Using the Web for Nominal Anaphora Resolution

Katja Markert, Malvina Nissim
School of Informatics
University of Edinburgh
markert@inf.ed.ac.uk
mnissim@inf.ed.ac.uk

Natalia N. Modjeska
School of Informatics
University of Edinburgh and
Department of Computer Science
University of Toronto
natalia@cs.toronto.edu

Abstract

We present a novel method for resolving non-pronominal anaphora. Instead of using handcrafted lexical resources, we search the Web with shallow patterns which can be predetermined for the type of anaphoric phenomenon. In experiments for *other*-anaphora and bridging, our shallow, almost knowledge-free and unsupervised method achieves state-of-the-art results.

1 Introduction

After having focussed on pronominal anaphora, researchers are now devoting attention to other nominal anaphors as well (Harabagiu and Maiorano, 1999; Vieira and Poesio, 2000; Bierner, 2001; Modjeska, 2002; Ng and Cardie, 2002). These comprise such diverse phenomena as *coreference*, *bridging* (Clark, 1975) (see also Example (1)), and *other-anaphora* (Example (2)).[1]

(1) *The apartment* she shares with a 12-year-old daughter and her sister was rattled, books and crystal hit **the floor**, [...]

(2) You either believe Seymour can do it again or you don't. Beside *the designer's age*, **other risk factors for Mr. Cray's company** include the Cray-3's [...] chip technology.

[1] In all examples the anaphor is typed in bold face and the antecedent in italics. All examples in this paper are from the Wall Street Journal (WSJ), Penn Treebank, release 2.

In Example (1), the definite NP "the floor" can be felicitously used because a related entity, "the apartment" has already been introduced, and a part-of relation between the two entities can be established. In other-anaphora, *other* provides a set-complement to an entity already evoked in the discourse model. In Example (2), the NP "other risk factors for Mr. Cray's company" refers to a set of risk factors *excluding* the designer's age, and can be paraphrased as "other risk factors (for Mr. Cray's company) than the designer's age".

There is evidence that grammatical salience plays a lesser role for resolving anaphors with full lexical heads, than for pronominal anaphora (Strube and Hahn, 1999; Modjeska, 2002). Instead, a large and diverse amount of lexical or world knowledge is necessary to understand examples like (1) and (2): for example, that floors are parts of apartments or that age can be viewed as a risk factor. Therefore, the state-of-the-art resolution systems that handle these phenomena rely heavily on handcrafted resources, such as the WordNet lexical hierarchy (Fellbaum, 1998).

Using WordNet suffers from several major drawbacks. Many expressions (e.g., risk factor), word senses and lexical relations (e.g., floor as a part of an apartment) are missing from the database. The hierarchy is structured such that the desired information might not be straightforward to retrieve. For example, in WordNet, "floor" is encoded as part of "building", but not as part of "apartment", although "apartment" is itself a meronym of "apartment building" which in turn is a hyponym of "building" (see also (Vieira and Poesio, 2000)). Moreover, manually built resources are expensive and time-consuming to build and maintain.

There have been efforts to extract missing lexical relationships from corpora in order to build new knowledge sources and enrich existing ones (Hearst, 1992; Berland and Charniak, 1999; Poesio et al., 2002). In our view there are two main problems with these approaches.

Firstly, the size of the used corpora still leads to data sparseness (Berland and Charniak, 1999) and the extraction procedure can therefore require extensive smoothing. Secondly, it is not clear *how much* and *which* knowledge to include in a fixed context-independent ontology, whether manually built or derived from a corpus. Thus, should metonymy, underspecified and point-of-view dependent hyponymy relations (Hearst, 1992) be included? Should age, for example, be classified as a hyponym of risk factor independent of context?

To solve the first problem, we propose using the Web, which with approximately 968M pages[2] is the largest corpus available to the NLP community. Using the web has proved successful in several fields of NLP, e.g., machine translation (Grefenstette, 1999) and bigram frequency estimation (Keller et al., 2002). In particular, (Keller et al., 2002) have shown that using the Web handles data sparseness better than smoothing. However, to our knowledge, the Web has not been used for anaphora resolution yet.

We do not offer a solution to the second problem, but instead claim that, for our task, we *do not need* a predetermined fixed ontology at all. In Example (2), we do not need to have and fix the knowledge that age is always a risk factor, but only that, among the possible NP antecedents "Seymour", "designer", and "(designer's) age", the latter is the most likely to be viewed as a risk factor.

In the next section, we introduce a method that uses shallow lexico-syntactic patterns and their web frequencies instead of a fixed ontology to achieve this comparison between several possible antecedents. We then present two experiments on other-anaphora and bridging, respectively. Our shallow technique, used on a noisy and unprocessed corpus like the Web, achieves results comparable with state-of-the-art methods using handcrafted knowledge bases. Finally, the results are discussed and compared to related work.

2 The Basic Idea

In the phenomena we consider, the relation between anaphor and antecedent is implicitly expressed, i.e., anaphor and antecedent do not stand in a structural or grammatical relationship. However, they are linked by a strong semantic relation that is likely to be *structurally explicitly expressed* in other texts. We exploit this insight by adopting the following procedure:

1. Dependent on the anaphoric phenomenon, we determine which lexical relationships usually hold between anaphor and antecedent. For example, in other-anaphora, a hyponymy/similarity relation between the lexical heads of anaphor and antecedent is stipulated by the context,[3] e.g. age is viewed as a risk factor.

2. We select patterns that structurally explicitly express the same lexical relationships. For example, `NP`$_1$ `and other NP`$_2$ is a pattern that usually expresses hyponymy/similarity relations between the hyponym NP_1 and its hypernym NP_2 (Hearst, 1992).

3. If the implicit lexical relationship between anaphor and antecedent is strong, then it is likely that anaphor and antecedent also frequently cooccur in the selected explicit patterns. We extract all possible antecedents for each anaphor, and instantiate the explicit pattern for all anaphor/antecedent pairs. In Example (2) the pattern `NP`$_1$ `and other NP`$_2$ can be instantiated with `Seymour and other risk factors`, `designer and other risk factors`, and `age and other risk factors`.[4] The instantiation of a pat-

[2]`http://www.searchengineshowdown.com/stats/sizeest.shtml`, data from March 2002.

[3]From now on, we will often use "anaphor/antecedent" instead of the more cumbersome "lexical heads of the anaphor/antecedent".

[4]These simplified instantiations serve as an example and are not the final instantiations we use; see Section 3.2.

tern can be searched in any corpus to determine its frequency. We now follow the rationale that the most frequent of these instantiated patterns determines the correct antecedent.

4. As the patterns can be quite elaborate, most corpora will be too small to determine the corresponding frequencies reliably. The instantiation `age and other risk factors`, for example, does not occur at all in the British National Corpus (BNC), a 100M words corpus of British English.[5] Therefore we use the largest corpus available, the Web.[6] We submit all instantiated patterns as queries to the Web making use of the GOOGLE API technology and, as a first approximation, select the instantiation yielding the highest number of hits. Here, `age and other risk factors` yields over 400 hits, whereas the other two instantiations for this example yield 0 hits each.

3 Experiment I: Other-anaphora

Here we restrict other-anaphora to referential lexical NPs with the modifiers *other* or *another* and non-structurally given antecedents, as in Example (2).[7] The distance between an other-anaphor and its antecedent can be large; (Modjeska, 2002) observed a dependency that spans over 17 sentences.

3.1 Data Collection and Preparation

We tested our method on 120 samples of other-anaphors from the *Wall Street Journal* corpus (Penn Treebank release 2, first three sections). These samples are part of the dataset reported in (Modjeska, 2002). We used the samples in which the antecedents are NPs realized within a two-sentence window, i.e., either in the sentence containing the anaphor or in the preceding one.

Antecedent Extraction. For each anaphor we extracted the set of all potential NP-antecedents in the two-sentence window. This was done in three steps. First, we extracted all base NPs, i.e., NPs that contain no further NPs within them. NPs containing a possessive modifier, e.g. "the designer's age" were split into the possessor NP, "the designer's" and the possessed phrase, "age". Second, we filtered out null elements (tagged -NONE-) and pronouns. Pronouns can be antecedents of other-anaphors but our method cannot deal with them since they are lexically empty. Third, we split coordinated NPs, e.g., "risk, technology and innovation", into their constituting parts using simple heuristics. We also split proper names followed by a common noun; therefore "Mips computers" was automatically split into an antecedent "Mips" and an antecedent "Mips computers".

We call $\mathcal{A}$ the list of possible antecedents, and *ana* the anaphor. For Example (2), this results in $\mathcal{A}$={*Seymour, the designer's, age*} and *ana=other risk factors for Mr. Cray's company*.

Antecedent Preparation and Named Entity (NE) Recognition. All modification was eliminated to avoid data sparseness. In addition only the rightmost noun of compounds was kept.

For Example (2), this results in $\mathcal{A}$={*Seymour, designer, age*} and *ana=factors*.

Using patterns containing NEs (like "Mr. Pickens" in Example (4)) also leads to data sparseness.

(4) Koito has refused to grant *Mr. Pickens* seats on its board, asserting *he* is a greenmailer trying to pressure Koito's **other shareholders** [...]

We resolved NEs in two steps. First, we processed the data using ANNIE, an IE software, which is part of the GATE2 software package.[8] We only used its classification into the ENAMEX MUC-7 categories (Chinchor, 1997): PERSON, ORGANIZATION and LOCATION. Second, we used some heuristics to automatically obtain more fine-grained distinctions for the categories LOCATION

[5] `http://info.ox.ac.uk/bnc`

[6] The Web is a constantly growing, changing and updating resource. On the one hand, its size and changing potential are an advantage as we can have access to such a large corpus without having to create it. On the other hand, one has no control over its content.

[7] In contrast, in Example (3), the (split) antecedent "jam and cocoa" is the coordinated constituent to the left of the conjunction "and". Thus the antecedent is given *structurally*.

(3) [...] it enabled her to buy *jam*, *cocoa* and **other war-rationed goodies**.

[8] `http://gate.ac.uk`

and ORGANIZATION, whenever possible. We classified LOCATIONS into COUNTRY, (US) STATE, CITY, RIVER, LAKE and OCEAN, using mainly gazetteers.[9] If an entity classified by GATE as ORGANIZATION contained an indication of the organization type, we used this as a subclassification; therefore "Bank of America" is classified as BANK. No further distinctions were developed for the category PERSON. For numeric and times entities we used simple heuristics to classify them further into DAY, MONTH, YEAR as well as DOLLAR or simply NUMBER.

Disregarding numeric and time entities, our dataset included 262 possible NE antecedents. Our method recognised 216 as proper names (82% recall). Of these, 202 (93% precision) were correctly classified.

Finally, all elements of $\mathcal{A}$ were lemmatized.

For Example (2), this results in $\mathcal{A}$={*person [=Seymour], designer, age*} and *ana=factor*.

3.2 Pattern Selection and Query Generation

We use the following pattern for other-anaphora:[10]

(O1) `(`N$_1${*sg*} `OR` N$_1${*pl*}`) and other` N$_2${*pl*}

For common noun antecedents, we instantiate the pattern by substituting N$_1$ with a possible antecedent, an element of $\mathcal{A}$, and N$_2$ with *ana*, as normally N$_1$ is a *hyponym* of N$_2$ in (O1), and the antecedent is a hyponym of the anaphor. An instantiated pattern for Example (2) is `(age OR ages) and other factors` (see I_1^c in Table 1).[11]

For NE antecedents we instantiate (O1) by substituting N$_1$ with the NE category of the antecedent, and N$_2$ with *ana*. An instantiated pattern for Example (4) is `(person OR persons) and other shareholders` (see I_1^p in Table 1). In this instantiation, N$_1$ ("person") is not a hyponym of N$_2$ ("shareholder"), instead N$_2$ is a hyponym of N$_1$. This is a consequence of the substitution of the antecedent ("Mr. Pickens") with its NE category ("person") (see also Figure 1). Such an instantiation is normally not very frequent, since it violates standard relations within (O1). Therefore, we also instantiate (O1) by substituting N$_1$ with *ana*, and N$_2$ with the NE category of the antecedent (see I_2^p in Table 1).

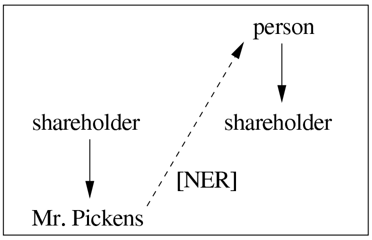

Figure 1: NER and Hyponymy Relation

Furthermore, for NE antecedents, we use an additional pattern (O2):

(O2) N$_1$ `and other` N$_2${*pl*}

We instantiate it by substituting N$_1$ with the original NE antecedent, and N$_2$ with *ana* (see I_3^p in Table 1).

Patterns and instantiations are summarised in Table 1. We instantiate the patterns for each anaphor/antecedent pair and submit these instantiations as queries to the GOOGLE search engine.

3.3 Scoring Method

For each antecedent *ant* in $\mathcal{A}$ we obtain the raw frequencies of all instantiations it occurs in (I_1^c for common nouns, or I_1^p, I_2^p, I_3^p for proper names) from the Web, yielding $freq(I_1^c)$, or $freq(I_1^p)$, $freq(I_2^p)$ and $freq(I_3^p)$. We compute the maximum M_{ant} over these frequencies for proper names. For common nouns M_{ant} corresponds to $freq(I_1^c)$. The instantiation yielding M_{ant} is then called $Imax_{ant}$.

We then use two scoring methods. In our first method, we select the antecedent with the highest M_{ant} as the correct antecedent.

The second method takes into account the individual frequencies of *ant* and *ana* by adapting *mutual information*. We call the first part of $Imax_{ant}$ (e.g. "age OR ages", or "shareholder OR shareholders") X_{ant}, and the second part (e.g. "factors" or "persons") Y_{ant}. We compute the probability of $Imax_{ant}$, X_{ant} and Y_{ant}, using GOOGLE to determine $freq(X_{ant})$ and $freq(Y_{ant})$.

[9] We extracted the gazetteers from the Web. Small gazetteers, containing in all about 500 entries, are sufficient. This is the only external knowledge source we collected.

[10] In all the patterns in this paper, "OR" is the boolean operator, "N$_1$" and "N$_2$" are variables, and all other words are constants.

[11] Common noun instantiations are marked by a superscript "c" and proper name instantiations are marked by a superscript "p".

Table 1: Patterns and Instantiations for other-anaphora

ANTECEDENT	PATTERN	INSTANTIATIONS
common noun	(O1): `(N`$_1\{sg\}$ `OR N`$_1\{pl\}$ `and other N`$_2\{pl\}$	I_1^c: "(age OR ages) and other factors"
proper name	(O1): `(N`$_1\{sg\}$ `OR N`$_1\{pl\}$`) and other N`$_2\{pl\}$	I_1^p: "(person OR persons) and other shareholders"
		I_2^p: "(shareholder OR shareholders) and other persons"
	(O2): `N`$_1$ `and other N`$_2\{pl\}$	I_3^p: "Mr. Pickens and other shareholders"

$$Pr(Imax_{ant}) = \frac{M_{ant}}{\text{number of GOOGLE pages}}$$

$$Pr(X_{ant}) = \frac{freq(X_{ant})}{\text{number of GOOGLE pages}}$$

$$Pr(Y_{ant}) = \frac{freq(Y_{ant})}{\text{number of GOOGLE pages}}$$

We then compute the final score MI_{ant}.

$$MI_{ant} = log\frac{Pr(Imax_{ant})}{Pr(X_{ant})Pr(Y_{ant})}$$

We resolve to the antecedent with the highest MI_{ant}.

In both methods, if two antecedents achieve the same score, a recency based tie-breaker chooses the antecedent closest to the anaphor in the text.

3.4 Results and Error Analysis

We postulate three categories for classifying the results: (i) *correct*, when the antecedent selected is the correct one; (ii) *lenient*, when the antecedent selected refers to the same entity as the correct antecedent; (iii) *wrong* in all other cases.

In Table 2, we compare our results with those obtained by the algorithm LEX (Modjeska, 2002). Although LEX makes extensive use of WordNet, our algorithm achieves comparable results.

Our algorithm's mistakes are due to several factors.

NEs. As 58 (48.3%) out of the total 120 anaphors (48.3%) have NE antecedents, NE resolution is crucial for our algorithm. The low recall of our NE recognition module has a significant impact on our algorithm's performance as it leads to missing instantiations. Moreover, incorrect NE classifications yield incorrect instantiations, although this problem is not very frequent as the precision of our NE resolution module is relatively high (93%).

Table 2: Results for other-anaphora

	Web-raw	Web-mut	LEX
corr	50	58	54
len	7	5	7
tot corr	**57**	**63**	**61**
tot wrong	**63**	**57**	**59**
tot	120	120	120

Vague Anaphors. Anaphors that are semantically vague, such as "issue" or "problem", are not informative enough for a purely semantic-oriented method.

Split Antecedents. Our algorithm cannot handle split antecedents (6 cases in our dataset). In Example (5), the antecedent of "other contract months" is a set of referents consisting of "May" and "July".

(5) The *May* contract, which also is without restraints, ended with a gain of 0.45 cent to 14.26 cents. The *July* delivery rose its daily permissible limit of 0.50 cent a pound to 14.00 cent, while **other contract months** showed near-limit advances.

Our algorithm is not able to distinguish between proper cases of split antecedents (e.g. Example (5)), where a tie-breaker should *not* be applied, and cases where two similar entities are mentioned but only one is the actual antecedent. It is normally necessary to resort to sophisticated inference techniques to make this distinction. As we always apply a tie-breaker, examples such as (5) cannot be handled: only “July” (the most recent antecedent) is selected and the result counts as wrong.

Pronouns. Our algorithm cannot handle pronoun antecedents as they are lexically empty. Contrary to intuition, this does not constitute a significant limitation as only 2 anaphors in our dataset have pronominal antecedents that refer to an entity that is not additionally mentioned by a full NP within the 2 sentence window. In contrast, in Example (4), the referent of the pronoun “he” is also mentioned by the full NP “Mr. Pickens”, and the algorithm is able to resolve the anaphor to “Mr. Pickens”, thus yielding a lenient correct result.

4 Experiment II: Bridging

For the scope of this paper, a bridging anaphor is a definite NP that can be felicitously used only because it refers to an entity which stands in a meronymic relation with an already explicitly introduced entity (see Example (1)).

We use the corpus described in (Poesio et al., 2002), restricting ourselves to the examples classified as meronymy. Unfortunately, this yields only 12 examples so that the current experiment is only a very small pilot study to explore the extension of our method to other nominal anaphora. In addition, we profit from the a priori knowledge that we are dealing with an instance of meronymy (one of the many relations bridging can express), so that we do not operate in a completely realistic scenario. This contrasts with our experiment on other-anaphora, where the modifier *other* (together with the absence of a structurally given antecedent) reliably signals the presence of an anaphor and the lexical relations expressed are more constrained.

For each anaphor, all possible NP antecedents in a 5-sentence window have been already extracted by Renata Vieira and Massimo Poesio, who also had already deleted NEs from the original dataset. Again, we stripped modification so that only the heads of possible antecedents and of the anaphors were used.

Meronymy is often explicitly expressed by the following patterns:

(B1) `(N1{sg} OR N1{pl}) of (a OR an OR the OR each OR every OR any)* N2{sg}`

Table 3: Results for bridging

	Web-raw	Poesio-WN	Poesio-BNC
corr	**7**	**3**	**8**
wrong	**5**	**9**	**4**
tot	12	12	12

(B2) `N1{pl} of (the OR all)* N2{pl}`

We instantiate both patterns by equating N_1 with the anaphor and N_2 with the antecedent. In Example (1) the instantiations for the antecedent “apartment” are `(floor OR floors) of (a OR an OR the OR each OR every OR any)* apartment` and `floors of (the OR all)* apartments`.

For each antecedent *ant* we obtain the raw frequencies of all instantiations it occurs in from the Web. We then compute the maximum M_{ant} over these frequencies and select the antecedent with the highest M_{ant} as the correct one. We will use mutual information for bridging on a larger dataset.

Table 3 compares our results with those obtained by the algorithms used in (Poesio et al., 2002), of which one relies on WordNet and the other on knowledge a priori extracted from a parsed version of the BNC.

In this preliminary study, our results outperform the Wordnet method and are comparable to those obtained from corpus-based knowledge extraction, although we do not linguistically process the web pages returned by our search.

5 Related Work

Most of the current resolution algorithms for non-pronominal anaphora make heavy use of handcrafted ontologies. The COCKTAIL system for coreference resolution (Harabagiu and Maiorano, 1999) combines sortal constraints and conceptual glosses from WordNet with co-occurence information from a treebank. (Vieira and Poesio, 2000)’s system for definite descriptions (covering, inter alia, coreference and bridging) also makes use of WordNet, as well as handcrafted constraints and consistency checks. LEX, a resolution algorithm developed for other-anaphors (Modjeska, 2002), employs lexical information from WordNet and heuristics for resolving anaphora with NE antecedents, presupposing they have previously been classified into MUC-7 categories.[12] All these approaches suffer from the shortcomings that we outlined in Section 1.

In contrast, we do not use any external handcrafted knowledge. In addition, we use only shallow search patterns and do not process the pages that GOOGLE returns in any way.[13] We achieve results comparable to knowledge- or processing-intensive methods. Moreover, we only compare the likelihood of several given antecedents to cooccur in a given pattern with a given anaphor instead of assuming context-independent lexical relations. Thus, we can resolve anaphor/antecedent relations that might or should not be included in a lexical hierarchy (e.g. risk factor/age).

Our results confirm results by (Keller et al., 2002) and (Grefenstette, 1999), who use the Web successfully for other NLP applications (for an overview of successful usage of the Web in NLP, see (Keller et al., 2002)). In line with these results, ours also show that the large amount of data available on the Web overcomes its intrinsic noise as well as the lack of linguistic processing. To our knowledge, ours is the first attempt to tackle anaphora resolution using the Web.

6 Contributions and Future Work

We have proposed a novel method for non-pronominal anaphora resolution, using simple Web searches with shallow linguistic patterns. We show that the large amount of data available on the Web makes anaphora resolution without handcrafted lexical knowledge feasible.

In particular, we have described two experiments carried out on two different anaphoric phenomena, namely other-anaphora and bridging. Exploiting free text achieves results comparable to those obtained when using rich and structured handcrafted resources.

Given the shallow techniques used and the state-of-the-art results obtained, our method is promising. There is still room for improvement in several directions. Our NE resolution module, for example, is extremely simple and has a low recall (82%). As a large number of antecedents for other-anaphora are NEs (48.3%), including a state-of-the-art NE resolution system would certainly improve our algorithm's performance.

For each lexical relation we use variations of a single pattern. In the future we will explore the use of additional substantially different patterns.

As this is pilot study, the datasets we use are small. We are currently testing our algorithm on a larger dataset of other-anaphora.

[12]In the current evaluation of LEX, NEs are manually annotated.

[13]This is in contrast to other approaches that extract relations from corpora and that use chunkers/taggers/parsers for this purpose (Hearst, 1992; Poesio et al., 2002).

Acknowledgments

Natalia N. Modjeska is supported by grant No. GR/M75129 from the Engineering and Physical Sciences Research Council to the University of Edinburgh. Katja Markert is supported by an Emmy Noether Fellowship of the Deutsche Forschungsgemeinschaft. We thank Mark Chignell at the University of Toronto for providing Natalia N. Modjeska with a good work environment in the Interactive Media Lab.

We also would like to thank Massimo Poesio for providing the bridging data, as well as Bonnie Webber, Johan Bos and two anonymous reviewers for helpful comments and suggestions.

References

Matthew Berland and Eugene Charniak. 1999. Finding parts in very large corpora. In *Proc. of ACL-99*, pages 57–64.

Gann Bierner. 2001. Alternative phrases and natural language information retrieval. In *Proc. of ACL-01*.

Nancy Chinchor. 1997. MUC-7 Named Entity Task definition. In *Proc. of MUC-7, 1997*.

Herbert H. Clark. 1975. Bridging. In *Proc. of the Conference on Theoretical Issues in NLP, 1975*, pages 169–174.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, Mass.

Gregory Grefenstette. 1999. The WWW as a resource for example-based MT tasks. In *Proc. of ASLIB'99 Translating and the Computer 21*.

Sanda Harabagiu and Steven Maiorano. 1999. Knowledge-lean coreference resolution and its relation to textual cohesion and coherence. In *Proc. of the ACL-99 Workshop on the Relation of Discourse and Dialogue Structure and Reference*, pages 29–38.

Marti Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proc. of COLING-92*.

Frank Keller, Maria Lapata, and Olga Ourioupina. 2002. Using the Web to overcome data sparseness. In *Proc. of EMNLP-02*, pages 230–237.

Natalia N. Modjeska. 2002. Lexical and grammatical role constraints in resolving other-anaphora. In *Proc. of DAARC-02*.

Vincent Ng and Claire Cardie. 2002. Improving machine learning approaches to coreference resolution. In *Proc. of ACL-02*, pages 104–111.

Massimo Poesio, Tomonori Ishikawa, Sabine Schulte im Walde, and Renata Viera. 2002. Acquiring lexical knowledge for anaphora resolution. In *Proc. of LREC-02*, pages 1220-1224.

Michael Strube and Udo Hahn. 1999. Functional centering — grounding referential coherence in information structure. *Computational Lingustics*, 25(3):309–344.

Renata Vieira and Massimo Poesio. 2000. An empirically-based system for processing definite descriptions. *Computational Linguistics*, 26(4).