Incorporating Contextual Cues in Trainable Models for Coreference Resolution

Ryu Iida, Kentaro Inui, Hiroya Takamura and Yuji Matsumoto
Graduate School of Information Science,
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara, 630-0192, Japan
{ryu-i,inui,hiroya-t,matsu}@is.aist-nara.ac.jp

Abstract

We propose a method that incorporates various novel contextual cues into a machine learning for resolving coreference. Distinct characteristics of our model are (i) incorporating more linguistic features capturing contextual information that is more sophisticated than what is offered in Centering Theory, and (ii) a tournament model for selecting a referent. Our experiments show that this model significantly outperforms earlier machine learning approaches, such as Soon et al. (2001).

1 Introduction

Computational approaches to coreference resolution have been roughly evolving in two different but complementary directions. One is theory-oriented rule-based approaches and the other is empirical corpus-based approaches.

In rule-based approaches (Mitkov, 1997; Baldwin, 1995; Nakaiwa and Shirai, 1996; Okumura and Tamura, 1996), efforts have been directed to manual encoding of various linguistic cues into a set of rule . Such cues include, for example, the syntactic role of each target noun phrase, the appearance order of antecedent candidates, and the semantic compatibility between an anaphor and a candidate. Most rule-based approaches are also influenced, to a greater or less extent, by theoretical linguistic work, such as Centering Theory (Grosz et al., 1995; Walker et al., 1994; Kameyama, 1986) and the systemic theory (Halliday and Hasan, 1976). The best-achieved performance in MUC-7 [1] was around 70% precision with 60% recall, which is still far from being satisfactory for many practical applications. Worse still, a rule set tuned for a particular domain is unlikely to work equally for another domain due to domain-dependent properties of coreference patterns. Given these facts, further manual refinements of rule-based models will be prohibitively costly.

Corpus-based empirical approaches, such as (Soon et al., 2001; Ng and Cardie, 2002), on the other hand, are cost effective, while having achieved a performance comparable to the best-performing rule-based systems for the coreference task test sets of MUC-6 and MUC-7. However, they tend to lack an appropriate reference to theoretical linguistic work on coherence and coreference. Given this background, one of the challenging issues we should explore next is to make a good marriage between theoretical linguistic findings and corpus-based empirical methods.

In this paper, we report our attempt to enhance existing trainable coreference resolution models by incorporating such theoretical findings as the features utilized in Centering Theory. In Section 2, we review the decision tree-based model proposed by Soon et al. (2001) and one of its successors devised by Ng and Cardie (2002), the latter of which is referred to as the baseline model in our

[1] The Seventh Message Understanding Conference (1998): www.itl.nist.gov/iaui/894.02/related_projects/muc/

empirical evaluation. In Section 3, we discuss a significant drawback of Ng and Cardie's model and propose two solutions: (a) implementing the centering factors as what we call *centering features*, and (b) introducing a novel searching model, which we call a *tournament model*. We then report the results of our experiments on Japanese zero-anaphora resolution in Section 4. We finally discuss remaining problems and future directions in Section 5.

2 A baseline model

Among previous machine learning approaches to coreference resolution, it is probably reasonable to take the work done by Soon et al. (2001) and Ng and Cardie (2002) as our departing point because their models are reported to have reached a level of performance comparable to state-of-the-art knowledge-based systems.

Soon et al.'s model is designed to operate by recasting anaphora resolution (i.e. detection of the antecedent of a given anaphor) as a classification task. Let us see Figure 1 in order to go into the detail. The figure illustrates a situation where there are eight noun phrases, NP_1 through NP_8, which precede the anaphoric noun phrase *ANP* in question. NP_2 and NP_4, NP_3 and NP_5, and NP_6 and NP_7 are coreferent respectively, and NP_5 (and its coreferent NP_3) is the antecedent of *ANP*. Under this situation, the model detects the antecedent by answering a sequence of candidate-wise boolean classification questions: whether or not NP_i is *ANP*'s antecedent for each $i \in \{1, \ldots, 8\}$.

More precisely, for training, Soon et al.'s model creates a positive instance from an anaphor and its closest antecedent (NP_5-*ANP*) and a negative instance from each of the intervening NPs paired with the anaphor (NP_6-*ANP*, NP_7-*ANP* and NP_8-*ANP*). Analogously, given a target NP for resolution in the test phase (see the box "Resolution process" in Figure 1), the model processes each of its preceding NPs in the right-to-left order, answering a classification question of whether or not it is coreferent, until a positive answer comes up. If all the preceding NPs are classified in the negative, the target NP is judged to be non-anaphoric. For classifier induction, Soon et al. used the C5.0 decision tree induction system, an updated version of C4.5 (Quinlan, 1993), in their experiments with a very limited feature set consisting of mere twelve features.

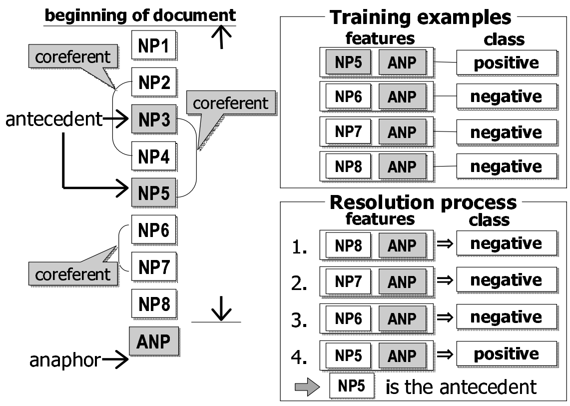

Figure 1: Training example extraction in candidate-wise classification models (Soon et al., 2001; Ng and Cardie, 2002)

Following Soon et al.'s work, Ng and Cardie (2002) improved upon the model by (a) expanding the feature set, and (b) introducing a new search algorithm that searches for the NP with the highest coreference likelihood value. Let us return to Figure 1 for the sake of explanation. In this revised model, while training examples are extracted in the same manner, antecedent detection is done by selecting the highest scored NP from the candidates NP_1 to NP_7. If no candidate is classified in the positive, the target noun phrase is judged to be non-anaphoric. According to Ng and Cardie (2002), their model outperforms the original one, which has also been supported by our experiment on Japanese zero-anaphora resolution reported later. In the rest of the paper, we thus refer to this revised model as the baseline of our empirical evaluation.

Although we do not have the space to go into the detail of the feature set used in Ng and Cardie's experiments, it should be pointed out that their model does not capture an important aspect of local context that has been proved useful for coreference interpretation in the literature of discourse analysis. We elaborate this flaw and propose two solutions in the next section.

3 Incorporating of contextual cues

3.1 A flaw of the baseline model

Consider the following two discourses:

(1) a. Mary went to see John$_i$.

b. He$_i$ was playing baseball.

(2) a. Tom$_i$ went to see John.

b. He$_i$ tried to explain what happened to him yesterday.

In (1), the subject of sentence (b), *He*, refers to the object of sentence (a), *John*. In (2), on the other hand, it is not the case although *He* and *John* fills the same syntactic role, respectively. An explanation for this difference derived from Centering Theory can be briefed as follows. In (2), *Tom* is chosen to be the preferred antecedent of *he* because:

(a) *Tom*, being the subject role filler, is the preferred center (i.e. the highest ranked entity of the forward looking centers) assigned in (a),

(b) *Tom* is thus most likely to be the backward looking center of (b), and

(c) if so is *Tom*, it must be realized as a pronoun.

In (1), on the other hand, *Mary*, the preferred center, violates the gender constraint imposed by *He*, and therefore the second ranked entity *John* is interpreted as the antecedent.

The essence of the above explanation is that it is derived from a model that takes into account the preference between candidates. Whether or not *John* is coreferent depends on the appearance of other entities, such as *Mary* and *Tom*, in its local context. This crucial property of local coherence is, however, not properly captured in Ng and Cardie's model because it views antecedent detection as a set of *candidate-wise* boolean classification problems.

3.2 Two solutions

Among various possibilities one may think of as a solution to the problem argued above, we have empirically examined two novel solutions.

3.2.1 Centering features

A straightforward solution is to augment the number of features that implement local contextual factors. For example, one may introduce a feature that indicates whether or not the antecedent candidate in question is the present preferred center. This feature can also be enhanced so that it can indicate whether or not the candidate is ranked the highest among the forward-looking centers while satisfying gender and number constraints. Such a feature would help the classification model to distinguish the two *John*s in the previous examples. Note that the computation of such features requires the use of additional devices, such as a list for storing forward-looking centers, which has never been used in previous trainable models. We refer to such features as *centering features* for capturing centering state transitions. The centering features we used in our experiments will be presented in the next section.

3.2.2 The tournament model

Recall that what we wanted in *John*'s examples was a model that compares the first *John* with its opponent *Mary* and the second *John* with *Tom*. Our second solution is to implement a pairwise comparison between two candidates in reference to *ANP* as a binary classification problem (i.e. which candidate wins) and to conduct a tournament to check against the candidate. A tournament consists of a series of matches in which candidates compete with each other and the one that prevails through the final round is declared the

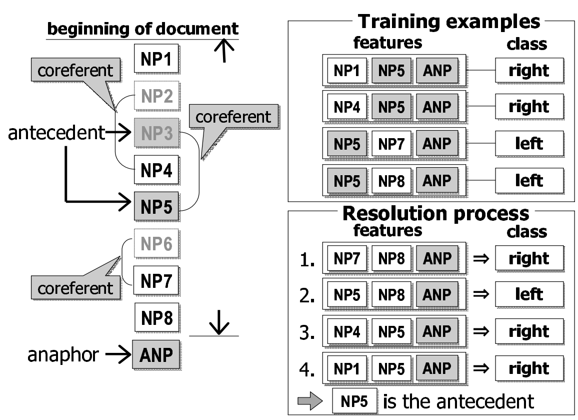

Figure 2: The tournament model

winner, namely, identified as the antecedent. We call this new model the *tournament model*.

Observe the situation given in Figure 1 again, which we have reillustrated here as Figure 2. Now, due to the coreference chains, we have five candidates: NP_1, NP_4 (and its coreferent NP_2), NP_5 (NP_3), NP_7 (NP_6) and NP_8.

Let us first consider the training process . In the tournament, the correct antecedent NP_5 (NP_3) must prevail over any of the other four candidates. We thus extract four training examples from the present case as illustrated in the figure. The class *right* denotes that the succeeding one of a given pair of candidates prevails against (i.e. is more likely to be the antecedent than) the preceding one. Likewise, the class *left* denotes that the preceding candidate prevails over the succeeding one. Finally, we induce from a set of extracted training examples a pair-wise classifier that classifies a given feature vector into either *right* or *left*.

In the test phase, the model conducts a tournament for each given anaphor. In each tournament, it processes the antecedent candidates in the right-to-left order. In the first round, the model consults the trained classifier to judge which of the right-most (closest th *ANP*) two candidates is more likely to be the antecedent. Suppose anew that we are trying to resolve the problem illustrated in Figure 2. As shown in the "resolution process" part of the figure, the first match is arranged between the right-most two candidates NP_8 and NP_7. Here, we assume that NP_8 wins as shown in the figure. Then, each of the following matches is arranged in turn between the winner of the previous match and a right-most new challenger. In the case shown in the figure, the second match is arranged between the current winner NP_8 and the right-most new challenger NP_5. If NP_5 wins, it is next matched against all next challenger NP_4. This process is repeated until the left-most candidate participates. The model selects the candidate that prevails through the final round as the answer.

The introduction of the pairwise classification as above can incorporate the learning of centering factors, such as the expected center order; for example, the model may learn from *Tom* and *John*'s example that the subject role filler is preferred to the object role filler. The tournament model can also encode relational properties between candidates into features. One may, for example, add a feature that indicates the relative distance between a given candidate pair, expecting a tendency that the succeeding candidate is more likely to win when the relative distance between two candidates is longer.

4 Experiments

We made an empirical evaluation taking on Japanese zero-anaphora resolution. Japanese is characterized by an extensive use of zero-pronouns, which behave like pronouns in English texts. The resolution of zero-anaphora has been receiving interest from an increasing number of researchers (Kameyama, 1986; Nariyama, 2002; Nakaiwa and Shirai, 1996; Seki et al., 2002; Yamamoto and Sumita, 1998).

4.1 Training and test sets

We extracted training and test data sets from a corpus with GDA-tagged[2] newspaper articles, which is annotated with coreference relation tags as well as various syntactic and semantic tags. The corpus contains over 25,000 sentences with roughly 20,000 coreference tags annotated. In the experiment, we preliminarily restricted our experiments for resolving subject zero-anaphors, 2,155 instances in total, and conducted five-fold cross-validation on that data set.

4.2 Feature set

We used five types of features as summarized in Table 1: (i) grammatical, (ii) semantic, (iii) positional, (iv) heuristic and (v) centering features. The features of types (i) to (iv) are defined so as to simulate Ng and Cardie's feature set, except the following three features:

- LOG_LIKE: indicates the largest value among the log-likelihood coefficients (Dunning, 1993) of the pairs of a noun in the coreference chain including the candidate and the predicate of the anaphor. Those coefficients are calculated with about ten millions of

[2]The GDA (Global Document Annotation (Hasida, 2002)) tag set is designed to be a standard tag set which allows machines to automatically recognize the semantic and pragmatic structures of documents.

Feature types	Feature names	Descriptions
Grammatical	POS	The part-of-speech of NP_i such as 'proper noun' and 'sahen noun'.
	DEFINITE	Y if NP_i is 'sore', 'soko', 'sono', 'sonna', etc; else N.
	DEMONSTRATIVE	Y if NP_i is 'kore', 'soko', 'ano', 'asoko', etc; else N.
	PARTICLE	The case marker attached to NP_i such as 'wa', 'ga' and 'o'.
Semantic	NE	Named entity class of NP_i: PERSON, ORGANIZATION, LOCATION, ARTIFACT, DATE, TIME, MONEY, PERCENT or N/A.
	EDR_HUMAN	Y if NP_i has the human attribute of EDR dictionary; else N.
	SELECT_REST	C if NP_i-ANP pair satisfies the selectional restriction; else I.
	LOG_LIKE	Five degree of the log-like coefficient of the NP_i-ANP pair.
	ANIMACY	Y if NP_i has the PERSON or ORGANIZATION class; else N.
	ANIMACY_COMP*	NP_1 if NP_1 has ANIMACY feature and NP_2 doesn't; else NP_2 if the opposite relation.
Positional	SENTNUM_ANP	Distance between NP_i and ANP in terms of sentences.
	SENTNUM_NPS*	Distance between NP_1 and NP_2 in terms of sentences.
	DEP_MAIN	Y if NP_i depends on the main clause; else N.
	EMBEDDED	Y if NP_i locates in an embedded clause; else N.
	BEGINNING	Y if NP_i locates in the beginning of the sentence; else N.
Heuristic	CHAIN_LENGTH	Length of a cohesive chain of NP_i
Centering	SRL_ORDER	The priority rank of NP_i in SRL.
	SRL_ORDER_COMP*	NP_1 if NP_1 is higher ranked than NP_2 in SRL; else NP_2
	GA_REF	Y if NP_i is the subject of a subordinate clause of a particuler conjunctive type and ANP is the subject of its matrix clause; else N.

Table 1: Feature Set

ANP is an anaphor, and $NP_{i\in\{1,2\}}$ is an antecedent candidate. The feature set contains relational and non-relational features. Non-relational features test some property P of NP_i under consideration and take on a value of YES or NO depending on whether P holds. Relational features test whether some property P holds for the NP_1-NP_2 or NP_2-ANP pair under consideration and indicates whether the pair is COMPATIBLE or INCOMPATIBLE w.r.t. P; a value of NOT APPLICABLE is used when property P does not apply. Features with an asterisk are used only in the tournament model.

NOUN-VERB pairs extracted from other corpora (Nikkei Shimbun, 1990-2000; Mainichi Shimbun, 1991-1999).

- SELECT_REST: indicates whether or not a candidate satisfies selectional restrictions in Nihongo Goi Taikei (Japanese Lexicon) (Ikehara, et al., 1997).
- CHAIN_LENGTH: indicates the number of all the preceding nouns in the coreference chain including the candidate.

We also introduce ANIMACY feature as in Ng's feature set, because an animate noun tends to be salient. ANIMACY indicates whether or not the candidate is an animate noun. A noun is regarded as animate if the noun is classified as PERSON or ORGANIZATION by a named entity tagger or the noun is included in PERSON or ORGANIZATION class of Nihongo Goi Taikei (Ikehara, et al., 1997).

To define centering features, we adopted a Japanese anaphora resolution model proposed by Nariyama (2002) as the underlying theory. Nariyama's method is an expansion of Kameyama's work on the application of Centering Theory to Japanese anaphora (Kameyama, 1986). Nariyama expanded the original forward-looking center list into Salience Reference List (SRL) in order to take into account broader contextual information from preceding sentences. Analogous to common centering models, in SRL, discourse entities are stored in the salience order: TOPIC (marked by *wa*-particle) > SUBJ (*ga*) > I_OBJ (*ni*) > D_OBJ (*o*) > OTHERS. In the experiment, we introduced two features, SRL_ORDER and SRL_ORDER_COMP, to reflect the SRL-related contextual factors. The definition of them is given in Table 1. Nariyama's method is also devised to deal with state transitions in complex sentences, which was originally not handled in Kameyama's model on Japanese. We partially implemented this extension as another feature, GA_REF, expecting the strong tendency of coreference that some conjunctives convey.

In the experiment, all the features are automatically computed with the help of the following NLP systems: the Japanese morphological analyzer

ChaSen (Matsumoto, 2000), the Japanese dependency structure analyzer *CaboCha* (Kudoh and Matsumoto, 2000), and the named entity chunker *Yanee* (Yamada, 2002).

4.3 Results

While Ng and Cardie used the C4.5 decision tree induction system, we adopted Support Vector Machines (Vapnik, 1998) for classifier induction because of their state-of-the-art performance and considerable generalization ability, which had been proven for various NLP tasks.

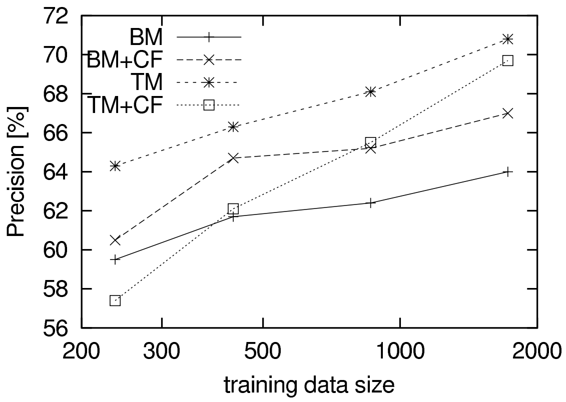

Figure 3: Learning curves
BM: Baseline model, BM+CF: Baseline model using centering features TM: Tournament model, TM+CF: Tournament model using centering features

The results are shown in Figure 3. We can see the positive effects for introducing the centering features by comparing the learning curves of BM+CF with BM, and TM+CF with TM. Likewise, the differences between BM and TM show that the introduction of the tournament model significantly improved the performance regardless of the size of training data. That is to say, when the tournament and centering features (TM+CF) are incorporate, precision is always higher than without, except for TM+CF in small data size. However, the improvement ratio of this model against the data size is, in fact, the best of all, which suggest that it will become the best method as the data size increases.

One can also introduce the notion of decision confidence into the tournament model. With a good confidence measure, one can effectively improve precision just by slightly sacrificing recall.

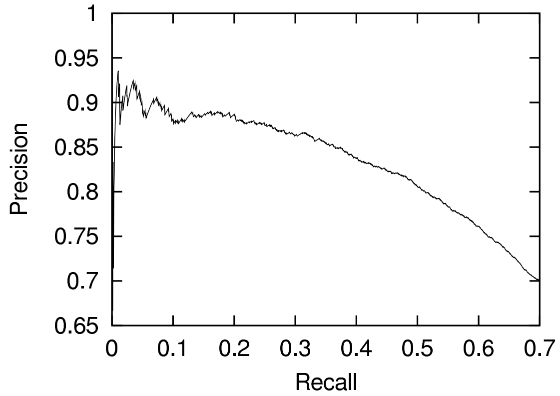

Figure 4: Precision-recall curve obtained with the tournament model

In case of the tournament model, the likelihood (i.e. the degree of confidence) that the decision for a match is correct can be heuristically estimated by, for example, the absolute value of the SVM classifier's discrimination function for the corresponding classification problem. The likelihood that the winner of a tournament is correct is then given by the confidence value of the closest match the winner have played. Given such a confidence measure, one can obtain a recall-precision curve by moving the threshold of confidence values. Working of this is shown in Figure 4, which presents the recall-precision curve obtained by testing this heuristic measure.

5 Related work

There have been an increasing number of reports on corpus-based empirical approaches to coreference resolution. For example, besides the models proposed by Soon et al. (2001) and Ng and Cardie (2002), which we have referred to as the baseline model, one can find a diversity of trainable models for the resolution of pronouns (Ge and Charniak, 1998), definite NPs (Aone and Bennett, 1995; McCarthy and Lehnert, 1995; Strube et al., 2002), and Japanese zero-pronouns (Yamamoto and Sumita, 1998; Seki et al., 2002).

Surprisingly enough, however, very few of these models has an explicit reference to such theoretical work as Centering Theory. In fact, none of them incorporates a training feature that captures local context as well as the centering features we proposed in this paper. Furthermore, the previous models are all designed to estimate for each candidate how "likely" it is the coreferent without referring to others. . The deficiency of such a candidate-wise estimation model is just as we discussed in Section 3.1. The experimental results reported above confirmed that our tournament model has the potential for overcoming it.

6 Summary and future work

In this paper, we presented a trainable coreference resolution model that is designed to incorporate contextual cues by means of centering features and a tournament-based search algorithm. These two improvements worked effectively in our experiments on Japanese zero-anaphora resolution.

As future work, we address three remaining issues: (i) identification of relations between the topics and the subtopics, (ii) analysis of complex and quoted sentences and (iii) refinement of selectional restrictions. With regard to (i), *wa*-marked subtopics are often incorrectly selected because the present model cannot capture topic-subtopic structures. Our next step will be to encode such hierarchical structure as a centering feature. Since a topic-subtopic relation holds between two NPs, it may be effective in the tournament model. As for (ii), a half of zero-anaphors in GDA have the closest antecedent in the same sentence because the great majority of sentences in newspaper articles are complex sentences. Thus, correct dependency analysis of complex sentences is necessary. However, the dependency analyzer we used often make errors with complex sentences. We hope for a parallel progress of dependency structure analysis and antecedent identification. Finally, we need to make selectional restrictions more effective for resolving anaphora. We examined the misclassified examples. Most errors are related to quoted sentences, which are not dealt with under the framework of the Centering. The corpus include many tagging errors. Enhancing the quality of corpus as well as the robustness of learning framework will be also our next step.

Acknowledgements

We thank Kôiti Hasida (AIST) for allowing us to use the GDA-tagged corpus, Hiroyasu Yamada (JAIST) for his NE tagging tool YANEE, and Taku Kudo (NAIST) for his Japanese dependency structure analyzer CaboCha and TinySVM. We also thank Shigeko Nariyama (Melbourne University / NAIST) for helpful comments on this work.

References

C. Aone and S. W. Bennett. 1995. Evaluating automated and manual acquisition of anaphora resolution strategies. In *Proceedings of 33th Annual Meeting of the Association for Computational Linguistics(ACL)*, 122-129.

B. Baldwin. 1995. CogNIAC: a discourse processing engine. *Ph.D. Thesis, Department of Computer and Information Sciences, University of Pennsylvania.*

T. Dunning. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61-74.

N. Ge, J. Hale, and E. Charniak. 1998. A statistical approach to anaphora resolution. In *Proceedings of the Sixth Workshop on Very Large Corpora*, 161-170.

B. J. Grosz, A. K. Joshi, and S. Weinstein. 1995. Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203-226.

M. Halliday and R. Hasan. 1976. *Cohesion in English.* Longman.

K. Hasida. 2002. Global Document Annotation (GDA). http://i-content.org/

K. Yamamoto and E. Sumita. 1998. Feasibility Study for Ellipsis Resolution in Dialogues by Machine-Learning Technique. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*

S. Ikehara, M. Miyazaki, A. Yokoo, S. Shirai, H. Nakaiwa, K. Ogura, Y. Ooyama, and Y. Hayashi. 1997. *Nihongo Goi Taikei - A Japanese Lexicon.* Iwanami Shoten. (In Japanese).

M. Kameyama. 1986. A property-sharing constraint in centering. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, 200-206.

M. Kameyama. 1997. Intrasentential centering: a case study. In *Centering Theory in Discourse*, chap. 6, Oxford University Press.

T. Kudoh and Y. Matsumoto. 2000. Japanese dependency analysis based on Support Vector Machines. In *Proceedings of EMNLP/VLC2000.*

Mainichi Shimbunsha. 1991-1999. *Mainichi Shimbun CD-ROM*.

Y. Matsumoto, A. Kitauchi, T. Yamashita, Y. Hirano, H. Matsuda, K. Takaoka, and M. Asahara. 2000. *Morphological Analysis System ChaSen version 2.2.1 Manual*.

J. F. McCarthy and W. G. Lehnert. 1995. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*.

R. Mitkov. 1997. Factors in anaphora resolution: they are not the only things that matter. A case study based on two different approaches. In *Proceedings of the ACL'97/EACL'97 Workshop on Operational Factors in Practical, Robust Anaphora Resolution*.

C. Müller, S. Rapp, and M. Strube. 2002. Applying Co-Training to reference resolution. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (ACL 2002)*.

M. Murata, H. Isahara, and M. Nagao. 1999. Resolution of indirect anaphora in Japanese sentences using examples "X no Y (Y of X)". In *Proceedings of the ACL'99 Workshop on Coreference and Its Applications.*

H. Nakaiwa and S. Shirai. 1996. Anaphora resolution of Japanese zero pronouns with deictic reference. In *Proceedings of the 16th International Conference on Computational Linguistics*, 812-817.

S. Nariyama 2002. Grammar for ellipsis resolution in Japanese. In *Proceedings of the 9th International Conference on Theoretical and Methodological Issues in Machine Translation*, 135-145.

V. Ng and C. Cardie. 2002. Improving machine learning approaches to coreference resolution. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics(ACL)*, 104-111.

Nikkei Shimbunsha. 1990-2000. *Nikkei Shimbun CD-ROM.*

M. Okumura and K. Tamura. 1996. Zero pronoun resolution in Japanese discourse based on centering theory. In *Proceedings of the 16th Annual Meeting of the Association for Computational Linguistics*, 871-876.

J. R. Quinlan. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA.

K. Seki, A. Fujii, and T. Ishikawa. 2002. A probabilistic method for analyzing Japanese anaphora integrating zero pronoun detection and resolution, In *Proceedings of the 19th International Conference on Computational Linguistics*, 911-917.

W. M. Soon, H. T. Ng, and D. C. Y. Lim. 2001. A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27(4):521-544.

M. Strube, S. Rapp, and C. Müller. 2002. The influence if minimum edit distance on reference resolution. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*.

V. Vapnik. 1998. *Statistical Learning Theory*. John Wiley.

M. Walker, M. Iida, and S. Cote. 1994. Japanese discourse and the process of centering. *Computational Linguistics*, 20(2):193-233.

H. Yamada, T. Kudo, and Y. Matsumoto. 2002. Japanese named entity extraction using Support Vector Machine. *IPSJ Journal*, Vol.43, No.1, (in Japanese).

T. Yokoi. 1995. The EDR electric dictionary. *Communications of the ACM*, Vol.38, No.11, November 1995, 42-44.