Corpus-based Development and Evaluation of a System for Processing Definite Descriptions

Renata Vieira
Universidade do Vale do Rio dos Sinos
Av. Unisinos 950 - Cx. Postal 275
93022-000 São Leopoldo RS Brazil.
renata@exatas.unisinos.br

Massimo Poesio
University of Edinburgh
HCRC and Informatics
Edinburgh, Scotland
Massimo.Poesio@ed.ac.uk

Abstract

We present an implemented system for processing definite descriptions. The system is based on the results of a corpus analysis previously reported, which showed how common discourse-new descriptions are in newspaper corpora, and identified several problems to be dealt with when developing computational methods for interpreting bridging descriptions. The annotated corpus produced in this earlier work was used to extensively evaluate the proposed techniques for matching definite descriptions with their antecedents, discourse segmentation, recognizing discourse-new descriptions, and suggesting anchors for bridging descriptions.

1 Motivation

In previous work (Poesio and Vieira, 1998) we reported the results of corpus annotation experiments in which the subjects were asked to classify the uses of definite descriptions in Wall Stree Journal articles according to a scheme derived from work by Hawkins (1978) and Prince (1981) and including three classes: DIRECT ANAPHORA, DISCOURSE-NEW, and BRIDGING DESCRIPTION (Clark, 1977). This study showed that about half of the time, definite descriptions are used to introduce a new entity in the discourse, rather than to refer to an object already mentioned. We also observed that our subjects didn't always agree on the classification of a given definite; the problem was especially acute for bridging descriptions.

In this paper, we present an implemented system for processing definite descriptions based on the results of that earlier study. In our system, techniques for recognising discourse-new descriptions play a role as important as techniques for identifying the antecedent of anaphoric ones. The system also incorporates robust techniques for processing bridging descriptions.

A fundamental characteristic of our system is that it was developed so that its performance could be evaluated using the annotated corpus. In the paper, we discuss how we arrived at the optimal version of the system by measuring the performance of each method in this way. Because of the problems observed in our previous study concerning agreement between annotators, we evaluated the system both by measuring precision/recall against a 'gold standard' and by measuring the agreement between the annotation it produces and the annotators.

2 General Overview

At the moment, the only systems engaged in semantic interpretation whose performance can be evaluated on fairly unrestricted text such as the Wall Street Journal articles are based on a shallow-processing approach, i.e., that do not rely on extensive amounts of hand-coded commonsense knowledge (Carter, 1987; Appelt, 1995; Humphreys et al., 1998).[1] Our system is of this type: it only relies on structural information, on the information provided by pre-existing lexical sources such as WordNet (Fellbaum, 1998), on minimal amounts of general hand-coded information, and on information that can be acquired automatically from a corpus. Although we believe that quantitative evaluations of the performance of a system on a large number of examples are the only true assessment of its performance, and therefore a shallow processing approach is virtually unavoidable for implemented systems until better sources of commonsense knowledge become available, we do know that this approach limits the performance of a system on those instances of definite descriptions which do require commonsense knowledge for their resolution. (We grouped these in what we call the 'bridging' class.) We

---

[1]Most systems participating in the Message Understanding Conference (MUC) evaluations are customized to specific domains by adding hand-coded commonsense knowledge.

nevertheless developed heuristic techniques for processing these types of definites as well, which may provide a baseline against which the gains in performance due to the use of commonsense knowledge can be assessed more clearly.

Our system attempts to classify each definite description as either DIRECT ANAPHORA, DISCOURSE-NEW, and BRIDGING DESCRIPTION. The first class includes definite descriptions whose head is identical to that of their antecedent, as in *a house ... the house*. The second includes definite descriptions that refer to objects not already mentioned in the text and not related to any such object. (Some of these definite descriptions refer to objects whose existence is widely known, such as discourse-initial references to *the pope*; other instances of discourse-new descriptions refer to objects that can be assumed to be unique, even if unfamiliar, such as *the first woman to climb all Scottish Munros*.) Finally, we classify as bridging descriptions all definite descriptions whose resolution depends on knowledge of *relations* between objects, such as definite descriptions that refer to an object related to an entity already introduced in the discourse by a relation other than identity (Prince's 'inferrables'), as in *the flat ... the living room*; and definite descriptions that refer an object already introduced, but using a different predicate, as in *the car ... the vehicle*. In addition to this classification, the system tries to identify the antecedents of anaphoric descriptions and the anchors (Fraurud, 1990) of bridging ones. Accordingly, we developed three types of heuristics:

- for resolving directly anaphoric descriptions. These include heuristics for dealing with segmentation and to handle modification.

- for identifying discourse-new descriptions. Some of these heuristics attempt to recognize semantically functional definite descriptions (Hawkins, 1978; Loebner, 1987), whereas others try to recognize definite descriptions that are anchored via their modification (Clark and Marshall, 1981; Prince, 1981).

- for identifying the anchor of a bridging description and the semantic relation between the bridging description and its anchor. WordNet is accessed, and heuristics for named entity recognition were also developed.

The final configuration of the system was arrived at on the basis of an extensive evaluation of the heuristics using the corpus annotated in our previous work (Poesio and Vieira, 1998). The evaluation was used both to determine which version of each heuristic worked better, and to identify the best order in which to try them.

The corpus we used consists of 34 texts from the Penn Treebank I included in the ACL/DCI CD-rom. 20 of these texts were treated as 'training corpus'; this corpus contains 1040 definite descriptions, of which 312 are anaphoric, 492 discourse-new, and 204 bridging. 14 more texts were used as 'test corpus'; these include 464 definite descriptions, of which 154 have been classified as anaphoric, 218 as discourse-new, and 81 as bridging.

3 The Heuristics And Their Performance

3.1 Resolving Anaphoric Definites

We discuss heuristics for two subproblems of the task of resolving anaphoric definites: limiting the accessibility of discourse entities (segmentation), and taking into account the information given by pre- and post-modifiers. See (Vieira, 1998) for a discussion of the other heuristics used by the system.

Segmentation In general, discourse entities have life-spans limited to pragmatically determined SEGMENTS that may be nested (see, e.g., (Reichman, 1985; Grosz and Sidner, 1986; Fox, 1987)). E.g., in our corpus we found that about 10% of direct anaphoric definite descriptions have more than one possible antecedent if segmentation is not taken into account (Vieira and Poesio, 1999). Recognizing the hierarchical structure of segments in a text is, however, still pretty much an open problem, as it involves reasoning about intentions;[2] better results have been achieved on the simpler task of 'chunking' the text into approximate segments, generally by means of lexical density measures (Hearst, 1997) . In fact, the methods to limit the lifespan of discourse entity we considered for our system were even simpler. One type of heuristics we looked at are window-based techniques, i.e., considering as potential antecedents only the discourse entities within fixed-size windows of previous sentences, allowing however for some discourse entities to take a longer life span: we call this method LOOSE SEGMENTATION. More specifically, a discourse entity is considered as potential antecedent for a definite

---

[2]See, however, (Marcu, 1999).

description when the antecedent's head is identical to the description's head, and

- the potential antecedent's distance from the description is within the established window, or else

- the potential antecedent is itself a subsequent mention, or else

- the definite description and the antecedent are identical NPs (including the article).

We also considered an even simpler RECENCY heuristic: this involves keeping a table indexed by the heads of potential antecedents, such that the entry for noun N contains the index of the last occurrence of an antecedent with head N. Finally, we considered combinations of segmentation and recency.

The best results were obtained with a combination of the recency and segmentation heuristics: just one potential antecedent for each different head noun is available for resolution, the last occurrence of that head noun. The resolution still respects the segmentation heuristic (loose version). The recall (R), precision (P) and F-measure (F) results for the two heuristics are presented in Table 1.[3]

Combined heuristics	R	P	F
4 sentences + recency	75.96%	87.77%	81.44%
8 sentences + recency	77.88%	84.96%	81.27%

Table 1: Combining loose segmentation and recency heuristics

The version with higher F value in Table 1 (4-sentence window plus recency) was chosen and used in the tests discussed in the rest of this section. *Noun Modifiers* In general, when matching a definite description with a potential antecedent the information provided by the prenominal and the postnominal part of the noun phrase also has to be taken into account: so, for example, *a blue car* cannot serve as the antecedent for *the red car*, or *the house on the left* for *the house on the right*. Taking proper care of the semantic contribution of these premodifiers would, in general, require commonsense rea-

---

[3]The standard definitions of precision and recall from information retrieval were used: R = number of objects of type A correctly identified by the system / total number of objects of type A, P = number of correct identifications of objects of type A / total number of objects of type A identified by the system, F = RP / R+P.

soning; for the moment, we only developed heuristic solutions to the problem, including:

- allowing an antecedent to match with a definite description if the premodifiers of the description are a subset of the premodifiers of the antecedent. This heuristic deals with definites which contain less information than the antecedent, such as *an old Victorian house... the house*, and prevents matches such as *the business community... the younger, more activist black political community*.

- allowing a non-premodified antecedent to match with any same head definite. This second heuristic deals with definites that contain additional information, such as *a check... the lost check*.

The results of our premodifier matching algorithm are presented in Table 2. In that Table we also show the results obtained with a modified matching algorithm including a third rule, that allows a premodified antecedent to match with a definite whose set of pre-modifiers is a superset of the set of modifiers of the antecedent (an elaboration of rule 2). We tested each of these three heuristics alone and their combinations. (The fourth line simply repeats the results shown in Table 1.)

Antecedents selection	R	P	F
1. Ant-set/Desc-subset	69.87%	91.21%	79.12%
2. Ant-empty	55.12%	88.20%	67.85%
3. Ant-subset/Desc-set	64.74%	88.59%	74.81%
1 and 2 (basic v.)	75.96%	87.77%	81.44%
1 and 3	75.96%	87.13%	81.16%
None	78.52%	81.93%	80.19%

Table 2: Evaluation of the heuristics for premodification (version 1)

The best precision is achieved by the matching algorithm that does not allow for new information in the anaphoric expression, but the best results overall are again obtained by combining rule 1 and rule 2, although either 2 or 3 works equally well when combined with 1.
Overall results for anaphoric definite descriptions To summarize, the version of the system that achieves the best results as far as anaphoric definite descriptions are concerned includes :

1. combined segmentation and recency,

2. 4-sentence window,

3. considering indefinites, definites and possessives as potential antecedents (Vieira, 1998),

4. the premodification of the description must be contained in the premodification of the antecedent when the antecedent has no premodifiers.

3.2 Heuristics for Recognizing Discourse-New Descriptions

As mentioned above, a central characteristic of our system is that it also includes heuristics for recognising discourse-new descriptions (i.e., definite descriptions that introduce new discourse entities) on the basis of syntactic and lexical features of the noun phrase. Our heuristics are based on the discussion by Hawkins (1978), who identified a number of correlations between certain types of syntactic structure and discourse-new descriptions, particularly those that he called 'unfamiliar' definites (i.e., those whose existence cannot be expected to be known on the basis of generally shared knowledge), including:

- the presence of 'special predicates':[4]

  - the occurrence of pre-modifiers such as *first* or *best* when accompanied with full relatives, e.g., *the first person to sail to America* (Hawkins calls these 'unexplanatory modifiers'; Loebner (1987) showed how these predicates may license the use of definite descriptions in an account of definite descriptions based on functionality);
  - a head noun taking a complement such as *the fact that there is life on Earth* (Hawkins calls this subclass 'NP complements');

- the presence of restrictive modification, as in *the inequities of the current land-ownership system*.

Our system attempts to recognize these syntactic patterns; in addition, it considers as unfamiliar some definites occurring in

---

[4]This list was developed by hand; more recently, Bean and Riloff (1999) proposed methods for automatically extracting from a corpus such special predicates, i.e., heads that correlate well with discourse novelty.

- appositive constructions (e.g., *Glenn Cox, the president of Phillips Petroleum Co.*);

- copular constructions (e.g.,*the man most likely to gain custody of all this is a career politician named David Dinkins*).

In our corpus study (Poesio and Vieira, 1998) we found that our subjects did better at identifying discourse-new descriptions all together (K=.68) than they did at distinguish 'unfamiliar' from 'larger situation' (Hawkins, 1978) cases (K = .63). This finding was confirmed by our implementation: although each of the heuristics is designed, in principle, to identify only one of the uses (larger situation or unfamiliar), they work better when used all together to the class of discourse new descriptions.

The overall recall and precision results for the heuristics for identifying discourse new descriptions are shown in Table 3. In this Table we do not distinguish between the two types of discourse-new descriptions, 'unfamiliar' and 'larger-situation'. The column headed by (#) represents the number of cases of descriptions classified as discourse new in the standard annotation; + indicates the total number of discourse-new descriptions correctly identified; - the number of errors. These results are for the version of the system (version 1) that uses the best version of the heuristics for dealing with anaphoric descriptions discussed above, and that doesn't attempt to resolve bridging descriptions .

Discourse new	#	+	-	R	P	F
Training data	492	368	60	75%	86%	80%
Test data	218	151	58	69%	72%	70%

Table 3: Evaluation of the heuristics for identifying discourse new descriptions

3.3 Bridging Descriptions

Bridging descriptions are the class of definite descriptions which a shallow processing system is least equipped to handle, and therefore the most crucial indicator of where commonsense knowledge is actually needed. We knew from the start that in general, a system can only resolve certain types of bridging descriptions when supplied with an adequate knowledge base; in fact, the typical way of implementing a system for resolving bridging references has been to restrict the domain and feed the system with hand-coded world knowledge (see, e.g., (Sidner, 1979) and especially (Carter, 1987)).

Furthermore, the relation between bridging descriptions and their anchors may be arbitrarily complex (Clark, 1977; Sidner, 1979; Prince, 1981; Strand, 1996) and our own results indicate that the same description may relate to different anchors in a text, which makes it difficult to decide what the intended anchor and the intended link are (Poesio and Vieira, 1998). Nevertheless, we feel that trying to process these definite descriptions is the only way to discover which types of commonsense knowledge are actually needed. .

We began by developing a classification of bridging descriptions according to the kind of information needed to resolve them, rather than on the basis of the possible relations between descriptions and their anchors as usually done in the literature (Vieira, 1998). This allowed us to get an idea of what types of bridging descriptions our system might be able to resolve. We classified definite descriptions as follows:

- cases based on well-defined lexical relations, such as synonymy, hypernymy and meronymy, that can be found in a lexical database such as WordNet (Fellbaum, 1998)–as in *the flat . . . the living room*;

- bridging descriptions in which the antecedent is a proper name and the description a common noun, whose resolution requires some way of recognizing the type of object denoted by the proper name (as in *Bach . . . the composer*);

- cases in which the anchor is not the head noun but a noun modifying an antecedent, as in *the company has been selling discount packages . . . the discounts*

- cases in which the antecedent (anchor) is not introduced by an NP but by a VP, as in *Kadane oil is currently drilling two oil wells. The activity* . . .

- descriptions whose the antecedent is not explicitly mentioned in the text, but is implicitly available because it is a discourse topic–e.g., *the industry* in a text referring to oil companies;

- cases in which the relation with the anchor is based on more general commonsense knowledge, e.g., about cause-consequence relations.

We developed heuristics for handling the first three of these classes: lexical bridges, bridges based on names, and bridges to entities introduced by non-head nouns in a compound nominal. We refer the reader to (Vieira, 1998) for discussion of the heuristics for this last class.

Our system attempts to resolve lexical bridges by consulting WordNet to determine if there is a semantic relation between the head noun of the description and the head noun of one of the NPs in the previous five sentences. The results of this search for our training corpus, in which 204 descriptions are classified as bridging, are shown in Table 4. It is interesting to note that the semantic relations found in this automatic search were not always those observed in our manual analysis.

Bridging Class	Relations Found	Right Anchors	% Right
Synonimy	11	4	36%
Hyponimy	59	18	30%
Meronimy	6	2	33%
Sister	30	6	20%
Total	106	30	28%

Table 4: Evaluation of the search for anchors using WordNet

We developed a simple heuristic method for assigning types to named entities. Our method identified entity types for 66% (535/814) of all names in the corpus (organizations, persons and locations). The precision was 95%. We could have had a better recall if we had adopted more comprehensive lists of cue words, or consulted dictionaries of names as done for the systems participating in MUC-6. There, recall in the named entity task varies from 82% to 96%, and precision from 89% to 97%.[5]

4   Overall Evaluation of the System

The order of application of heuristics is as important as the heuristics themselves. The final order of application was also arrived at on the basis of an extensive evaluation (Vieira, 1998), and is based on the following strategy:[6]

---

[5]A more recent version of the system using the named entity recognition software developed by HCRC for the MUC-7 competition (Mikheev et al., 1999) is discussed in (Ishikawa, 1998).

[6]We also attempted to learn the best order of application of the heuristics automatically by means of decision tree learning algorithms (Quinlan, 1993), without however observing a significant difference in performance. See (Vieira, 1998) for details.

1. eliminate first some non-anaphoric cases using 'safe' heuristics (first two tests);

2. if that fails, try to find a same head antecedent;

3. if that doesn't work either, look for an indication that the description is discourse new

4. only then try to interpret the definite description as a bridge.

Two versions of the system were evaluated. Version 1 only attempts to resolve anaphoric definite descriptions and to recognize discourse-new ones, using the techniques discussed in 3.1 and 3.2. Version 2 of the system also attempts to recognize bridging descriptions and to identify their anchors. (We evaluated this version separately because we expected it to get worse results, given the problems mentioned in 3.3.) Both versions of the system were evaluated using the corpus obtained in (Poesio and Vieira, 1998), which, as discussed earlier, was divided in a training set of 1,040 definite descriptions and a test set of 464 definite descriptions not looked at while developing the heuristics. We evaluated the systems both by means of precision and recall figures, and by measuring the agreement of the system with our annotators by means of the K measure (Carletta, 1996).

4.1 Recall and Precision Figures

Version 1 of the system assigned a classification to 698 out of 1040 definite descriptions in the training corpus; 342 were not classified. Of these 698, 611 were classified correctly, 87 incorrectly. In the test corpus, 324 definites were classified, 140 weren't; 254 definites were classified correctly. Per-class and overall precision and recall figures were as follows:

Task	#	+	-	R	P	F
Anaphora classification						
Training data	312	243	27	78%	90%	83%
Test data	154	103	12	67%	90%	77%
Anaphora resolution						
Training data	312	237	33	76%	88%	81%
Test data	154	96	19	62%	83%	71%
Discourse new						
Training data	492	368	60	75%	86%	80%
Test data	218	151	58	69%	72%	70%
Overall (Classification)						
Training Data (total)	1040	611	87	59%	88%	70%
Training Data (in domain)	804	611	87	76%	87%	81%
Test Data (total)	464	254	70	55%	78%	64%
Test Data (in domain)	372	254	70	68%	78%	73%

Recall was only 59% on the training data when considering all definite descriptions, but 87% when considering only the definite descriptions that our system was actually equipped to handle; if the definite descriptions that the system couldn't classify were treated as discourse-new, we would get R = P = 70%.

Version 2 of the system couldn't be evaluated automatically, as the suggestions concerning bridging descriptions have to be evaluated by hand. The system correctly classified 264 out of 464 definite descriptions in the test corpus, for a recall of 57%, but also incorrectly resolved 113 definite descriptions, reducing the overall precision to 70%.

4.2 Agreement with Annotators

We also evaluated the system by comparing the class it assigned to each definite with all the classes assigned to that definite by the annotators, i.e., without fixing a 'golden standard'. Version 1 of the system finds a classification for 318 out of 464 definite descriptions in the test data. If all the definite descriptions that the system cannot classify are treated as discourse-new, the agreement between system and the three subjects that annotated this corpus on the two classes first mention (= discourse-old) and subsequent mention (= discourse-new or bridges) is K=.7; this should be compared with an agreement of K=.77 between the three annotators themselves. If instead of counting these definite descriptions as discourse-new we simply do not include them in our measure of agreement, then the agreement between the system and the annotators is K=.78, as opposed to K=.81 between the annotators.

Version 2 finds a classification for 355 out of 464 definite descriptions; however, its agreement figures are worse. If we count the cases that the system can't classify as discourse-new, the agreement between the system and the three annotators for three classes is K=.57; if we count them as bridges, K=.63; if we just discard those cases, K=.63 again. (By comparison, the agreement among annotators on the three classes was K=.68 overall, K=.70 on just the cases that the system was able to classify.) The cases that the system can't handle are mainly discourse-new descriptions.

5 Conclusions

We have presented a domain independent system for definite description interpretation whose development was based on an empirical study of definite description use that included multi-annotator experiments. The main novelty of this system is that it does not only attempt to find an antecedent for a

definite description; it also uses methods for recognizing discourse-new descriptions, which our previous studies revealed to be the largest class of definite descriptions in our corpus. Because our system only relies on 'shallow' information, it encounters problems when commonsense reasoning is actually needed; on the other hand, it can be tested on any domain without extensive hand-coding.

We believe that ultimately, the real value of such a system will be that it provides a baseline against which to measure the impact of commonsense knowledge, once heuristics such as those we developed for premodifier matching can be replaced by robust methods for commonsense inference. We already started preliminary work on replacing a hand-crafted lexical knowledge base like WordNet with automatically acquired lexical bases (Poesio et al., 1998; Ishikawa, 1998).

Acknowledgments

References

D. et al.. Appelt. 1995. SRI International FASTUS system MUC-6 test results and analysis. In *Proc. of the Sixth Message Understanding Conference*, pages 237–248, Columbia, Maryland, November.

D. L. Bean and E. Riloff. 1999. Corpus-based identification of non-anaphoric noun phrases. In *Proc. of the 37th ACL*, pages 373–380, University of Maryland.

J. Carletta. 1996. Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, 22(2):249–254.

D. M. Carter. 1987. *Interpreting Anaphors in Natural Language Texts*. Ellis Horwood, Chichester, UK.

H. H. Clark and C. R. Marshall. 1981. Definite reference and mutual knowledge. In A. Joshi, B. Webber, and I. Sag, editors, *Elements of Discourse Understanding*. Cambridge University Press, New York.

H. H. Clark. 1977. Inferences in comprehension. In D. Laberge and S. J. Samuels, editors, *Basic Process in Reading: Perception and Comprehension*. Lawrence Erlbaum.

C. Fellbaum, editor. 1998. *WordNet: An electronic lexical database*. The MIT Press.

B. A. Fox. 1987. *Discourse Structure and Anaphora*. Cambridge University Press, Cambridge, UK.

K. Fraurud. 1990. Definiteness and the processing of NPs in natural discourse. *Journal of Semantics*, 7:395–433.

B. J. Grosz and C. L. Sidner. 1986. Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

J. A. Hawkins. 1978. *Definiteness and Indefiniteness*. Croom Helm, London.

M. A. Hearst. 1997. TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1):33–64.

K. Humphreys, R. Gaizauskas, S. Azzam, C. Huyck, B. Mitchell, and H. Cunningham Y. Wilks. 1998. University of Sheffield: description of the LaSIE-II System as used for MUC-7. In *Proc. of the Seventh Message Understanding Conference (MUC-7)*.

T. Ishikawa. 1998. Acquisition of associative information and resolution of bridging descriptions. Master's thesis, University of Edinburgh, Department of Linguistics, Edinburgh, Scotland.

S. Loebner. 1987. Natural language and generalised quantifier theory. In P. Gärdenfors, editor, *Generalized Quantifiers*, pages 93–108. D. Reidel, Dordrecht.

D. Marcu. 1999. A decision-based approach to rhetorical parsing. In *Proc. of the 37th ACL*, pages 365–372, University of Maryland, June. ACL.

A. Mikheev, M. Moens, and C. Grover. 1999. Named Entity recognition without gazetteers. In *Proc. of EACL*, Bergen, Norway. EACL.

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

M. Poesio, S. Schulte im Walde, and C. Brew. 1998. Lexical clustering and definite description interpretation. In *Proc. of the AAAI Spring Symposium on Learning for Discourse*, pages 82–89, Stanford, CA, March. AAAI.

E. F. Prince. 1981. Toward a taxonomy of given-new information. In P. Cole, editor, *Radical Pragmatics*, pages 223–256. Academic Press, New York.

J. R. Quinlan. 1993. *C4.5: programs for machine learning*. Morgan Kaufmann, San Mateo, CA.

R. Reichman. 1985. *Getting Computers to Talk Like You and Me*. The MIT Press, Cambridge, MA.

C. L. Sidner. 1979. *Towards a computational theory of definite anaphora comprehension in English discourse*. Ph.D. thesis, MIT.

K. Strand. 1996. A taxonomy of linking relations. Manuscript. A preliminary version presented at the Workshop on Indirect Anaphora, Lancaster University, 1996.

R. Vieira and M. Poesio. 1999. Processing definite descriptions in corpora. In S. Botley and T. McEnery, editors, *Corpus-based and Computational Approaches to Discourse Anaphora*. John Benjamins.

R. Vieira. 1998. *Definite Description Resolution in Unrestricted Texts*. Ph.D. thesis, University of Edinburgh, Centre for Cognitive Science, February.