Robust pronoun resolution with limited knowledge

Ruslan Mitkov
School of Languages and European Studies
University of Wolverhampton
Stafford Street
Wolverhampton WV1 1SB
United Kingdom
R.Mitkov@wlv.ac.uk

Abstract

Most traditional approaches to anaphora resolution rely heavily on linguistic and domain knowledge. One of the disadvantages of developing a knowledge-based system, however, is that it is a very labour-intensive and time-consuming task. This paper presents a robust, knowledge-poor approach to resolving pronouns in technical manuals, which operates on texts pre-processed by a part-of-speech tagger. Input is checked against agreement and for a number of antecedent indicators. Candidates are assigned scores by each indicator and the candidate with the highest score is returned as the antecedent. Evaluation reports a success rate of 89.7% which is better than the success rates of the approaches selected for comparison and tested on the same data. In addition, preliminary experiments show that the approach can be successfully adapted for other languages with minimum modifications.

1. Introduction

For the most part, anaphora resolution has focused on traditional linguistic methods (Carbonell & Brown 1988; Carter 1987; Hobbs 1978; Ingria & Stallard 1989; Lappin & McCord 1990; Lappin & Leass 1994; Mitkov 1994; Rich & LuperFoy 1988; Sidner 1979; Webber 1979). However, to represent and manipulate the various types of linguistic and domain knowledge involved requires considerable human input and computational expense.

While various alternatives have been proposed, making use of e.g. neural networks, a situation semantics framework, or the principles of reasoning with uncertainty (e.g. Connoly et al. 1994; Mitkov 1995; Tin & Akman 1995), there is still a strong need for the development of robust and effective strategies to meet the demands of practical NLP systems, and to enhance further the automatic processing of growing language resources.

Several proposals have already addressed the anaphora resolution problem by deliberately limiting the extent to which they rely on domain and/or linguistic knowledge (Baldwin 1997; Dagan & Itai 1990; Kennedy & Boguraev 1996; Mitkov 1998; Nasukawa 1994; Williams et al. 1996). Our work is a continuation of these latest trends in the search for inexpensive, fast and reliable procedures for anaphora resolution. It is also an example of how anaphors in a specific genre can be resolved quite successfully without any sophisticated linguistic knowledge or even without parsing. Finally, our evaluation shows that the basic set of antecedent tracking indicators can work well not only for English, but also for other languages (in our case Polish and Arabic).

2. The approach

With a view to avoiding complex syntactic, semantic and discourse analysis (which is vital for real-world applications), we developed a robust, knowledge-poor approach to pronoun resolution which does not parse and analyse the input in order to identify antecedents of anaphors. It makes use of only a part-of-speech tagger, plus simple noun phrase rules (sentence constituents are identified at the level of noun phrase at most) and operates on the basis of antecedent-tracking preferences (referred to hereafter as "antecedent indicators"). The approach works as follows: it takes as an input the output of a text processed by a part-of-speech tagger, identifies the noun phrases which precede the anaphor within a distance of 2 sentences, checks them for gender and number agreement with the anaphor and then applies the genre-specific antecedent indicators to the remaining candidates (see next section). The noun phrase with the highest aggregate score is proposed as antecedent; in the rare event of a tie, priority is given to the candidate with the higher score for immediate reference. If immediate reference has not been identified, then priority is given to the candi-

date with the best collocation pattern score. If this does not help, the candidate with the higher score for indicating verbs is preferred. If still no choice is possible, the most recent from the remaining candidates is selected as the antecedent.

2.1 Antecedent indicators

Antecedent indicators (preferences) play a decisive role in tracking down the antecedent from a set of possible candidates. Candidates are assigned a score (-1, 0, 1 or 2) for each indicator; the candidate with the highest aggregate score is proposed as the antecedent. The antecedent indicators have been identified empirically and are related to salience (definiteness, givenness, indicating verbs, lexical reiteration, section heading preference, "non-prepositional" noun phrases), to structural matches (collocation, immediate reference), to referential distance or to preference of terms. Whilst some of the indicators are more genre-specific (term preference) and others are less genre-specific ("immediate reference"), the majority appear to be genre-independent. In the following we shall outline some the indicators used and shall illustrate them by examples.

Definiteness

Definite noun phrases in previous sentences are more likely antecedents of pronominal anaphors than indefinite ones (definite noun phrases score 0 and indefinite ones are penalised by -1). We regard a noun phrase as definite if the head noun is modified by a definite article, or by demonstrative or possessive pronouns. This rule is ignored if there are no definite articles, possessive or demonstrative pronouns in the paragraph (this exception is taken into account because some English user's guides tend to omit articles).

Givenness

Noun phrases in previous sentences representing the "given information" (theme)[1] are deemed good candidates for antecedents and score 1 (candidates not representing the theme score 0). In a coherent text (Firbas 1992), the given or known information, or theme, usually appears first, and thus forms a co-referential link with the preceding text. The new information, or rheme, provides some information about the theme.

[1]We use the simple heuristics that the given information is the first noun phrase in a non-imperative sentence.

Indicating verbs

If a verb is a member of the Verb_set = {discuss, present, illustrate, identify, summarise, examine, describe, define, show, check, develop, review, report, outline, consider, investigate, explore, assess, analyse, synthesise, study, survey, deal, cover}, we consider the first NP following it as the preferred antecedent (scores 1 and 0). Empirical evidence suggests that because of the salience of the noun phrases which follow them, the verbs listed above are particularly good indicators.

Lexical reiteration

Lexically reiterated items are likely candidates for antecedent (a NP scores 2 if is repeated within the same paragraph twice or more, 1 if repeated once and 0 if not). Lexically reiterated items include repeated synonymous noun phrases which may often be preceded by definite articles or demonstratives. Also, a sequence of noun phrases with the same head counts as lexical reiteration (e.g. "toner bottle", "bottle of toner", "the bottle").

Section heading preference

If a noun phrase occurs in the heading of the section, part of which is the current sentence, then we consider it as the preferred candidate (1, 0).

"Non-prepositional" noun phrases

A "pure", "non-prepositional" noun phrase is given a higher preference than a noun phrase which is part of a prepositional phrase (0, -1). Example:

> Insert the cassette$_i$ into the VCR making sure it$_i$ is suitable for the length of recording.

Here "the VCR" is penalised (-1) for being part of the prepositional phrase "into the VCR".

This preference can be explained in terms of salience from the point of view of the centering theory. The latter proposes the ranking "subject, direct object, indirect object" (Brennan et al. 1987) and noun phrases which are parts of prepositional phrases are usually indirect objects.

Collocation pattern preference

This preference is given to candidates which have an identical collocation pattern with a pronoun (2,0). The collocation preference here is restricted to the patterns "noun phrase (pronoun), verb" and "verb,

noun phrase (pronoun)". Owing to lack of syntactic information, this preference is somewhat weaker than the collocation preference described in (Dagan & Itai 1990). Example:

> Press the key$_i$ down and turn the volume up... Press it$_i$ again.

Immediate reference

In technical manuals the "immediate reference" clue can often be useful in identifying the antecedent. The heuristics used is that in constructions of the form "...(You) V_1 NP ... con (you) V_2 it (con (you) V_3 it)", where con $\in$ {and/or/before/after...}, the noun phrase immediately after V_1 is a very likely candidate for antecedent of the pronoun "it" immediately following V_2 and is therefore given preference (scores 2 and 0).

This preference can be viewed as a modification of the collocation preference. It is also quite frequent with imperative constructions. Example:

> To print the paper, you can stand the printer$_i$ up or lay it$_i$ flat.
> To turn on the printer, press the Power button$_i$ and hold it$_i$ down for a moment.
> Unwrap the paper$_i$, form it$_i$ and align it$_i$, then load it$_i$ into the drawer.

Referential distance

In complex sentences, noun phrases in the previous clause[2] are the best candidate for the antecedent of an anaphor in the subsequent clause, followed by noun phrases in the previous sentence, then by nouns situated 2 sentences further back and finally nouns 3 sentences further back (2, 1, 0, -1). For anaphors in simple sentences, noun phrases in the previous sentence are the best candidate for antecedent, followed by noun phrases situated 2 sentences further back and finally nouns 3 sentences further back (1, 0, -1).

Term preference

NPs representing terms in the field are more likely to be the antecedent than NPs which are not terms (score 1 if the NP is a term and 0 if not).

[2]Identification of clauses in complex sentences is done heuristically.

As already mentioned, each of the antecedent indicators assigns a score with a value {-1, 0, 1, 2}. These scores have been determined experimentally on an empirical basis and are constantly being updated. Top symptoms like "lexical reiteration" assign score "2" whereas "non-prepositional" noun phrases are given a negative score of "-1". We should point out that the antecedent indicators are preferences and not absolute factors. There might be cases where one or more of the antecedent indicators do not "point" to the correct antecedent. For instance, in the sentence "Insert the cassette into the VCR$_i$ making sure it$_i$ is turned on", the indicator "non-prepositional noun phrases" would penalise the correct antecedent. When all preferences (antecedent indicators) are taken into account, however, the right antecedent is still very likely to be tracked down - in the above example, the "non-prepositional noun phrases" heuristics (penalty) would be overturned by the "collocational preference" heuristics.

2.2 Informal description of the algorithm

The algorithm for pronoun resolution can be described informally as follows:

1. Examine the current sentence and the two preceding sentences (if available). Look for noun phrases[3] only to the left of the anaphor[4]
2. Select from the noun phrases identified only those which agree in gender and number[5] with the pronominal anaphor and group them as a set of potential candidates
3. Apply the antecedent indicators to each potential candidate and assign scores; the candidate with the highest aggregate score is proposed as

[3]A sentence splitter would already have segmented the text into sentences, a POS tagger would already have determined the parts of speech and a simple phrasal grammar would already have detected the noun phrases

[4]In this project we do not treat cataphora; non-anaphoric "it" occurring in constructions such as "It is important", "It is necessary" is eliminated by a "referential filter"

[5]Note that this restriction may not always apply in languages other than English (e.g. German); on the other hand, there are certain collective nouns in English which do not agree in number with their antecedents (e.g. "government", "team", "parliament" etc. can be referred to by "they"; equally some plural nouns (e.g. "data") can be referred to by "it") and are exempted from the agreement test. For this purpose we have drawn up a comprehensive list of all such cases; to our knowledge, no other computational treatment of pronominal anaphora resolution has addressed the problem of "agreement exceptions".

antecedent. If two candidates have an equal score, the candidate with the higher score for immediate reference is proposed as antecedent. If immediate reference does not hold, propose the candidate with higher score for collocational pattern. If collocational pattern suggests a tie or does not hold, select the candidate with higher score for indicating verbs. If this indicator does not hold again, go for the most recent candidate.

3. Evaluation

For practical reasons, the approach presented does not incorporate syntactic and semantic information (other than a list of domain terms) and it is not realistic to expect its performance to be as good as an approach which makes use of syntactic and semantic knowledge in terms of constraints and preferences. The lack of syntactic information, for instance, means giving up c-command constraints and subject preference (or on other occasions object preference, see Mitkov 1995) which could be used in center tracking. Syntactic parallelism, useful in discriminating between identical pronouns on the basis of their syntactic function, also has to be forgone. Lack of semantic knowledge rules out the use of verb semantics and semantic parallelism. Our evaluation, however, suggests that much less is lost than might be feared. In fact, our evaluation shows that the results are comparable to syntax-based methods (Lappin & Leass 1994). We believe that the good success rate is due to the fact that a number of antecedent indicators are taken into account and no factor is given absolute preference. In particular, this strategy can often override incorrect decisions linked with strong centering preference (Mitkov & Belguith 1998) or syntactic and semantic parallelism preferences (see below).

3.1 *Evaluation A*

Our first evaluation exercise (Mitkov & Stys 1997) was based on a random sample text from a technical manual in English (Minolta 1994). There were 71 pronouns in the 140 page technical manual; 7 of the pronouns were non-anaphoric and 16 exophoric. The resolution of anaphors was carried out with a success rate of 95.8%. The approach being robust (an attempt is made to resolve each anaphor and a proposed antecedent is returned), this figure represents both "precision" and "recall" if we use the MUC terminology. To avoid any terminological confusion, we shall therefore use the more neutral term "success rate" while discussing the evaluation.

In order to evaluate the effectiveness of the approach and to explore if / how far it is superior over the baseline models for anaphora resolution, we also tested the sample text on (i) a Baseline Model which checks agreement in number and gender and, where more than one candidate remains, picks as antecedent the most recent subject matching the gender and number of the anaphor (ii) a Baseline Model which picks as antecedent the most recent noun phrase that matches the gender and number of the anaphor.

The success rate of the "Baseline Subject" was 29.2%, whereas the success rate of "Baseline Most Recent NP" was 62.5%. Given that our knowledge-poor approach is basically an enhancement of a baseline model through a set of antecedent indicators, we see a dramatic improvement in performance (95.8%) when these preferences are called upon.

Typically, our preference-based model proved superior to both baseline models when the antecedent was neither the most recent subject nor the most recent noun phrase matching the anaphor in gender and number. Example:

> Identify the drawer$_i$ by the lit paper port LED and add paper to it$_i$.

The aggregate score for "the drawer" is 7 (definiteness 1 + givenness 0 + term preference 1 + indicating verbs 1 + lexical reiteration 0 + section heading 0 + collocation 0 + referential distance 2 + non-prepositional noun phrase 0 + immediate reference 2 = 7), whereas aggregate score for the most recent matching noun phrase (“the lit paper port LED”) is 4 (definiteness 1 + givenness 0 + term preference 1 + indicating verbs 0 + lexical reiteration 0 + section heading 0 + collocation 0 + referential distance 2 + non-prepositional noun phrase 0 + immediate reference 0 = 4).

From this example we can also see that our knowledge-poor approach successfully tackles cases in which the anaphor and the antecedent have not only different syntactic functions but also different semantic roles. Usually knowledge-based approaches have difficulties in such a situation because they use preferences such as "syntactic parallelism" or "semantic parallelism". Our robust approach does not use these because it has no information about the syntactic structure of the sentence or about the syntactic function/semantic role of each individual word.

As far as the typical failure cases are concerned, we anticipate the knowledge-poor approach to have difficulties with sentences which have a more complex syntactic structure. This should not be surpris-

ing, given that the approach does not rely on any syntactic knowledge and in particular, it does not produce any parse tree. Indeed, the approach fails on the sentence:

> The paper through key can be used to feed [a blank sheet of paper]$_i$ through the copier out into the copy tray without making a copy on it$_i$.

where "blank sheet of paper" scores only 2 as opposed to the "the paper through key" which scores 6.

3.2 *Evaluation B*

We carried out a second evaluation of the approach on a different set of sample texts from the genre of technical manuals (47-page Portable Style-Writer User's Guide (Stylewriter 1994). Out of 223 pronouns in the text, 167 were non-anaphoric (deictic and non-anaphoric "it"). The evaluation carried out was manual to ensure that no added error was generated (e.g. due to possible wrong sentence/clause detection or POS tagging). Another reason for doing it by hand is to ensure a fair comparison with Breck Baldwin's method, which not being available to us, had to be hand-simulated (see 3.3).

The evaluation indicated 83.6% success rate. The "Baseline subject" model tested on the same data scored 33.9% recall and 67.9% precision, whereas "Baseline most recent" scored 66.7%. Note that "Baseline subject" can be assessed both in terms of recall and precision because this "version" is not robust: in the event of no subject being available, it is not able to propose an antecedent (the manual guide used as evaluation text contained many imperative zero-subject sentences).

In the second experiment we evaluated the approach from the point of view also of its "critical success rate". This measure (Mitkov 1998b) applies only to anaphors "ambiguous" from the point of view of number and gender (i.e. to those "tough" anaphors which, after activating the gender and number filters, still have more than one candidate for antecedent) and is indicative of the performance of the antecedent indicators. Our evaluation established the critical success rate as 82%.

A case where the system failed was when the anaphor and the antecedent were in the same sentence and where preference was given to a candidate in the preceding sentence. This case and other cases suggest that it might be worthwhile reconsidering/refining the weights for the indicator "referential distance".

Similarly to the first evaluation, we found that the robust approach was not very successful on sentences with too complicated syntax - a price we have to pay for the "convenience" of developing a knowledge-poor system.

The results from experiment 1 and experiment 2 can be summarised in the following (statistically) slightly more representative figures.

	Robust approach	Baseline subject	Baseline most recent
Success rate (=Precision/ Recall)	89.7%	31.55% / 48.55%	65.95%

The lower figure in "Baseline subject" corresponds to "recall" and the higher figure - to "precision".

If we regard as "discriminative power" of each antecedent indicator the ratio "number of successful antecedent identifications when this indicator was applied"/"number of applications of this indicator" (for the non-prepositional noun phrase and definiteness being penalising indicators, this figure is calculated as the ratio "number of unsuccessful antecedent identifications"/"number of applications"), the immediate reference emerges as the most discriminative indicator (100%), followed by non-prepositional noun phrase (92.2%), collocation (90.9%), section heading (61.9%), lexical reiteration (58.5%), givenness (49.3%), term preference (35.7%) and referential distance (34.4%). The relatively low figures for the majority of indicators should not be regarded as a surprise: firstly, we should bear in mind that in most cases a candidate was picked (or rejected) as an antecedent on the basis of applying a number of different indicators and secondly, that most anaphors had a relatively high number of candidates for antecedent.

In terms of frequency of use ("number of non-zero applications"/"number of anaphors"), the most frequently used indicator proved to be referential distance used in 98.9% of the cases, followed by term preference (97.8%), givenness (83.3%), lexical reiteration (64.4%), definiteness (40%), section heading (37.8%), immediate reference (31.1%) and collocation (11.1%). As expected, the most frequent indicators were not the most discriminative ones.

3.3 Comparison to similar approaches: comparative evaluation of Breck Baldwin's CogNIAC

We felt appropriate to extend the evaluation of our approach by comparing it to Breck Baldwin's CogNIAC (Baldwin 1997) approach which features "high precision coreference with limited knowledge

and linguistics resources". The reason is that both our approach and Breck Baldwin's approach share common principles (both are knowledge-poor and use a POS tagger to provide the input) and therefore a comparison would be appropriate.

Given that our approach is robust and returns antecedent for each pronoun, in order to make the comparison as fair as possible, we used CogNIAC's "resolve all" version by simulating it manually on the same training data used in evaluation B above.

CogNIAC successfully resolved the pronouns in 75% of the cases. This result is comparable with the results described in (Baldwin 1997). For the training data from the genre of technical manuals, it was rule 5 (see Baldwin 1997) which was most frequently used (39% of the cases, 100% success), followed by rule 8 (33% of the cases, 33% success), rule 7 (11%, 100%), rule 1 (9%, 100%) and rule 3 (7.4%, 100%).

It would be fair to say that even though the results show superiority of our approach on the training data used (the genre of technical manuals), they cannot be generalised automatically for other genres or unrestricted texts and for a more accurate picture, further extensive tests are necessary.

4. Adapting the robust approach for other languages

An attractive feature of any NLP approach would be its language "universality". While we acknowledge that most of the monolingual NLP approaches are not automatically transferable (with the same degree of efficiency) to other languages, it would be highly desirable if this could be done with minimal adaptation.

We used the robust approach as a basis for developing a genre-specific reference resolution approach in Polish. As expected, some of the preferences had to be modified in order to fit with specific features of Polish (Mitkov & Stys 1997). For the time being, we are using the same scores for Polish.

The evaluation for Polish was based technical manuals available on the Internet (Internet Manual, 1994; Java Manual 1998). The sample texts contained 180 pronouns among which were 120 instances of exophoric reference (most being zero pronouns). The robust approach adapted for Polish demonstrated a high success rate of 93.3% in resolving anaphors (with critical success rate of 86.2%).

Similarly to the evaluation for English, we compared the approach for Polish with (i) a Baseline Model which discounts candidates on the basis of agreement in number and gender and, if there were still competing candidates, selects as the antecedent the most recent subject matching the anaphor in gender and number (ii) a Baseline Model which checks agreement in number and gender and, if there were still more than one candidate left, picks up as the antecedent the most recent noun phrase that agrees with the anaphor.

Our preference-based approach showed clear superiority over both baseline models. The first Baseline Model (Baseline Subject) was successful in only 23.7% of the cases, whereas the second (Baseline Most Recent) had a success rate of 68.4%. Therefore, the 93.3% success rate (see above) demonstrates a dramatic increase in precision, which is due to the use of antecedent tracking preferences.

We have recently adapted the approach for Arabic as well (Mitkov & Belguith 1998). Our evaluation, based on 63 examples (anaphors) from a technical manual (Sony 1992), indicates a success rate of 95.2% (and critical success rate 89.3 %).

5. Conclusion

We have described a robust, knowledge-poor approach to pronoun resolution which operates on texts pre-processed by a part-of-speech tagger. Evaluation shows a success rate of 89.7% for the genre of technical manuals and at least in this genre, the approach appears to be more successful than other similar methods. We have also adapted and evaluated the approach for Polish (93.3 % success rate) and for Arabic (95.2% success rate).

References

Baldwin, Breck. 1997. "CogNIAC: high precision coreference with limited knowledge and linguistic resources". *Proceedings of the ACL'97/EACL'97 workshop on Operational factors in practical, robust anaphora resolution,* 38-45, Madrid, Spain.

Brennan, S., M. Fridman and C. Pollard. 1987. A centering approach to pronouns. *Proceedings of the 25th Annual Meeting of the ACL (ACL'87)*, 155-162, Stanford, CA, USA.

Carbonell, James G. & Ralf D. Brown. 1988. "Anaphora resolution: a multi-strategy approach". *Proceedings of the 12. International Conference on Computational Linguistics (COLING'88)*, Vol.I, 96-101, Budapest, Hungary.

Carter, David M. 1987. *Interpreting anaphora in natural language texts*. Chichester: Ellis Horwood

Connoly, Dennis, John D. Burger & David S. Day. 1994. "A Machine learning approach to anaphoric reference". *Proceedings of the International Conference "New Methods in Language Processing"*, 255-261, Manchester, United Kingdom.

Dagan, Ido & Alon Itai. 1990. "Automatic processing of large corpora for the resolution of anaphora references". *Proceedings of the 13theInternational Conference on*

Computational Linguistics (COLING'90), Vol. III, 1-3, Helsinki, Finland

Firbas, Jan. 1992. *Functional sentence perspective in written and spoken communication*. Cambridge: Cambridge University Press

Hobbs, Jerry R. 1978 "Resolving pronoun references". *Lingua*, 44, 339-352.

Ingria, Robert J.P. & David Stallard. 1989. "A computational mechanism for pronominal reference". *Proceedings of the 27th Annual Meeting of the ACL*, 262-271, Vancouver, British Columbia.

Internet Manual. 1994. *Translation of Internet Manual Internet i okolice: Przewodnik po swiatowych sieciach komputerowych*. Tracy LaQuey, Jeanne C. Ryer Translated by Monika Zielinska, BIZNET Poland.

Java Manual. 1998. *Jezyk Java.* Clico, Krakow.

Kennedy, Christopher & Branimir Boguraev, 1996. "Anaphora for everyone: pronominal anaphora resolution without a parser". *Proceedings of the 16th International Conference on Computational Linguistics (COLING'96)*, 113-118. Copenhagen, Denmark

Lappin, Shalom & Michael McCord. 1990. "Anaphora resolution in slot grammar". *Computational Linguistics*, 16:4, 197-212.

Lappin, Shalom & Herbert Leass. 1994. "An algorithm for pronominal anaphora resolution". *Computational Linguistics*, 20(4), 535-561.

Minolta. 1994. *Minolta Operator's Manual for Photocopier EP5325*. Technical Manual Minolta Camera Co., Ltd., Business Equipment Division 3-13, 2-Chome, Azuchi, -Machi, Chuo-Ku, Osaka 541, Japan

Mitkov, Ruslan. 1994. "An integrated model for anaphora resolution". *Proceedings of the 15th International Conference on Computational Linguistics (COLING'94)*, 1170-1176, Kyoto, Japan.

Mitkov, Ruslan. 1995. "Un uncertainty reasoning approach for anaphora resolution". *Proceedings of the Natural Language Processing Pacific Rim Symposium (NLPRS'95)*, 149-154, Seoul, Korea.

Mitkov, Ruslan. 1998a. "Pronoun resolution: the practical alternative". In T. McEnery, S. Botley(Eds) *Discourse Anaphora and Anaphor Resolution*. John Benjamins.

Mitkov, Ruslan. 1998b. “Evaluating anaphora resolution approaches”. *Proceedings of the Discourse Anaphora and Anaphora Resolution Colloquium (DAARC'2),* (forthcoming) Lancaster, UK

Mitkov, Ruslan & Malgorzata Stys. 1997. "Robust reference resolution with limited knowledge: high precision genre-specific approach for English and Polish". *Proceedings of the International Conference "Recent Advances in Natural Language Proceeding" (RANLP'97)*, 74-81, Tzigov Chark, Bulgaria.

Mitkov, Ruslan & Lamia Belguith. 1998. “Pronoun resolution made simple: a robust, knowledge-poor approach in action”. *Proceedings of the International Conference “Traduction Automatique et Langage Naturel” (TALN'98*) (forthcoming). Paris, France.

Rich, Elaine & Susann LuperFoy. 1988. "An architecture for anaphora resolution". *Proceedings of the Second Conference on Applied Natural Language Processing (ANLP-2)*, 18-24, Texas, U.S.A.

Sidner, Candy L. 1979. *Towards a computational theory of definite anaphora comprehension in English discourse*. Technical Report No. 537. M.I.T., Artificial Intelligence Laboratory.

Sony. 1992. *Video cassette recorder*. Operating Instructions. Sony Corporation.

Stylewriter 1994. *Portable StyleWriter*. User's guide. Apple Computers.

Tin, Erkan & Varol, Akman. 1994. "Situated processing of pronominal anaphora". *Proceedings of the KONVENS"94 Conference*, 369-378, Webber, Bonnie L. 1979. *A formal approach to discourse anaphora*. London: Garland Publishing.

Williams, Sandra, Mark Harvey & Keith Preston. 1996. "Rule-based reference resolution for unrestricted text using part-of-speech tagging and noun phrase parsing". *Proceedings of the International Colloquium on Discourse Anaphora and Anaphora Resolution (DAARC)*, 441-456. Lancaster, UK.