, it must be possible to say that X usually has Y *and* that Y usually is a part of X. The meronymic relation implies (spatial, temporal or abstract) inclusion and can only hold between entities of the same ontological types (individuals, events etc.). Following (Winston et al., 1987), we assume various types of meronymic relations (whole/Part, Whole/Piece, Individual/Stuff, Collection/Member, Place/area, Event/Subevent) – for a more precise definition of each of these relations, we refer the reader to (Winston et al., 1987). Additionally, we assume an INDIVIDUAL/FUNCTION meronymic relation (e.g., a club/the president) which involves a definitional bridge holding between individuals with one of the related individuals being described by his profession or function wrt the other.

Contrary to the meronymic relations, the two other types of definitional bridging relations do not imply inclusion but a simple implication relation (a teacher implies some audience, a surface implies an object etc.). More specifically, INDIVIDUAL/ATTRIBUTES pairs (e.g., a person/the age) involve a definitional bridge holding between individuals with one of the related individuals being a feature (i.e., something that takes a value within a finite domain) while the ENTITY/ASSOCIATE pairs may involve two individual, one individual and an event or between two events.

Co-participants. There are cases where the relation holding between target and anchor is given by the definitions of both the anchor and the target. For instance the pair "trip/seat" is related by the relation "in vehicle used for" which can be reconstructed from the definition of the target ("a seat is a place reserved for sitting in a vehicle or a room") and of the anchor ("a trip is a deplacement of persons by some means of transport). In such cases, the definitions of the target and the anchor involve two properties P_a and P_t which stand in a subsumption relation (here, "vehicle" is subsumed by "means of transport").

Non lexical. Finally, there are cases such as (5) where no amount of lexical knowledge will help and where the relation holding between target and anchor is given either by discourse structure (**circumstantial**) or by our knowledge of the world and of how things work (**WKL**).

(5) a. Grenoble/The region
b. Fights/The dead

3 Corpus processing

The corpus used to investigate the nature of bridging relations in real text is a 65 000 words subcorpus extracted from the french PAROLE corpus (Lecomte, 1997)[2]. This corpus consists of articles taken from the newspaper "Le Monde" and covers a wide range of topics (sports, culture, politics, economics and leisure). It is annotated at the morpho-syntactic level in accordance with the annotation scheme Multitag/Multext of the GRACE project (Beaumont et al., 1998; Lecomte, 1997). In particular, each determiner is marked as either definite, indefinite, contracted (i.e., contracting a preposition and a determiner), partitive, demonstrative, possessive, relative, exclamative or interrogative.

Starting from this corpus, we used various tools and scripts to automate the identification and annotation of definite descriptions.

3.1 Identifying definite descriptions

In a first step, we used the Gsearch system to identify and mark the definite descriptions occurring in the PAROLE corpus.

The Gsearch system (Corley et al., 2001) is a tool designed to facilitate the investigation of syntactic phenomena in unparsed corpora. It allows users to search for given linguistic structures by processing a query on the basis of a user defined grammar where the terminals of this grammar are regular expressions over elements in the corpus (e.g., words, lemmas, part of speech tags).

Customising Gsearch to the PAROLE corpus involved[3]:

- writing a filter which translates the PAROLE corpus format into the UIF (Uniform Input Format) expected by Gsearch,
- specifying a grammar for definite NPs in the Gsearch grammar format,
- writing a filter to eliminate spurious analyses produced by Gsearch (if two NP are embedded, Gsearch gives the sentence containing these two NPs twice in its output, if three are embedded, it gives it three times, etc..).
- writing a filter to adapt the output format of Gsearch to the input format of MMAX, i.e. XML text, tagged word by word and with all the definite description phrases tagged as markables to be highlighted in the MMAX window.

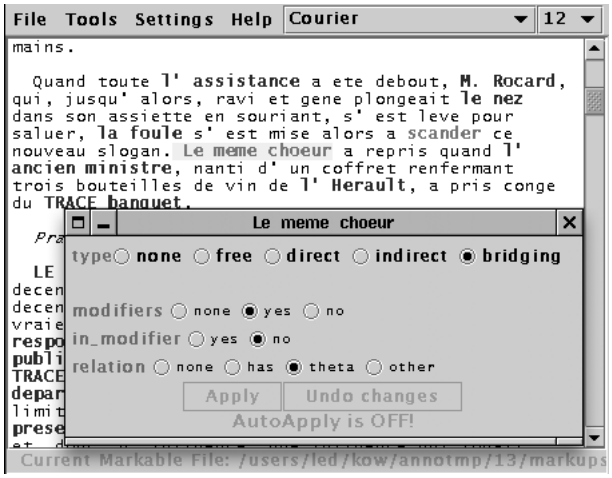

FIG. 2 – *The MMAX annotation tool*

3.2 Annotating definite descriptions

To annotate definite descriptions and their relation to their antecedent, we used the MMAX tool (Mueller and Strube, 2001). Designed to support the annotation of anaphoric and bridging relations in written text, MMAX takes as input XML encoded text corpora whereby the structure of each of the XML element types is described by a DTD (Document Type Definition). The XML elements representing markables (i.e., antecedent, coreferential and bridging NPs) have a closed set of fixed system attributes which can be complemented with user defined attributes as required by the annotation scheme.

The system is equiped with a graphical interface which allows the annotator to select the values of user or predefined system attributes by clicking within an adaptive attribute window. By means of this simple mouse-click system, the annotator can thus insert XML tags which relate an anaphor to

2. The PAROLE corpus was created by the CNRS research unit ATILF (Analyse et Traitement Informatique de la Langue Française) and was made available to us in the context of a collaboration between ATILF and the LORIA research unit.

3. This part of the work was carried out by Eric Kow.

its antecedent and indicate the type of anaphoric relation holding between these.

3.3 Post-processing of annotated corpus

The output of MMAX is an XML file. To facilitate the analysis of the annotated corpus, we wrote XSL stylesheets and shell scripts which transform this XML file into several HTML files each of which contains all the information collected about a given phenomenon (e.g, bridging definite descriptions). Specifically, each of this HTML file lists, for each anaphor/antecedent pair in the considered category, its linguistic context (i.e., the sentence containing the antecedent, the sentence containig the anaphor, and all the text in between).

By using Microsoft Excel, we can then draw on the HTML files to count, sort and extend the data. As a result, all collected information is classified within tables in an easy-to-use-and-to-exchange format.

4 Corpus annotation

The annotation phase proceeded in two steps. First, all definite descriptions were annotated as either first mention, coreferential or bridging. Second, bridging definite descriptions were annotated for the categories described in section 2.3.

In this first phase of the project, we did not carry out multiple annotation and inter annotator agreement statistics but instead worked in tandem trying to agree on the correct category.

4.1 Annotating definite descriptions

To decide on the anaphoric status of the definite description (first mention, coreferential or bridging), the annotator follows the decision tree sketched in Figure 4.1. She should first decide on whether a nominal or verbal antecedent can be found in the text which contributes to the interpretation of the definite description to be annotated. If not, the definite description is classified as “first mention”, else the annotator must decide whether anaphor and antecedent are coreferential (i.e., designates the same entity or set of entities). If yes, the definite description is annotated as “coreferential”, else as bridging. The results of this first annotation phase are given in table 4.1. As can be seen, the proportion of “first mention” is very high (almost

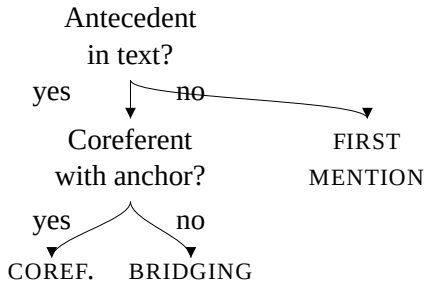

FIG. 3 – *Decision tree for first annotation phase*

80%). In comparison (Poesio and Vieira, 1998) records a rate of 48%. There are several factors which might be responsible for this difference. First, many country names, institution acronyms and idioms in French involve a definite article (e.g., la France, la Côte d’Ivoire for country names; le CNRS, la CNCL for acronyms and “avoir la main” for idioms). Second, we classified repeated use of similar definite descriptions separated from each other by a long distance as “first mention”. Third, we found a high number of generic uses. Fourth, the corpus contains a very high percentage (19.63%) of containing inferrables (i.e., definite descriptions such as “the heat of the sun” which are familiar through their explicit relation to a known entity). Note also that in her analysis of definite descriptions in Swedish, (Fraurud, 1990) found that 60.9% of the definite descriptions are first mention. Again by comparison with (Poesio and Vieira, 1998), the percentage of bridging cases is relatively low (4.7% here against 11% in (Poesio and Vieira, 1998)). This might be explained by the fact that contrary to (Poesio and Vieira, 1998), we require a strictly nominal or verbal antecedent for bridging descriptions thus excluding event or discourse deictic anaphora.

Relation	Number	Percentage
First mention	6892	78.40%
Coreferential	1481	16.85%
Bridging	416	4.73%

FIG. 4 – *Types of anaphoric relation*

4.2 Annotating bridging definite descriptions

Class	Nb of occurences	Proportion
Set membership	21	5.8%
Thematic	19	5.3%
Definitional	283	80.8%
Indiv./Attribute	32	9.0%
Associate/Indiv.	64	17.8%
Meronymic relns	187	52.0%
Whole/Part	89	24.7%
Whole/Piece	0	
Indiv/Stuff	0	
Collection/Mb	22	6.1%
Place/Area	26	7.2%
Event/Subevent	16	4.4%
Indiv./Function	34	9.4%
Co-participants	8	2.2%
Non lexical	28	7.8%
Circumstantial	17	4.7%
WKL	11	3.0%

FIG. 5 – *Bridging relations*

In a second annotation pass, we then classified the 359 bridging definite descriptions found in the PAROLE corpus according to the typology presented in section 2.3. The results are given in table 5 (the total percent adds up to slightly over 100% due to arbitrary approximations when rounding up decimals).

5 Results and Discussion

These first results on the nature of the bridging relation suggest the following preliminary conclusions.

First, the importance of the **meronymic** relation is confirmed[4] since 52% of the bridging definite descriptions involve this relation. Since, moreover, the meronymy relation is encoded in WordNet (henceforth, WN), this suggests that many cases of bridging definite descriptions could be processed using WordNet (Fellbaum, 1998). We thus did a first manual search through WordNet, checking for each bridging definite description involving a meronymic relation encoded in WN (i.e., whole/part, collection/member and Indiv/Stuff) whether it was related by a direct or indirect (i.e., inherited through a hyponym) meronymic link to its anchor. Unfortunately, we found that only 38 of our 187 meronymic cases were present in WordNet. However, a closer look at the date shows that the subtypes (town parts, country parts, entreprise parts etc.) of meronymic relations involved in corpora are actually restricted to a relatively small number which again suggest that it should not be very difficult to extend Wordnet with the meronymic information necessary to process most of the bridging definite descriptions involving this relation; or alternatively, to develop the appropriate meronymic knowledge given a specific domain and sublanguage.

Second, and again this is important for processing purposes, the number of cases involving **non-lexical knowledge** is relatively small with 4.7% of the definite descriptions involving a circumstantial relation (i.e., non knowledge based spatial or temporal inclusion e.g., laguna/the inhabitants) and 3% involving world knowledge (no lexical relation can be found between anchor and target e.g., war/survivors, fight/dead). In such cases, the relation between target and anchor can be found either (in the first case) through discourse structure (the structure of discourse determines in some way the relation between predicates, arguments and modifiers) or (in the second type of cases) through some complex reasoning (a fight can result in a person being hurt; one form of being hurt is to be dead etc.).

Third, an important class of bridging that does not appear in the literature but that turned out to be quantitatively non negligeable is the class of **Associate/Indiv** (17.8%). This class covers cases where the lexicographic definition of the target implies the existence of a target related entity whose sort subsumes the sort of the anchor. The bridging relation in such cases is the relation given by the lexicographic definition (cf. examples 3). For computational processing, the Associate/Indiv class is problematic because it presupposes the availability of lexicographic definitions usable computationally.

Finally, the **thematic** class which represents roughly 6% of the found bridging definite descriptions, could be processed using a tool such as Fra-

4. In the literature, this relation is often taken as the canonical example of bridging relation.

meNet (Baker et al., 1998) in which words are associated with a frame (or script) specifying the frame elements (aka thematic roles) likely to participate in the scenario evoked by that frame. A preliminary manual search shows that this is indeed the case –for 14 of the 19 thematic cases, we found a frame containing target and anchor as frame element. How exactly to automatically query and use FrameNet to reconstruct the missing bridging relation remains an open question however.

In summary, it seems that for the data found in the PAROLE corpus, roughly 65% of bridging definite descriptions could be processed using either framenet, wordnet or some limited form of lexical reasoning. The remaining 35% requires either lexicographic definitions (17.8%), essential attribute information (9%), discourse structure information (4.7%) or deep knowledge based reasoning (3%).

6 Conclusion

We have proposed a typology for bridging relations and classified the bridging definite descriptions occurring in the french PAROLE corpus according to this typology. In so doing we have achieved two subgoals which we see as initial steps in a more complete analysis of bridging definite descriptions.

On the one hand, we have developed the tools and infrastructure necessary for carrying a large scale analysis of definite descriptions in corpora. As explained in sections 4 and 3, this permits a rapid and user friendly analysis of morpho-syntactically annotated corpora. As further work has shown, it is moreover rapidly portable to another domain or corpora. Thus we use the same infrastructure to investigate not only bridging definite descriptions in the french PAROLE corpus, but also demonstrative and possessive NPs. In that study, the aim is to investigate the conditions under which each of these NP forms can be used. Furthermore, the tools and the ontology developed to investigate bridging definite descriptions in french are now being reused and tested on the german NEGRA corpus (Skut et al., 1997).

On the other hand, we have defined and tested a typology of the bridging relations involved in definite descriptions which contrary to the proposals in the existing literature (Clark, 1977; Strand, 1997; Kleiber, 1997) is both exhaustive and operational (the criteria used are precised enough that two annotators can rapidly agree on a given case). Further work will concentrate on refining and validating this ontology by applying it to German and English.

References

C. F. Baker, C. J. Fillmore, and J. B. Lowe. 1998. The berkeley framenet project. In *Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics*.

C. Beaumont, J. Lecomte, and N. Hatout. 1998. Etiquetage morpho-syntaxique du corpus "le monde" pour les besoins du projet parole. Technical report, INALF, Nancy.

H. H. Clark. 1977. Bridging. In P. N. Johnson-Laird and P. C. Wason, editors, *Thinking: Readings in Cognitive Science*. Cambridge University Press, Cambridge.

S. Corley, M. Corley, F. Keller, M. W. Crocker, and S. Trewin. 2001. Finding syntactic structure in unparsed corpora: The Gsearch corpus query system. *Computers and the Humanities*, 35(2):81–94.

C. Fellbaum, editor. 1998. *WordNet. An electronic lexical database*. MIT Press, Cambridge/Mass.

K. Fraurud. 1990. Definiteness and the processing of nps in natural dicourse. *Journal of Semantics*.

G. Kleiber. 1997. *Des anaphores associatives méronymiques aux anaphores associatives locatives*, volume 19. Verbum.

J. Lecomte. 1997. Codage multext-grace pour l'action grace multitag. Technical report, INALF, Nancy.

G.A. Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11).

C. Mueller and M. Strube. 2001. Annotating anaphoric and bridging relations with mmax. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, pages 90–95, Aalborg, Denmark.

M. Poesio and R. Vieira. 1998. A corpus-based investigation of definite description use. *Computational Linguistics*, 24(2):183–216.

W. Skut, B. Krenn, T. Brants, and H. Uszkoreit. 1997. An annotation scheme for free word order languages. In *Proceedings of the Fifth Conference on Applied Natural Language Processing ANLP-97*, Washington, DC.

K. Strand. 1997. A taxonomy of linking relations.

M. Winston, R. Chaffin, and D. Herrmann. 1987. A taxonomy of part-whole relations. *Cognitive Science*.