THE LATEST IN ANAPHORA RESOLUTION: GOING MULTILINGUAL

Ruslan Mitkov

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

Abstract

Until recently, no anaphora resolution projects had looked at the multilingual aspects of the approaches that have been developed, or, in particular, how a specific approach could be used or adapted for other languages. In contrast to previous work in the field, this paper describes a multilingual anaphora resolution approach which was initially developed and evaluated for English, but we have also adapted and evaluated it for Polish and Arabic. The first results are very promising with success rates of around 90% and over.

1. Introduction

The recent increased interest in anaphora resolution is linked to the fact that, until a few years ago, most traditional approaches relied 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. Consequently, the need for inexpensive and robust systems, possibly suitable for unrestricted texts, fuelled renewed research efforts (Baldwin 1997; Dagan & Itai 1990; Ferrandez et al. 1997; Kennedy & Boguraev 1996; Mitkov 1996; Mitkov 1998; Nasukawa 1994; Williams et al. 1996) in the field and a clear trend towards corpus-based and knowledge-poor approaches was established.

The developments in anaphora resolution take place in the wider context of NLP where the search for multilingual applications is a live issue. Against the background of growing interest in multilingual work, it was surprising that until recently, no anaphora resolution projects had looked at the multilingual aspects of the approaches that have been developed, or, in particular, at how a specific approach could be used or adapted for other languages.

The last few months, however, have seen the emergence of the first multilingual anaphora resolution projects (Mitkov & Stys 1997; Mitkov et al. 1998; Azzam et al. 1998) and therefore, the establishment of a new trend towards multilinguality in the field.

The project described in this paper has a truly multilingual character. We have developed a knowledge-poor, robust approach which we propose as a platform for multilingual pronoun resolution in technical manuals. The approach was initially developed and tested for English, but we have also adapted and tested it for Polish and Arabic. We found that the approach could be adapted with minimum modification for both languages and moreover, even if used without any modification, it still delivers acceptable success rates. Evaluation shows a success rate of 89.7% for English, 93.3% for Polish and 95.8% for Arabic.

2. The approach: general overview

With a view to avoiding complex syntactic, semantic and discourse analysis, we developed a robust, knowledge-poor approach to pronoun resolution which does not make use of parsing, syntactic and semantic constraints or any other form of linguistic or non-linguistic knowledge. Instead, we rely on the efficiency of sentence segmentation, part-of-speech tagging, noun phrase identification and the high performance of the antecedent indicators (knowledge is limited to a small noun phrase grammar, a list of terms, a list of (indicating) verbs, a list of genre-specific synonyms, and a set of antecedent indicators).

The core of the approach lies in activating a list of multilingual[1] "antecedent indicators" after filtering candidates (from the current and

[1] We term the antecedent indicators "multilingual" because they work well not only for English, but also for other languages (in this case Arabic and Polish).

two preceding sentences) on the basis of gender and number agreement. Before that, the text is pre-processed by a sentence splitter which determines the sentence boundaries, a part-of-speech tagger which identifies the parts of the speech and a simple phrasal grammar which detects the noun phrases (In addition, in the case of complex sentences, heuristic "clause identification" rules track the clause boundaries). Non-anaphoric occurrences of "it" in constructions such as "It is important", "It is necessary" etc., are eliminated by a "referential filter".

After passing the "agreement filter" (for more details on the agreement rules for the languages involved, see Mitkov et al. 1998), the genre-specific antecedent indicators are applied to the remaining candidates (see section 2.1). 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 candidate with the best collocation pattern score. If this does not help, the candidate with the higher score for indicating verbs is preferred. If there is still no possible choice, the most recent of 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 on the basis of empirical studies of numerous hand-annotated technical manuals (referential links had been marked by human experts). These indicators can be related to salience (definiteness/indefiniteness, givenness, indicating verbs, indicating noun phrases, lexical reiteration, section heading preference, "non-prepositional" noun phrases, relative pronoun), to structural matches (collocation, immediate reference, sequential instructions), to referential distance or to preference of terms. Also, the indicators can be "impeding" (non-PP NPs, definiteness/indefiniteness), assigning negative scores to candidates or "boosting" (the rest), assigning positive scores. Whilst some of the indicators are more genre-specific (term preference) and others are less genre-specific ("immediate reference", "sequential instructions" and to a much lesser extent "indicating noun phrases"), the majority appear to be genre-independent. Below we shall outline the indicators used and shall illustrate some of them by examples. Most of the indicators are used in the same way for English, Polish and Arabic and we shall describe them in this work from the point of view of English; for details on indicators defined differently for Polish and Arabic and for details on language-specific indicators (e.g. "Relative pronoun indicator" for Arabic, see Mitkov et al. 1998).

Definiteness/Indefiniteness

Indefinite noun phrases in previous sentences are less likely antecedents of pronominal anaphors than definite ones and are penalised by -1. In English 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 guides tend to omit articles).

Givenness

Noun phrases in previous sentences representing the "given information" (theme)[2] are deemed good candidates for antecedents and score 1 (candidates not representing the theme score 0). In a coherent text, the given or known information, or theme, usually appears first, and thus forms a co-referential link with the preceding text (Firbas 1992). The new information, or rheme, provides some information.

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 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 it 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").

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

Section heading preference

If a noun phrase occurs in the heading of a 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 (here "the VCR" is penalised (-1) for being part of the prepositional phrase "into the VCR"):

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

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)" or "noun phrase (pronoun), verb, adjective/past participle" if the verb is "to be". Owing to lack of syntactic information, this preference is somewhat weaker than the collocation preference described in (Dagan & Itai 1990).

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

The collocation pattern preference has been extended to patterns "(un)V-NP/anaphor" and "NP/anaphor – (un)V", i.e. verbs with an "undoing action" meaning are considered to fall into collocation patterns along with their "doing action" counterparts. This new extended rule helps in cases such as "Loading a cassette or unloading it". Also, we would consider a certain pattern still a collocation, if the verb featured as a gerund (e.g. "When you plug in the power adapter, the print head moves to its protected position (you'll hear it moving)", Stylewriter 1994).

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 in imperative constructions.

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.

Sequential instructions

This new antecedent indicator has recently been incorporated for Arabic but it works equally well for English and will shortly be implemented in the English version as well. It states that in sequential instructions of the form "To V_1 NP_1, V_2 NP_2. (Sentence). To V_3 it, V_4 NP_4" the noun phrase NP_1 is the likely antecedent of the anaphor "it" (NP_1 is assigned a score of 2). Example:

To turn on the video recorder$_i$, press the red button. To programme it$_i$, press the "Programme" key.

Referential distance

In English complex sentences, noun phrases in the previous clause[3] are the best candidates 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).

[3]The 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 are very likely to be overturned by the "collocational preference" heuristics since the collocation "The VCR is turned on" is likely to appear previously in the text, being typical of video technical manuals.

3. Evaluation

As in any other NLP task, evaluation is of crucial importance to anaphora resolution. The MUC (Message Understanding Conference) initiatives suggested the measures "recall" and "precision" be used for evaluating the performance of coreference resolution. It is felt, however, that evaluation in anaphora resolution needs further attention. Measuring the success rate of an anaphora resolution system in terms of "recall" and "precision" is undoubtedly an important (and consistent) step towards assessing the efficiency of anaphora resolution approaches, but "recall" and "precision" cannot be seen as distinct measures for robust systems. In addition, it appears that they alone cannot provide a comprehensive overall assessment of an approach. In order to see how much a certain approach is "worth", it would be necessary to evaluate it against other "benchmarks", e.g. against other existing or baseline models.

In order to evaluate the effectiveness of the approach and explore whether it is superior to the baseline models for anaphora resolution and if so by how much, we also tested all sample texts (so far for English and Polish only) 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 (we shall refer to it as "Baseline Subject") and (ii) a Baseline Model which picks out as antecedent the most recent noun phrase that matches the gender and number of the anaphor (we shall refer to it as "Baseline Most Recent"). We have also introduced the measure "critical success rate" which exclusively accounts for the performance of the antecedent indicators since it is associated only with those anaphors which still have more than one candidate for antecedent after gender and number filters i.e. anaphors whose antecedents can be tracked down only on the basis of the antecedent indicators.

In addition, evaluation of the factors/indicators used in anaphora resolution is mandatory since it gives insights as to how the approach can be improved. We consider the evaluation which we carried on the discriminative power of each indicator (defined as the ratio number of successful identifications of the antecedent when the indicator applied / number of all applications of the indicator; for details, see Mitkov et al. 1998) to be of particular importance, since it is very revealing for further improvement and fine-tuning of the scores.

Finally, for a multilingual approach we deem it appropriate to evaluate its "multilinguality" which we measure in terms of the extent of its applicability to another language without modifying it at all. Our results show that even without modifying the approach for Polish and Arabic (see below), the results are highly satisfactory.

We have carried out evaluations on sample texts from technical user guides for English, Arabic and Polish and the results show comparable success rates. The success rate for Arabic is slightly higher and we should mention that in addition to tuning the approach for Arabic, the "Arabic improved" version uses 2 new indicators recently introduced which have not been included in the "Robust English" version yet.

3.1 English

The evaluation shows that the results are comparable to and even better than syntax-based methods (Lappin & Leass 1994). The evaluation results also show superiority over other knowledge-poor methods (Baldwin 1997; see also below)[4]. 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 or syntactic and semantic parallelism preferences (Mitkov 1998b)

The evaluation in English included texts from different technical manuals (Minolta Photocopier, Portable Style Writer, Alba Twin Speed Video Recorder, Seagate Medalist Hard Drive, Haynes Car Manual, Sony Video Re-

[4] This applies to the genre of technical manuals; for other genres results may be different

corder) which contained a total of 223 anaphoric pronouns. The robust approach correctly resolved 200 anaphors which gives a success rate of 89.9%[5]. The success rates were different for each of the technical manuals (Minolta Photocopier 95.8% based on 48 pronouns, Portable Style Writer 83.8% - 54 pronouns, Alba Twin Speed Video Recorder 100% - 13 pronouns, Seagate Medalist Hard Drive - 77.8% - 18 pronouns, Haynes Car Manual - 80% - 50 pronouns, Sony Video Recorder - 90.6% - 40 pronouns) which shows that even for texts belonging to the same genre, results may differ. Therefore, for "more definitive" success rate figures very large test data containing thousands of anaphors, are needed.

We used the data from the Portable Style Writer manual for a comparative evaluation with Breck Baldwin's knowledge-poor approach (Baldwin 1997) which scored 75% on the same data.

We also measured the critical success rate as 82% on the basis of the Portable Style Writer: this figure and the significantly lower success rates of the Baseline Most Recent (success rate 65.9% based on Minolta Photocopier and Portable Style Writer) and of the Baseline Subject (precision 48.6%, recall 31.6%) undoubtedly demonstrate the efficiency of the antecedent indicators.

3.2 Arabic

We evaluated the robust approach for Arabic operating in two modes: the first mode consisted of using the robust approach directly, without any adaptation for Arabic, whereas the second mode used an adapted/enhanced version which included modified rules designed to capture some of the specific aspects of Arabic plus a few new indicators (Mitkov & Belguith 1998).

The evaluation was based on 190 examples (anaphors) from a Sony video technical manual. The first mode (i.e. using the robust approach without any adaptation for Arabic - this version is referred to as "Arabic direct" in the table below) reported a success rate of 77.9% (148 out of 190 anaphors were correctly resolved).

The second evaluation mode (evaluating the version adapted and improved for Arabic which is referred to as "Arabic improved" in the table below) reported a success rate of 95.8% (182 out of 190 anaphors were correctly resolved).

The evaluation for Arabic also showed a very high "critical success rate". The robust approach used without any modification scored a "critical success rate" of 70.4%, whereas the improved Arabic version scored 94.4%.

3.3 Polish

The sample texts from technical manuals in Polish 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.

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, selects as the antecedent the most recent noun phrase that agrees with the anaphor.

The Polish version of our robust approach showed clear superiority over both Polish 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%. These results demonstrate the dramatic increase in precision, which is due to the use of antecedent tracking indicators.

The Polish version also showed a very high "critical success rate" of 86.2%. Used without any modification ("Polish direct"), the approach scored a 90% success rate.

The success rates obtained can be summarised as follows:

	Success rate
Robust English	89.7%
Polish direct	90%
Polish improved	93.3%
Arabic direct	77.9%
Arabic improved	95.8%

Table 1: Success rates of the robust approach

	Success rate
Baseline subject English	31.6% / 48.6%
Baseline most recent English	65.9%
Baseline subject Polish	23.7%
Baseline most recent Polish	68.4%

Table 2: Success rates of the baseline models

[5]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

4. Extension to other languages and genres

We are currently testing the approach for 3 new languages (Finnish, Russian and French) and the preliminary evaluation results confirm its multilingual features. Tested on 90 pronouns in Finnish, the approach scored a success rate of 77.8% (critical success rate 69.2%) when adapted and 70.0% when unmodified. The lower success rate for Finnish could be explained by its complex case system and the fact that the modification for Finish was minimal, but it should be noted that with a better NP extractor the Finnish version could have scored as high as 85.1%. For the evaluation in French (36 pronouns), we used the approach directly (without any language-specific adaptation) and therefore the success rate of 88.9% (81.8% critical success rate) looks quite promising. In Russian (sample of 60 pronouns) the approach performed even better with a 96.7% success rate and a 88.9% critical success rate. Future work includes adapting the approach for Spanish, Japanese (it looks like that the antecedent indicators will have to be modified significantly for Japanese) and Bulgarian as well as testing it on (and if necessary, modifying it to cover) a wider variety of genres.

As for extending the approach to other genres, it looks like the approach could be applied with success to other genres even without any modifications. A preliminary evaluation experiment based on a text from the genre of research papers (the text selected was the paper Mitkov & Lamia 1998) showed that out of 24 pronominal anaphors, our approach correctly resolved 19 pronouns which gives a success rate of 79.1%. We should note that the text chosen was not "easy from the point of view of anaphora resolution" at all: Baseline Most Recent scored only 29.1%!

5. Further work

We plan to extend our multilingual pronoun resolution approach into a multilingual platform for pronoun resolution in technical manuals featuring the same indicators for each language which in turn, will have different weights for each specific language. This will involve optimising and fine-tuning the scores of each indicator on the basis of large training data and to this end we are currently developing a multilingual corpora with annotated anaphoric links .

6. Conclusion

We have outlined a robust, knowledge-poor approach to pronoun and have shown its multilingual nature: we have adapted and tested the approach for Polish and Arabic. The evaluation reports success rates which are comparable to (and even better than) syntax-based methods and show superiority over other methods with limited knowledge. The good performance of the approach is due to its antecedent indicators which are truly multilingual: the results show that even if used unmodified, the antecedent indicators prove to be very successful for other languages.

References

Aiwa. 1996. *Operating Instructions for the Aiwa Compact Disc Stereo System.* Aiwa Corporation.

Alba. 1995. *Alba Twin Speed Video Recorder Instruction Manual.*

Azzam, Saliha, Kevin Humphreys & Robert Gaizauskas. 1998. "Coreference resolution in a multilingual information extraction". *Proceedings of the Workshop on Linguistic Coreference.* Granada, Spain.

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.

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

Dagan, Ido. John Justeson, Shalom Lappin, Herbert Leass & Amnon Ribak. 1995. Syntax and lexical statistics in anaphora resolution. *Applied Artificial Intelligence,* 9.

Ferrandez, Antonio, Manolo Palomar and Moreno L. 1998 "Slot unification grammar and anaphora resolution", 294-299. *Proceedings of the International Conference on Recent Advances in Natural Language Proceeding (RANLP'97),* Tzigov Chark, Bulgaria.

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

Haynes. 1994. *Haynes car manual.*

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

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 & 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. 1996. "Pronoun resolution: the practical alternative". In T. McEnery, S. Botley(Eds) Paper presented at the *Discourse Anaphora and Anaphor Resolution Colloqium (DAARC)*, Lancaster, UK.

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

Mitkov, Ruslan. 1998b. "Robust pronoun resolution with limited knowledge". *Proceedings of the 18.th International Conference on Computational Linguistics (COLING'98)/ACL'98 Conference*, (forthcoming). Montreal, Canada.

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)*,, 42-51. Paris, France.

Mitkov, Ruslan, Lamia Belguith & Malgorzata Stys. 1998. "Multilingual robust anaphora resolution". *Proceedings of the Third International Conference on Empirical Methods in Natural Language Processing (EMNLP-3)*, 7-16. Granada, Spain

Nasukawa, Tetsuya. 1994. "Robust method of pronoun resolution using full-text information". *Proceedings of the 15th International Conference on Computational Linguistics (COLING'94)*, 1157-1163, Kyoto, Japan.

Seagate. 1997. *Seagate Medalist Hard Drive Installation Guide*.

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

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

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.