Zero-Anaphora Resolution by Learning Rich Syntactic Pattern Features

RYU IIDA, KENTARO INUI, AND YUJI MATSUMOTO
Nara Institute of Science and Technology

We approach the zero-anaphora resolution problem by decomposing it into intrasentential and intersentential zero-anaphora resolution tasks. For the former task, syntactic patterns of zero-pronouns and their antecedents are useful clues. Taking Japanese as a target language, we empirically demonstrate that incorporating rich syntactic pattern features in a state-of-the-art learning-based anaphora resolution model dramatically improves the accuracy of intrasentential zero-anaphora, which consequently improves the overall performance of zero-anaphora resolution.

Categories and Subject Descriptors: I.2.7 [**Artificial Intelligence**]: Natural Language Processing — *Discourse*

General Terms: Algorithms, Design, Experimentation

Additional Key Words and Phrases: Anaphora resolution, zero-pronouns

ACM Reference Format:
Iida, R., Inui, K., and Matsumoto, Y. 2007. Zero-anaphora resolution by learning rich syntactic pattern features. ACM Trans. Asian Lang. Inf. Process. 6, 4, Article 12 (December 2007), 22 pages. DOI = 10.1145/1316457.1316458. http://doi.acm.org/10.1145/1316457.1316458.

1. INTRODUCTION

Zero-anaphora is a gap in a sentence that has an anaphoric function similar to a pro-form (e.g., pronoun) and is often described as "referring back" to an expression that supplies the information necessary for interpreting the sentence. For example, in the sentence "*There are two roads to eternity, a straight and narrow, and a broad and crooked*," the gaps in "*a straight and narrow (gap)*" and "*a broad and crooked (gap)*" have a zero-anaphoric relationship to "*two roads to eternity*."

Authors' addresses: R. Iida, Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara, 630-0192, Japan; email: ryu-i@is.naist.jp; K. Inui, Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara, 630-0192, Japan; email: inui@is.naist.jp; Y. Matsumoto, Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara, 630-0192, Japan; email: matsu@is.naist.jp.

The task of identifying zero-anaphoric relations in a given discourse, *zero-anaphora resolution*, is essential in a wide range of natural language processing (NLP) applications. This is the case particularly in a language such as Japanese, where even obligatory arguments of a predicate are often omitted when they are inferable from the context. In fact, in our Japanese newspaper corpus [Iida et al. 2007], for example, 45.5% of the nominative arguments of verbs are omitted. Since such gaps cannot be interpreted only by shallow syntactic parsing, a model specialized for zero-anaphora resolution needs to be built on top of both syntactic and semantic models.

Recent work on zero-anaphora resolution can be placed in two different research contexts. First, zero-anaphora resolution is studied in the context of anaphora resolution (AR), in which zero-anaphora is regarded as a subclass of anaphora. In AR, the research trend has been shifting from rule-based approaches [Baldwin 1995; Lappin and Leass 1994; Mitkov 1997] to empirical, or corpus-based, approaches [McCarthy and Lehnert 1995; Ge et al. 1998; Soon et al. 2001; Ng and Cardie 2002; Strube and Müller 2003; Yang et al. 2003; Ng 2004; Yang et al. 2005] because the latter are shown to be a cost-efficient solution achieving a performance that is comparable to best-performing rule-based systems (see the Coreference task in Message Understanding Conference (MUC)[1] and the Entity Detection and Tracking task in the Automatic Content Extraction (ACE) program[2]). The same trend is observed also in Japanese zero-anaphora resolution, where the findings made in rule-based or theory-oriented work [Kameyama 1986; Walker et al. 1994; Nakaiwa and Shirai 1996; Okumura and Tamura 1996; Murata and Nagao 1997] have been successfully incorporated in machine-learning-based frameworks [Seki et al. 2002; Iida et al. 2003; Isozaki and Hirao 2003].

Second, the task of zero-anaphora resolution has some overlap with Propbank[2]-style semantic role labeling (SRL), which has been intensively studied, for example, in the context of the CoNLL SRL task.[3] In this task, given a sentence "To attract younger listeners, Radio Free Europe intersperses the latest in Western rock groups," an SRL model is asked to identify the Noun Phrase (NP) *Radio Free Europe* as the A0 (Agent) argument of the verb *attract*. This can be seen as the task of finding the zero-anaphoric relationship between a nominal gap (the A0 argument of *attract*) and its antecedent (*Radio Free Europe*) under the condition that the gap and its antecedent appear in the same sentence.

In spite of this overlap between AR and SRL, there are some important findings that have yet to be exchanged between them, partly because the two fields have been evolving somewhat independently. The AR community has recently made two important findings:

—A model that identifies the antecedent of an anaphor by a series of comparisons between candidate antecedents has a remarkable advantage over

[1] http://www-nlpir.nist.gov/related_projects/muc/
[2] http://projects.ldc.upenn.edu/ace/
[3] http://www.lsi.upc.edu/~srlconll/

a model that estimates the absolute likelihood of each candidate independently of other candidates [Iida et al. 2003; Yang et al. 2003].

—An AR model that carries out antecedent identification *before* anaphoricity determination, the decision whether a given NP is anaphoric or not (i.e., discourse-new), significantly outperforms a model that executes those subtasks in the reverse order or simultaneously [Poesio et al. 2004; Iida et al. 2005].

To our best knowledge, however, existing SRL models do not exploit these advantages. In SRL, on the other hand, it is common to use syntactic features derived from the parse tree of a given input sentence for argument identification. A typical syntactic feature is the path on a parse tree from a target predicate to a noun phrase in question [Gildea and Jurafsky 2002; Carreras and Marquez 2005]. However, existing AR models deal with intra- and intersentential anaphoric relations in a uniform manner; that is, they do not use as rich syntactic features as state-of-the-art SRL models do, even in finding intrasentential anaphoric relations. We believe that the AR and SRL communities can learn more from each other.

Given this background, in this article, we show that combining the aforementioned techniques derived from each research trend makes significant impact on zero-anaphora resolution, taking Japanese as a target language. More specifically, we demonstrate the following:

—Incorporating rich syntactic features in a state-of-the-art AR model dramatically improves the accuracy of intrasentential zero-anaphora resolution, which consequently improves the overall performance of zero-anaphora resolution. This is to be considered as a contribution to AR research.

—Analogously to intersentential anaphora, decomposing the antecedent identification task into a series of comparisons between candidate antecedents works remarkably well also in intrasentential zero-anaphora resolution. We hope this finding will be adopted in SRL.

The rest of the article is organized as follows. Section 2 describes the task definition of zero-anaphora resolution in Japanese. In Section 3, we review previous approaches to AR. Section 4 described how the proposed model incorporates effectively syntactic features into the machine-learning-based approach. We then report the results of our experiments on Japanese zero-anaphora resolution in Section 5 and conclude in Section 6.

2. ZERO-ANAPHORA RESOLUTION

In this article, we consider only zero-pronouns that function as an obligatory argument of a predicate for two reasons:

—Providing a clear definition of zero-pronouns appearing in adjunctive argument positions involves awkward problems, which we believe should be postponed until obligatory zero-anaphora is well studied.

—Resolving obligatory zero-anaphora tends to be more important than adjunctive zero-pronouns in actual applications.

A zero-pronoun may have its antecedent in the discourse; in this case, we say the zero-pronoun is *anaphoric*. On the other hand, a zero-pronoun whose referent does not explicitly appear in the discourse is called a *nonanaphoric* zero-pronoun. A zero-pronoun may be nonanaphoric typically when it refers to an extralinguistic entity (e.g., the first or second person) or its referent is unspecified in the context.

The following are Japanese examples. In sentence (1), zero-pronoun ϕ_i is anaphoric as its antecedent, "*shusho* (prime minister)," appears in the same sentence. In sentence (2), on the other hand, ϕ_j is considered nonanaphoric if its referent (i.e., the first person) does not appear in the discourse.

(1)

shusho$_i$*-wa*	*houbeisi-te*	*,*	*ryoukoku-no*	*gaikou-o*
prime minister$_i$-TOP	visit-U.S.-CONJ	PUNC	both countries-BETWEEN	diplomacy-OBJ

(ϕ_i*-ga*)	*suishinsuru*	*houshin-o*	*akirakanisi-ta*	.
(ϕ_i-NOM)	promote-ADNOM	plan-OBJ	unveil-PAST	PUNC

The prime minister visited the United States and unveiled the plan to push diplomacy between the two countries.

(2)

(ϕ_j*-ga*)	*ie-ni*	*kaeri-tai*	.
(ϕ_j-NOM)	home-DAT	want to go back	PUNC

(I) want to go home.

Given this distinction, we consider the task of zero-anaphora resolution as the combination of two subproblems, antecedent identification and anaphoricity determination, which is analogous to NP-anaphora resolution:

> For each zero-pronoun in a given discourse, find its antecedent if it is anaphoric; otherwise, conclude it to be nonanaphoric.

3. PREVIOUS WORK

Anaphora resolution can be decomposed into two subtasks; antecedent identification and anaphoricity determination. In this section, we briefly show the previous work of each task respectively. In addition, we summarize related work that incorporates syntactic information into anaphora resolution.

3.1 Antecedent Identification

Previous machine-learning-based approaches to antecedent identification can be classified as either the *candidate-wise classification* approach or the *preference-based* approach. In the former approach [Soon et al. 2001; Ng and Cardie 2002], given a target anaphor, *TA*, the model estimates the absolute likelihood of each of the candidate antecedents (i.e., the NPs preceding *TA*), and selects the best-scored candidate. If all the candidates are classified negative, *TA* is judged nonanaphoric.

In contrast, the preference-based approach [Yang et al. 2003; Iida et al. 2003] decomposes the task into comparisons of the preference between candidates and selects the most preferred one as the antecedent. For example, Iida et al. [2003] proposed a method called the *tournament model*. This model conducts a tournament consisting of a series of matches in which candidate antecedents compete with each other for a given anaphor.

While the candidate-wise classification model computes the score of each single candidate independently of others, the tournament model learns the relative preference between candidates, which is empirically proved to be a significant advantage over candidate-wise classification [Iida et al. 2003].

3.2 Anaphoricity Determination

There are two alternative ways for anaphoricity determination: the *single-step model* and the *two-step model*. The single-step model [Soon et al. 2001; Ng and Cardie 2002] determines the anaphoricity of a given anaphor indirectly as a by-product of the search for its antecedent. If an appropriate candidate antecedent is found, the anaphor is classified as anaphoric; otherwise, it is classified as nonanaphoric. One disadvantage of this model is that it cannot employ the preference-based model because the preference-based model is not capable of identifying nonanaphoric cases.

The two-step model [Ng 2004; Poesio et al. 2004; Iida et al. 2005], on the other hand, carries out anaphoricity determination in a separate step from antecedent identification. Poesio et al. [2004] and Iida et al. [2005] claimed that the latter subtask should be done before the former. For example, given a target anaphor (*TA*), Iida et al.'s *selection-then-classification* (SCM) model:

(1) selects the most likely candidate antecedent (*CA*) of *TA* using the tournament model,
(2) classifies *TA* paired with *CA* as either *anaphoric* or *nonanaphoric* using an anaphoricity determination model. If the *CA*-*TA* pair is classified as *anaphoric*, *CA* is identified as the antecedent of *TA*; otherwise, *TA* is concluded to be *nonanaphoric*.

The anaphoricity determination model learns the nonanaphoric class directly from nonanaphoric training instances whereas the single-step model cannot use nonanaphoric cases in training.

4. PROPOSAL

In this section, we decompose the zero-anaphora resolution the problem into two subtasks, *intrasentential* and *intersentential* zero-anaphora resolution and then we propose a method that incorporates syntactic patterns as features into an intrasentential problem.

4.1 Task Decomposition

We approach the zero-anaphora resolution problem by decomposing it into two subtasks: intrasentential and intersentential zero-anaphora resolution. For the former problem, syntactic patterns in which zero-pronouns and their antecedents appear may well be useful clues, which, however, does not apply to the latter problem. We therefore build a separate component for each subtask, adopting Iida et al. [2005]'s selection-then-classification model for each component:

(1) *Intrasentential antecedent identification*. For a given zero-pronoun ZP in a given sentence S, select the most-likely candidate antecedent C_1^* from the candidates appearing in S by the intrasentential tournament model. The tournament model conducts a tournament consisting of a series of matches in which candidate antecedents compete with each other for a given anaphor. In the tournament, it processes the candidate antecedents in the right-to-left order. In the first round, the model consults a trained classifier to judge which of the rightmost two candidates is more likely to be the antecedent for the anaphor. The winner then plays a match with the third rightmost candidate. Likewise, each of the following matches is arranged in turn between the current winner and a rightmost new challenger until the winner of the final round is determined. The model selects the winner as the most likely candidate antecedent.
(2) *Intrasentential anaphoricity determination*. Estimate plausibility p_1 that C_1^* is the true antecedent, and return C_1^* if $p_1 \geq \theta_{\text{intra}}$ (θ_{intra} is a preselected threshold) or go to 3 otherwise.
(3) *Intersentential antecedent identification*. Select the most-likely candidate antecedent C_2^* from the candidates appearing outside of S by the intersentential tournament model.
(4) *Intersentential anaphoricity determination*. Estimate plausibility p_2 that C_2^* is the true antecedent, and return C_2^* if $p_2 \geq \theta_{\text{inter}}$ (θ_{inter} is a preselected threshold) or return nonanaphoric otherwise.

4.2 Representation of Syntactic Patterns

In the first two of the above four steps, we use syntactic pattern features. Analogously to SRL, we extract the parse path between a zero-pronoun to its antecedent to capture the syntactic pattern of their occurrence. Among many alternative ways of representing a path, in the experiments reported in the next section, we adopted a method as we will describe, leaving the exploration of other alternatives as future work.

Given a sentence, we first use a standard dependency parser, *CaboCha* [Kudo and Matsumoto 2002], to obtain the dependency parse tree, in which words are structured according to the dependency relation defined in the Kyoto Corpus[4] [Kurohashi and Nagao 1997]. Figure 1(a), for example, shows the dependency tree of sentence (1) in Section 2. In the figure, a content word node depends on the other content word node, sometimes involving functional word nodes. The node labeled "adnom" denotes the adnominal relation between *suishinsuru* (promote) and *houshin* (plan) and the "ϕ" node denotes a zero-pronoun. We then extract the path between a zero-pronoun and its antecedent as in Figure 1(b). Finally, to encode the order of siblings and reduce data sparseness, we further transform the extracted path as in Figure 1(c):

[4]http://nlp.kuee.kyoto-u.ac.jp/nl-resource/corpus.html

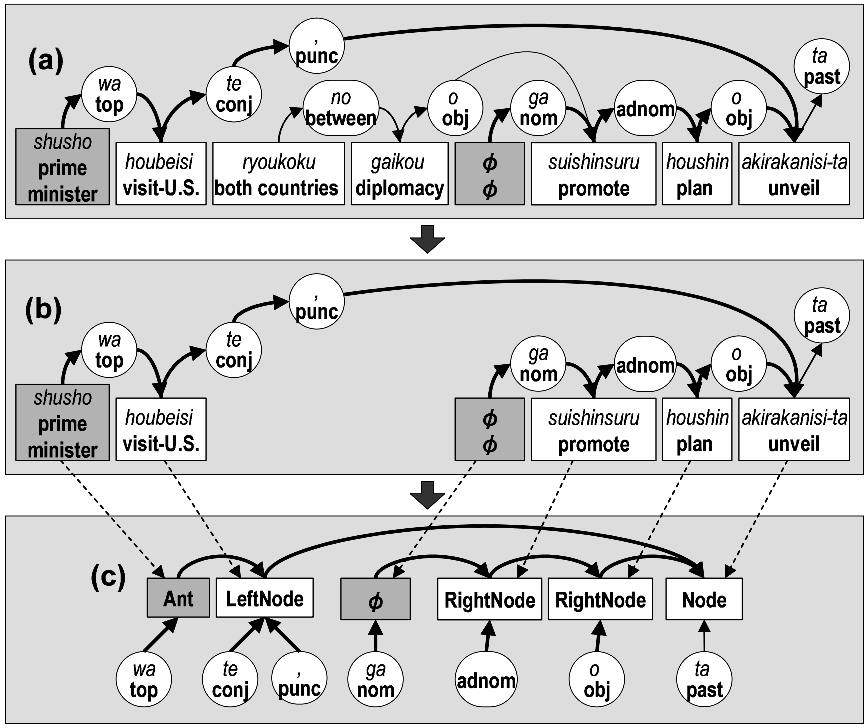

Fig. 1. Representation of the path between a zero-pronoun to its antecedent.

—A path is represented by a subtree consisting of backbone nodes: ϕ (zero-pronoun), Ant (antecedent), Node (the lowest common ancestor), LeftNode (node corresponding to a left-branch content word), and RightNode.

—Each backbone node has daughter nodes, each corresponding to a function word associated with it.

—Content words are deleted.

This way of encoding syntactic patterns is used in intrasentential anaphoricity determination. Our intrasentential anaphoricity determination model is trained by providing a set of labeled trees as a training set, where a label is either *anaphoric* or *nonanaphoric*. Each labeled tree consists of the path tree shown in Figure 1(c) and the node corresponding to the binary features in Table I, each of which is linked to the root node as illustrated in Figure 2. This way of organizing a labeled tree allows the model to learn, for example, the combination of a subtree of T_C and some of the binary features.

Analogously, in antecedent identification, the tournament model allows us to incorporate three paths, a path for each pair of a zero-pronoun and the left and right candidate antecedents. The model is trained by providing a set of labeled trees as a training set. A label is either *left* or *right*, indicating which candidate is more likely to be an antecedent than the other competing candidate. Each labeled tree has (1) path trees T_L, T_R, and T_I (as given in Figure 3)

Table I. Feature Set

Feature Type	Feature	Description
Lexical	HEAD_BF	characters of rightmost morpheme in *NP* (*PRED*).
Grammatical	PRED_IN_MATRIX	1 if *PRED* exists in the matrix clause; otherwise 0.
	PRED_IN_EMBEDDED	1 if *PRED* exists in the relative clause; otherwise 0.
	PRED_VOICE	1 if *PRED* contains passive auxiliaries such as "*(ra)reru*"; otherwise 0.
	PRED_AUX	1 if *PRED* contains auxiliaries such as "*(sa)seru*," "*hosii*," "*morau*," "*itadaku*," "*kudasaru*," "*yaru*," and "*ageru*."
	PRED_ALT	1 if PRED_VOICE is 1 or PRED_AUX is 1; otherwise 0.
	POS	Part-of-speech of *NP* followed by IPADIC [Asahara and Matsumoto 2003].
	DEFINITE	1 if *NP* contains the article corresponding to DEFINITE "the," such as "*sore*" or "*sono*"; otherwise 0.
	DEMONSTRATIVE	1 if *NP* contains the article corresponding to DEMONSTRATIVE "that" or "this," such as "*kono*," "*ano*"; otherwise 0.
	CASE_MARKER	Case marker followed by *NP*, such as "*wa* (topic)," "*ga* (subject)," "*o* (object)."
Semantic	NE	Named entity of *NP*: PERSON, ORGANIZATION, LOCATION, ARTIFACT, DATE, TIME, MONEY, PERCENT, or N/A.
	EDR_HUMAN	1 if *NP* is included among the concept "a human being" or "attribute of a human being" in EDR dictionary [Japan Electronic Dictionary Research Institute, Ltd. Japan 1995]; otherwise 0.
	EDR_ORG	1 if *NP* is included among the concept "organization" in EDR dictionary [Japan Electronic Dictionary Research Institute, Ltd. Japan 1995]; otherwise 0.
	PRONOUN_TYPE	Pronoun type of *NP*. (e.g. "*kare* (he)" → PERSON, "*koko* (here)" → LOCATION, "*sore* (this)" → OTHERS)
	SELECT_REST	1 if *NP* satisfies selectional restrictions in Nihongo Goi Taikei (Japanese Lexicon) [Ikehara et al. 1997]; otherwise 0.
	COOC	The score of well-formedness model estimated from a large number of triplets ⟨*Noun*, *Case*, *Predicate*⟩ proposed by Fujita et al. [2004]
Positional	SENTNUM	Distance between *NP* and *PRED*.
	BEGINNING	1 if *NP* is located in the beginning of sentence; otherwise 0.
	END	1 if *NP* is located in the end of sentence; otherwise 0.
	PRED_NP	1 if *PRED* precedes *NP*; otherwise 0.
	NP_PRED	1 if *NP* precedes *PRED*; otherwise 0.
	DEP_PRED	1 if NP_i depends on *PRED*; otherwise 0.
	DEP_NP	1 if *PRED* depends on NP_i; otherwise 0.
	IN_QUOTE	1 if *NP* exists in the quoted text; otherwise 0.
Heuristic	CL_RANK	A rank of *NP* in forward-looking-center list based on Centering Theory [Grosz et al. 1995]
	CL_ORDER	A order of *NP* in forward-looking-center list based on Centering Theory [Grosz et al. 1995]

NP stands for a bunsetsu-chunk of a candidate antecedent. *PRED* stands for a bunsetsu-chunk of a predicate which has a target zero-pronoun.

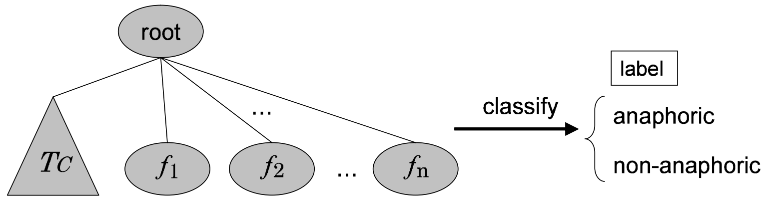

Fig. 2. Tree representation of features in intrasentential anaphoricity determination.

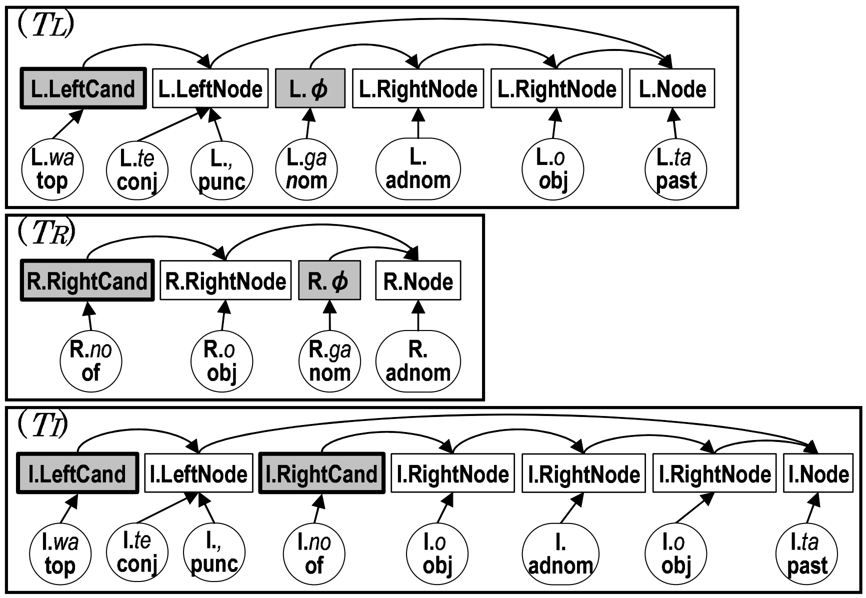

Fig. 3. Paths used in the tournament model.

and (2) a set of nodes corresponding to the binary features summarized in Table I, each of which is linked to the root node as illustrated in Figure 4.

Here, assume the situation considering two candidates, *shusho* (prime minister) and *ryoukoku* (both countries), for the zero-pronoun ϕ in Figure 1(a). For example, in order to represent the relation between the left candidate and the zero-pronoun, a path linking *shusho* to ϕ is extracted and then it is transformed into a syntactic pattern as in T_L in Figure 3. The patterns between the right candidate and the zero-pronoun and ones between two candidates are also created by the same way (T_R and T_I in Figure 3). Note that the label of each node is prefixed either with L, R or I to indicate which node belongs to which subtree.

Alternatively, learning structural information from phrase structures also effectively works for this kind of problem and has been paid attention by several research groups [Hobbs 1978; Luo and Zitouni 2005; Yang et al. 2006].

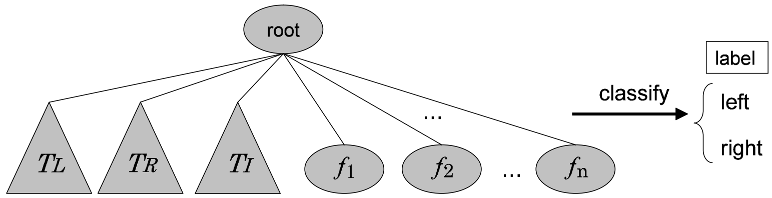

Fig. 4. Tree representation of features in antecedent identification.

If we take English as a target language, it is reasonable to deal with syntactic trees. In contrast, languages like Japanese involve grammatical properties such as word scrambling (e.g., objects followed by subjects), posing an obstacle to Japanese phrase structure analysis, so it is still difficult to construct phrase structure parsers. We, therefore, use a dependency tree in order to represent the structure of a sentence.

4.3 Learning Algorithm

As noted in Section 1, the use of zero-pronouns in Japanese is relatively less constrained by syntax compared, for example, with English. This forces the above way of encoding path information to produce an explosive number of different paths, which inevitably leads to serious data sparseness.

This issue can be addressed in several ways. The SRL community has devised a range of variants of the standard path representation to reduce the complexity [Carreras and Marquez 2005]. Applying kernel methods such as tree kernels [Collins and Duffy 2001] and hierarchical DAG kernels [Suzuki et al. 2003] is another strong option. The boosting-based algorithm proposed by Kudo and Matsumoto [2004] is designed to learn subtrees useful for classification.

Leaving the question of selecting learning algorithms open, in our experiments, we have so far examined Kudo and Matsumoto [2004]'s algorithm, which is implemented as the BACT system.[5] Given a set of training instances, each of which is represented as a tree labeled either positive or negative, the BACT system learns a list of weighted decision stumps with a boosting algorithm. Each decision stamp classifier is represented as a labeled ordered tree appearing in the training instances.

The tree classification problem in BACT is defined to induce a mapping function $f(\mathbf{x})$: $\mathcal{X} \rightarrow \{\pm 1\}$, from given training instances $T = \{\langle \mathbf{x}_i, y_i \rangle\}_{i=1}^{L}$, where $\mathbf{x}_i \in \mathcal{X}$ is a labeled ordered tree and $y_i \in \{\pm 1\}$ is a class label associated with each training data item. In each iteration of boosting, the decision stumps are trained to find a rule $\langle \hat{t}, \hat{y} \rangle$ that minimizes the error rate for the given training data $\{\langle \mathbf{x}_i, y_i \rangle\}_{i=1}^{L}$:

$$\langle \hat{t}, \hat{y} \rangle = \operatorname*{argmax}_{t \in \mathcal{F}, y \in \{\pm 1\}} \sum_{i=1}^{L} y_i d_i h_{\langle t, y \rangle}(\mathbf{x}_i), \tag{1}$$

[5]http://chasen.org/~taku/software/bact/

where $\mathcal{F}$ is a set of candidate trees, $d_i(\sum_{i=1}^{L} d_i = 1, d_i \geq 0, \forall i = 1, \ldots, L)$ is a weight of each instance, and $h_{\langle t,y \rangle}(\mathbf{x})$ is a decision stump classifier given by

$$h_{\langle t,y \rangle}(\mathbf{x}) \stackrel{\text{def}}{=} \begin{cases} y & t \subseteq \mathbf{x} \\ -y & \text{otherwise.} \end{cases} \quad (2)$$

At the classification step, we use the following mapping function:

$$f(\mathbf{x}) = sgn(\sum_{k=1}^{K} \alpha_k h_{\langle t,y \rangle}(\mathbf{x}_k)) \quad (3)$$

where α_k is a weight of each decision stumps classifier $h_{\langle t,y \rangle}(\mathbf{x}_k)$. In this algorithm, α_k is calculated based on a variant of boosting algorithm *Arc-GV* (see Breiman [1999]).

In our anaphoricity determination problem, given a set of positive (anaphoric) training trees and a set of negative (nonanaphoric) training trees, BACT induces a set of subtrees (decision stumps) that are useful for our binary classification. Each subtree is associated with weight $w \in [-1, 1]$ as shown in Table II and Table III. A subtree with a positive weight corresponds to a decision stump that votes for *anaphoric*. In the test phase, given an input tree involving a zero-pronoun and the most-likely candidate antecedent, BACT computes the score by summing up the weights of all the decision stumps that match with the input tree and concludes *anaphoric* if the score is positive, or *nonanaphoric* otherwise. Each choice between left or right in the tournament model for antecedent identification is carried out analogously.

The BACT algorithm has the important characteristic that the results of learning trees are more human readable than those learned from algorithms such as Support Vector Machines, because the result of each iteration is given as a pair of decision stumps $h_{\langle t,y \rangle}$ and weight α_k as shown in Table II and Table III. So, we can easily interpret what kinds of subtrees or features are useful for classification by viewing the weights of induced decision stumps.

5. EXPERIMENTS

We conducted an evaluation of our method using Japanese newspaper articles. The following four models were compared:

(1) BM: We employ Ng and Cardie's [2002] model, as a baseline model. It identifies antecedents by the candidate-wise classification model, and determines anaphoricity using the one-step model.
(2) BM_SYN: BM with the syntactic features such as those in Figure 1(c).
(3) SCM: The selection-then-classification model explained in Section 3.
(4) SCM_SYN: SCM with all types of syntactic features shown in Figure 3.

5.1 Setting

We are now developing the corpus annotated with predicate-argument and coreference relations, which is named the NAIST Text Corpus [Iida et al.

Table II. Patterns Used in Antecedent Identification

no	weight	pattern
0)	0.016784	(R_B)
1)	0.014208	(R_P(R_P_HEAD_BF=.)(R_B(R_B_HEAD_BF=*wa*)))
2)	0.012211	(ROOT(R_ANT_HEAD_BF_*kai*)(R_SELECT_REST))
3)	0.011848	(L_P(L_B(L_B(L_B_HEAD_BF=,))))
4)	0.011493	(ROOT(R_B(R_B_LEFT))(R_PRED_AUX))
5)	0.011333	(R_ANT_HEAD_POS=*noun-suffix-personname*)
6)	0.011300	(ROOT(R_ANT_HEAD_POS=*noun-suffix-personname*)(L_B(L_B_LEFT)))
7)	0.011005	(ROOT(R_ANT_IN_QUOTE)(L_ANT_HEAD_POS=*noun-general*)(L_ANT_PARTICLE=*wa*))
8)	0.010829	(ROOT(L_B(L_P))(L_ANT_NE=DATE))
9)	0.010422	(ROOT(R_B(R_P))(R_ANT_HEAD_POS=*noun-general*))
...	...	...
868)	-0.009995	(L_ANT(L_B))
869)	-0.010357	(ROOT(R_ANT_CASE=NULL)(L_B(L_ANT)))
870)	-0.010369	(ROOT(R_ANT_AC_COOC_MI_PLUS1)(L_A_NE_PERSON))
871)	-0.010463	(ROOT(L_ANT_PARTICLE=)(L_ANT_SELECT_REST)(L_ANT_COOC=2))
872)	-0.010744	(L_B_LEFT(L_B_LEFT(L_ANT(L_ANT_HEAD_BF=*kara*))))
873)	-0.011066	(ROOT(R_ANT_PARTICLE=*ga*)(L_ANT_HEAD_POS=*noun-suffix-general*)(L_NP_PRED))
874)	-0.014086	(L_P(L_P_HEAD_BF=.)(L_B(L_B_HEAD_BF=*te*)))
875)	-0.014424	(R_ANT_HEAD_POS=*noun-adjv*)
876)	-0.020973	(L_B)
877)	-0.025117	(ROOT(R_P_HEAD_BF=*tyushisuru*)(L_ANT_NE_GPE)(L_ANT_EDR_ORG))

Patterns containing the prefix `L_` (`R_`) mean the patterns extracted from the tree T_L (T_R) shown in Figure 3. Node expressed by `L_ANT` (`R_ANT`) stands for the node `L.LeftCand` (`R.RightCand`). `L_P` (`R_P`) node is the node depended on by the node containing `L.`ϕ (`R.`ϕ). `L_B` (`R_B`) means the node intervening between `L.LeftCand` (`R.RightCand`) and the node depended on by `L.`ϕ (`R.`ϕ). Patterns having more than a zero score support candidate antecedents appearing on right side in comparison of two candidates. Otherwise, patterns support on the left of them.

2007].[6] Unfortunately, the NAIST Text Corpus still contains many unreliable tags. Therefore, we limit experiments to include only a reliable data set: 287 articles annotated by two annotators and 60 by one. The agreement ratio between two annotators on the 287 articles was 84.0%, which indicated that the annotation was sufficiently reliable. We discarded from this data set the zero-pronouns for which the two annotators disagreed. Consequently, the data set contained 1,384 intrasentential anaphoric zero-pronouns, 1,128 intersentential anaphoric zero-pronouns, and 784 nonanaphoric zero-pronouns (3,306 zero-pronouns in total), with each anaphoric zero-pronoun annotated to be linked to its antecedent. For each experiment, we used 137 articles for training, 60 articles for optimizing ϕ_{intra} (threshold parameter of intrasentential zero-anaphora resolution), and 150 articles for testing.

All the features were automatically acquired with the help of the following NLP tools: the Japanese morphological analyzer *ChaSen*[7] and the Japanese dependency structure analyzer *CaboCha*,[8] which also carried out named-entity chunking.

[6] The corpus is publicly available for research purposes at: http://cl.naist.jp/nldata/corpus/
[7] http://chasen.naist.jp/hiki/ChaSen/
[8] http://chasen.org/~taku/software/cabocha/

Table III. Patterns Used in Anaphoricity Determination

no	weight	pattern
0)	0.017534	(ANT_HEAD_BF=*titioya*)
1)	0.017006	(ROOT(B(B_LEFT))(P_AUX))
2)	0.016847	(B)
3)	0.014374	(ROOT(ANT_HEAD_POS=*noun-general*)(ANT_PARTICLE=*to*)(ANT_NP_PRED))
4)	0.014249	(ROOT(P(P_HEAD_BF=*te*))(ANT_HEAD_POS=*noun-pronoun-general*))
5)	0.014008	(ANT_HEAD_POS=*noun-suffix-personname*)
6)	0.013653	(ROOT(P_VOICE=active)(ANT_HEAD_POS=*unknown*))
7)	0.012654	(ROOT(P_VOICE=active)(ANT_HEAD_POS=*noun-propernoun-location-country*)(ANT_NP_PRED))
8)	0.011584	(ROOT(ANT_PARTICLE=*ga*)(ANT_EDR_PERSON)(ANT_EDR_ORG))
9)	0.011419	(P_HEAD_BF=*aru*)
...	...	...
675)	-0.010275	(ROOT(ANT_HEAD_BF=*kaikaku*)(ANT_PARTICLE=*wa*))
676)	-0.010323	(ROOT(ANT_HEAD_POS=*noun-suffix-general*)(ANT_IN_QUOTE)(ANT_SELECT_REST))
677)	-0.010579	(ROOT(B(B_HEAD_BF=*iru*))(ANT_EDR_PERSON))
678)	-0.010616	(ROOT(P_HEAD_BF=*dekiru*)(ANT_COOC=1))
679)	-0.010887	(B(B_RIGHT)(B_RIGHT_HEAD_BF=*iru*)(B_RIGHT_HEAD_BF=*to*))
680)	-0.011002	(ANT_HEAD_BF=*naiyou*)
681)	-0.011625	(ROOT(B(P))(ANT_HEAD_POS=*noun-adverbial*))
682)	-0.012710	(ROOT(B(B_LEFT))(ANT_COOC=2))
683)	-0.013327	(P_HEAD_BF=*tari*)
684)	-0.014274	(ROOT(P_VOICE=active)(ANT_HEAD_POS=*noun-suffix-personname*)(ANT_PARTICLE=*no*))

Table IV. Accuracy of Antecedent Identification

BM	BM_SYN	SCM	SCM_SYN
0.340	0.559	0.691	**0.729**
(162/476)	(266/476)	(329/476)	(347/476)

5.2 Results on Intrasentential Zero-Anaphora Resolution

In both intra-anaphoricity determination and antecedent identification, we investigated the effect of introducing the syntactic features for improving the performance. First, the results of antecedent identification are shown in Table IV. The comparison between BM (SCM) with BM_SYN (SCM_SYN) indicates that introducing the structural information effectively contributes to this task. In addition, the large improvement from BM_SYN to SCM_SYN indicates that the use of the preference-based model has significant impact on intrasentential antecedent identification. This finding may well contribute to semantic role labeling because these two tasks have a large overlap as discussed in Section 1.

Second, to evaluate the performance of intrasentential zero-anaphora resolution, we plotted recall-precision curves altering threshold parameter and θ_{inter} for intra-anaphoricity determination as shown in Figure 5, where recall R and precision P were calculated by:

$$R = \frac{\text{\# of detected antecedents correctly}}{\text{\# of anaphoric zero-pronouns}},$$

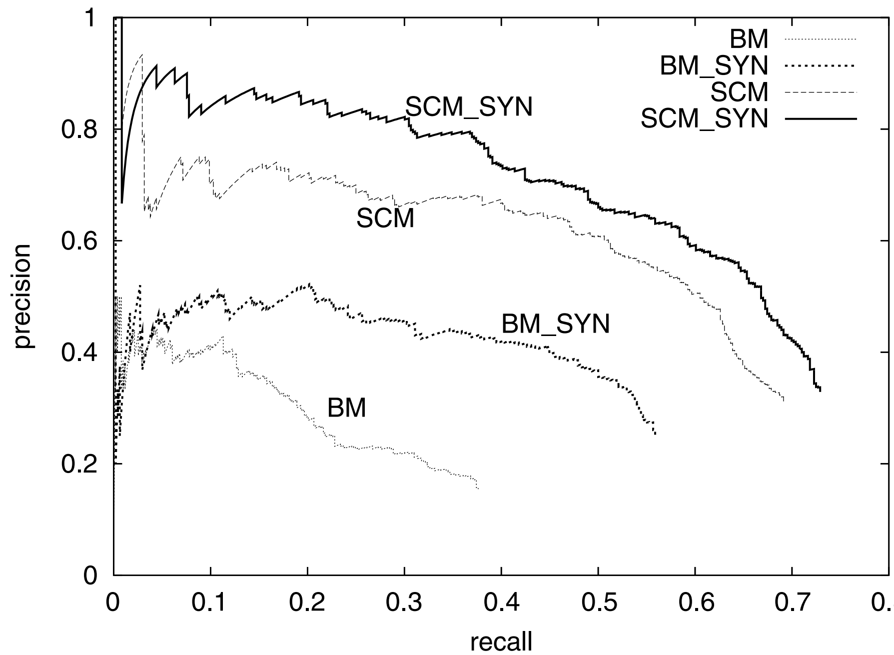

Fig. 5. Recall-precision curves of intrasentential zero-anaphora resolution.

Table V. Results of Intrasentential Zero-Anaphora Resolution Using Estimated Threshold Parameter

	BM	BM_SYN	SCM	SCM_SYN
Recall	0.311	0.398	0.613	0.596
Precision	0.213	0.419	0.487	0.595
F-measure	0.253	0.408	0.543	0.595

$$P = \frac{\text{\# of detected antecedents correctly}}{\text{\# of zero-pronouns classified as anaphoric}}.$$

The curves indicate the upper bound of the performance of these models; in practical settings, the parameters have to be trained beforehand.

Figure 5 shows that BM_SYN (SCM_SYN) outperforms BM (SCM), which indicates that incorporating syntactic pattern features works remarkably well for intrasentential zero-anaphora resolution. Futhermore, SCM_SYN is significantly better than BM_SYN. This is because the former has an advantage of learning nonanaphoric zero-pronouns as negative training instances in intrasentential anaphoricity determination, which enables it to reject nonanaphoric zero-pronouns more accurately than the others.

We also evaluate the performance in case that the threshold parameter is automatically estimated with the development data set. The result of each model is shown in Table V. Note that these thresholds are chosen when the F-measure of each model is maximized. Table V shows that the SCM_SYN achieved the best performance among them, indicating that our model has reasonable performance in practical situations.

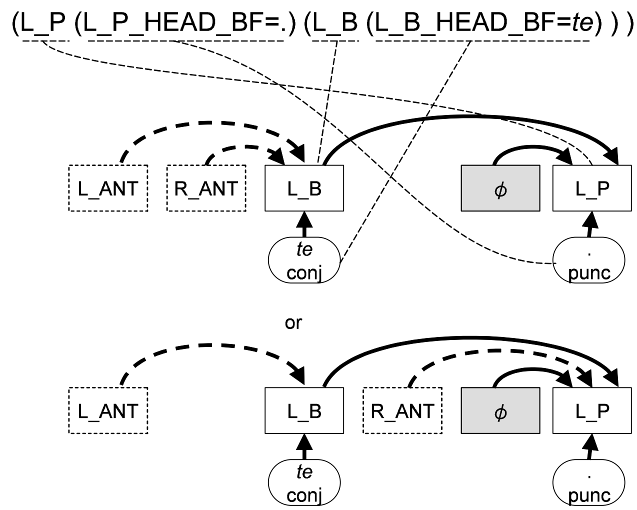

Fig. 6. Example of extracted syntactic pattern.

Next, we investigated what kinds of syntactic patterns effectively work by manually analyzing the decision stumps (syntactic patterns) induced by BACT. The extracted patterns for antecedent identification and anaphoricity determination are shown in Table II and Table III respectively. Table II shows some patterns used in antecedent identification; the patterns with positive weight supports the left candidate of the two competitors. Table III shows some patterns used in anaphoricity determination; the patterns with positive weight supports that a candidate anaphor is anaphoric.

Each syntactic pattern is expressed in S-expression form; a node *P* denotes a predicate that has a zero-pronoun, and *ANT* node denotes a candidate antecedent. For example, Pattern No. 874 from Table II is depicted in Figure 6. In this figure the node prefix *L_* indicates a pattern taken from the left antecedent path, likewise, *R_* indicates a pattern from the right antecedent path. The current left candidate (L_ANT) and right candidate (R_ANT) are not explicitly present in this decision stump. However, as L_ANT must be in the left candidate path, it should be a descendant of the L_B node. R_ANT must be to the right of L_ANT, so it should be a descendant of either the *L_B* or *L_P* nodes as the two graphs in Figure 6 show. Since it is given a negative weight, this pattern favors the left candidate. Interestingly enough, this rule can be interpreted linguistically as follows.

Zero-anaphoric phenomena have been extensively researched in the last two decades in the area of Japanese linguistics. Minami [1974], for example, manually examined what patterns between two predicates are helpful for interpreting zero-anaphoric relations. Minami argued that "If two predicates are connected with the conjunctive particle *te*, the nominative arguments of the two predicates tend to be shared in the same sentence" as the useful patterns. Also, Nariyama [2002] introduced such patterns into her centering-based algorithm to zero-anaphora resolution and showed that the patterns argued by

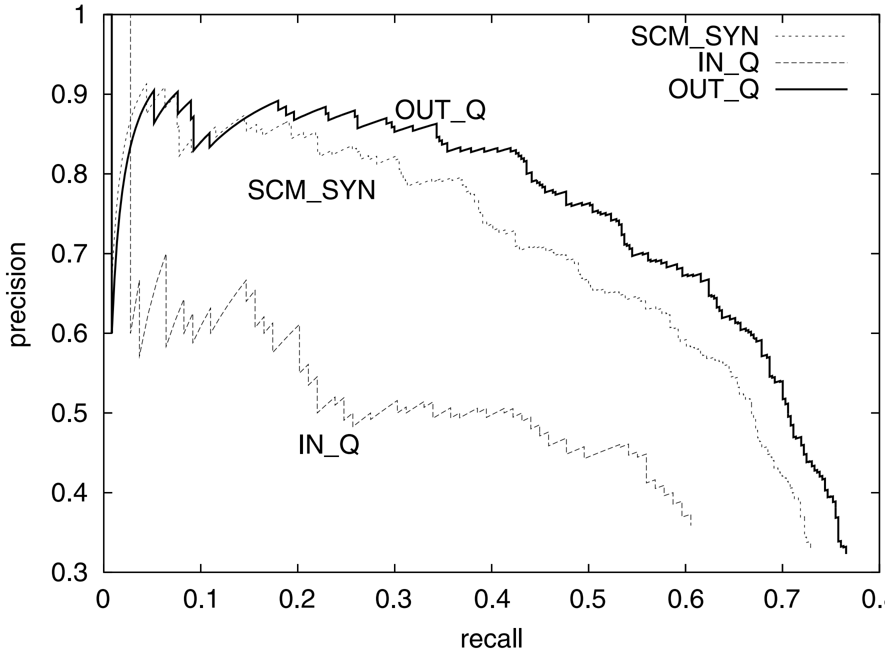

Fig. 7. Recall-precision curves of resolving in-quote and out-quote zero-pronouns.

Minami contributes to improving performance quantitatively in her manual evaluation. Note that the pattern described by Minami loosely corresponds to Pattern No. 874 described above. It is indirect evidence that learning syntactic patterns by BACT contributes to improving the performance of both antecedent identification and anaphoricity determination.

5.3 Discussion

Our error analysis reveals that a majority of errors can be attributed to the current way of handling quoted phrases and sentences. Figure 7 shows the difference in resolution accuracy between zero-pronouns appearing in a quotation (185 zero-pronouns) and the rest (871 zero-pronouns), where "IN_Q" denotes the former (in-quote zero-pronouns) and "OUT_Q" the latter. The accuracy on the IN_Q problems is considerably lower than that on the OUT_Q cases, which indicates that we should deal with in-quote cases with a separate model so that it can take into account the nested structure of discourse segments introduced by quotations.

To investigate the effect by separating in-quote zero-pronouns from out-quote ones, we trained two cases independently for the SCM model using syntactic pattern features. At the test phase, the system utilizes either a model that was trained from in-quote instances or one from out-quote cases depending on the context. Figure 8 shows recall-precision curves of such a separate model and the original SCM model. Figure 8 shows that the separate model does not work compared with the original one. It is because the relationship

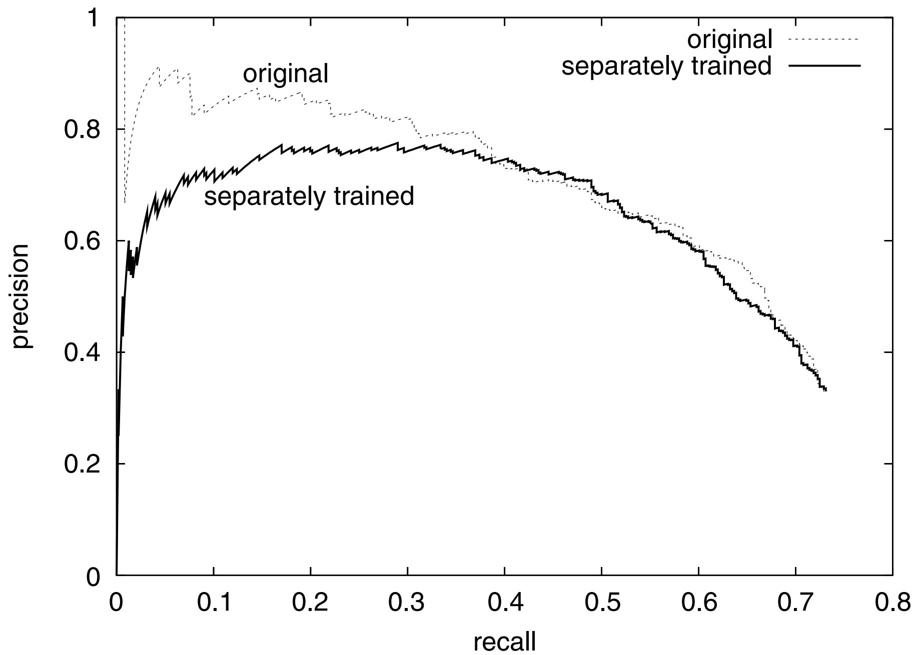

Fig. 8. Effect by separating in-quote zero-pronouns from out-quote ones.

between a zero-pronoun in a quoted sentence and its antecedent is too complicated to automatically extract effective syntactic patterns. As quote sentences, speakers are able to utter anything as long as the utterance is coherent to the preceding context. Thus, quotes can contain various types of expressions; word, clause, sentence, or text, causing severe data sparseness problems. To avoid them, we need to adapt the current model to each situation appearing in an in-quote zero-pronoun and its antecedent.

5.4 Impact on Overall Zero-Anaphora Resolution

We next evaluated the effects of introducing the proposed model on overall zero-anaphora resolution including intersentential cases.

As a baseline model, we implemented the original SCM, designed to resolve intrasentential zero-anaphora and intersentential zero-anaphora simultaneously with no syntactic pattern features. Here, we adopted Support Vector Machines [Vapnik 1998] to train the classifier on the baseline model and the intersentential zero-anaphora resolution in the SCM using structural information.

For the proposed model, we plotted several recall-precision curves by selecting different value for threshold parameters θ_{intra} and θ_{inter}. The results are shown in Figure 9, which indicates that the proposed model significantly outperforms the original SCM if θ_{intra} is appropriately chosen.

We then investigated the feasibility of parameter selection for θ_{intra} by plotting the area under the curve (AUC) values for different θ_{intra} values. Here, each AUC value is the area under a recall-precision curve. The results are

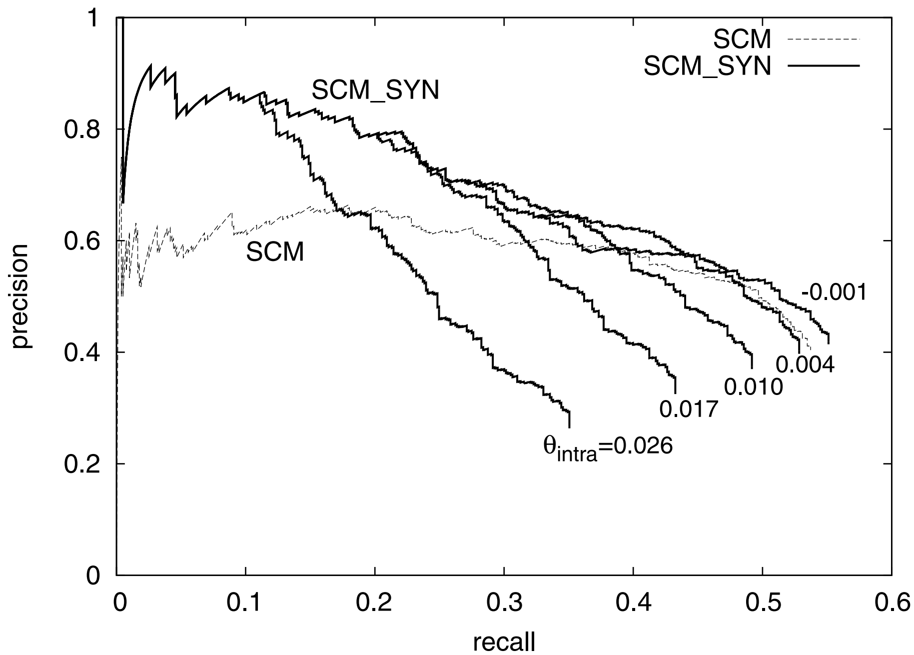

Fig. 9. Recall-precision curves of overall zero-anaphora resolution.

shown in Figure 10. Since the original SCM does not use θ_{intra}, the AUC value of it is constant, depicted by the SCM. As shown in Figure 10, the AUC-value curve of the proposed model does not have a well-defined peak, which indicates the selection of parameter θ_{intra} is not difficult.

6. CONCLUSION

In intrasentential zero-anaphora resolution, syntactic patterns of the appearance of zero-pronouns and their antecedents are useful clues. Taking Japanese as a target language, we have empirically demonstrated that incorporating rich syntactic pattern features in a state-of-the-art learning-based anaphora resolution model dramatically improves the accuracy of intrasentential zero-anaphora, which consequently improves the overall performance of zero-anaphora resolution. We approach the zero-anaphora resolution problem by decomposing it into two subtasks: intrasentential and intersentential zero-anaphora resolution. We built a separate component for each subtask, adopting Iida et al. [2005]'s selection-then-classification model. In intrasentential zero-anaphora resolution, the model treats syntactic patterns as features for both antecedent identification and anaphoricity determination.

In our next step, we are going to address the issue of how to find zero-pronouns, which requires us to design a broader framework that allows zero-anaphora resolution to interact with predicate-argument structure analysis. With regard to this problem, we need a verb dictionary such as VerbNet [Kipper 2005] or FrameNet [Baker et al. 1998] for verb sense disambiguation. In

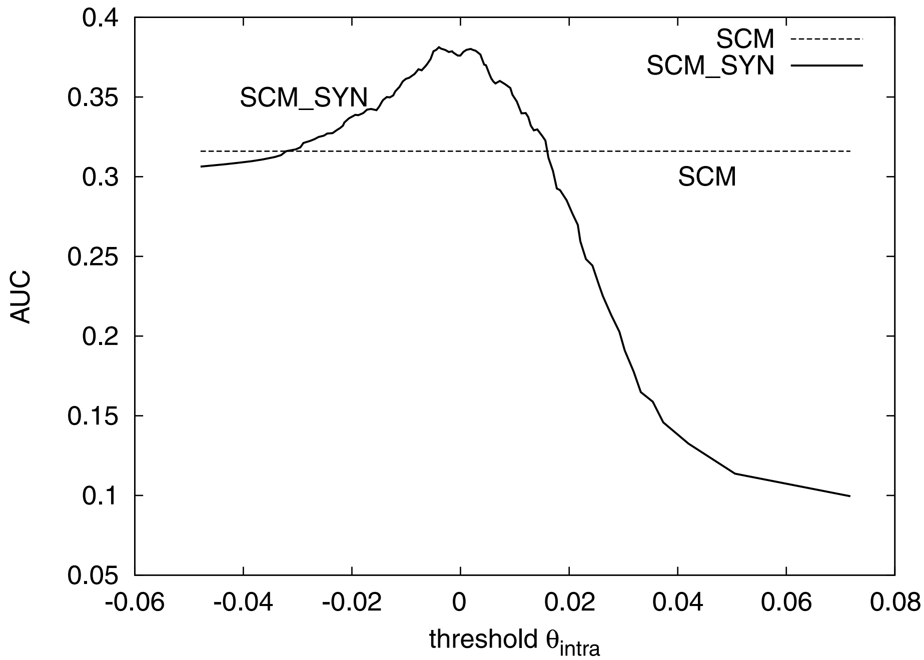

Fig. 10. AUC curves plotted by altering θ_{intra}.

the past decade, automatic verb frame construction methods have received increasing attention [Resnik 1993; Utsuro and Matsumoto 1997; Kawahara et al. 2000; Gildea 2002, etc.], which cluster verbs and arguments based on the similarity between instances in corpora. By adopting such a strategy, we obtain a scalable frame dictionary constructed automatically from large amounts of text data such as that available on the Web. Of course, dictionaries that are automatically constructed are noisy and not always fine grained; however, we are convinced that it is beneficial to introduce dictionaries into the process of anaphora resolution in light of these issues.

Another important issue is how to find a globally optimal solution to the set of zero-anaphora resolution problems in a given discourse. In the problem of zero-anaphora resolution, there are often more than one zero-pronoun for a given predicate. Thus, we have to consider the interdependency between zero-pronouns. For example, the model must resolve zero-anaphoric relations with the constraints that an NP which is the nominative case for a given predicate does not become the dative case. This issue leads us to explore methods as discussed by McCallum and Wellner [2003].

The computational cost of introducing syntactic patterns is also one of importance. Although using syntactic patterns improves the accuracy of both anaphoricity determination and antecedent identification, it involves a higher cost in matching between an input tree and each of its decision stumps. Efficient computation is beyond the scope of this article, but reducing computational cost without decreasing performance is an important issue in this area.

REFERENCES

ASAHARA, M. AND MATSUMOTO, Y. 2003. *IPADIC User Manual*. Nara Institute of Science and Technology, Japan.

BAKER, C. F., FILMORE, C. J., AND LOWE, J. B. 1998. The Berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics* (*COLING-ACL*). 86–90.

BALDWIN, B. 1995. Cogniac: A discourse processing engine. Ph.D. thesis, Department of Computer and Information Sciences, University of Pennsylvania.

BREIMAN, L. 1999. *Neural Computa. 11,* 7, 1493–1518.

CARRERAS, X. AND MARQUEZ, L. 2005. Introduction to the ConLL-2005 shared task: Semantic role labeling. In *Proceedings of the 9th Conference on Computational Natural Language Learning* (*CoNLL*). 152–164.

COLLINS, M. AND DUFFY, N. 2001. Convolution kernels for natural language. In *Proceedings of the Neural Information Processing Systems* (*NIPS*). 625–632.

FUJITA, A., INUI, K., AND MATSUMOTO, Y. 2004. Detection of incorrect case assignments in automatically generated paraphrases of Japanese sentences. In *Proceedings of the 1st International Joint Conference on Natural Language Processing* (*IJCNLP*). 14–21.

GE, N., HALE, J., AND CHARNIAK, E. 1998. A statistical approach to anaphora resolution. In *Proceedings of the 6th Workshop on Very Large Corpora*. 161–170.

GILDEA, D. 2002. Probabilistic models of verb-argument structure. In *Proceedings of the 19th International Conference on Computational Linguistics* (*COLING*). 308–314.

GILDEA, D. AND JURAFSKY, D. 2002. Automatic labeling of semantic roles. *Computa. Linguist.*, *28*, 3, 245–288.

GROSZ, B. J., JOSHI, A. K., AND WEINSTEIN, S. 1995. Centering: A framework for modeling the local coherence of discourse. *Computa. Linguist., 21,* 2, 203–226.

HOBBS, J. 1978. Resolving pronoun references. *Lingua 44*, 311–338.

IIDA, R., INUI, K., AND MATSUMOTO, Y. 2005. Anaphora resolution by antecedent identification followed by anaphoricity determination. *ACM Trans. Asian Lang. Inf. Process. 4,* 4, 417–434.

IIDA, R., INUI, K., TAKAMURA, H., AND MATSUMOTO, Y. 2003. Incorporating contextual cues in trainable models for coreference resolution. In *Proceedings of the 10th EACL Workshop on the Computational Treatment of Anaphora*. 23–30.

IIDA, R., KOMACHI, M., INUI, K., AND MATSUMOTO, Y. 2007. Annotating a Japanese text corpus with predicate-argument and coreference relations. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics* (*ACL*) *Workshop on Linguistic Annotation Workshop*, 132–139.

IKEHARA, S., MIYAZAKI, M.,YOKOO, A., SHERAI, S., NAKAIWA, H., OGURA, K., OOYAMA, Y., AND HAYASHI., Y. 1997. *Nihongo Goi Taikei* (in Japanese), Iwanami Shoten.

ISOZAKI, H. AND HIRAO, T. 2003. Japanese zero pronoun resolution based on ranking rules and machine learning. In *Proceedings of the Conference on Empirical Methods in Natural language Processing* (*EMNLP*). 184–191.

Japan Electronic Dictionary Research Institute, Ltd. Japan. 1995. *EDR Electronic Dictionary Technical Guide*. Japan Electronic Dictionary Research Institute, Ltd. Japan.

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

KAWAHARA, D., KAJI, N., AND KUROHASHI, S. 2000. Japanese case structure analysis by unsupervised construction of a case frame dictionary. In *Proceedings of the 18th International Conference on Computational Linguistics* (*COLING*). 432–438.

KIPPER, K. 2005. Verbnet: A broad-coverage, comprehensive verb lexicon. Ph.D. thesis, Department of Computer and Information Science University of Pennsylvania.

KUDO, T. AND MATSUMOTO, Y. 2002. Japanese dependency analysis using cascaded chunking. In *Proceedings of the 6th Conference on Natural Language Learning 2002* (*CoNLL*). 63–69.

KUDO, T. AND MATSUMOTO, Y. 2004. A boosting algorithm for classification of semi-structured text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (*EMNLP*). 301–308.

KUROHASHI, S. AND NAGAO, M. 1997. Kyoto university text corpus project. In *Proceedings of the 3rd Annual Meeting of the Association for Natural Language Processing* (*ANLP*) (in Japanese). 115–118.

LAPPIN, S. AND LEASS, H. J. 1994. An algorithm for pronominal anaphora resolution. *Computa. Linguist. 20,* 4, 535–561.

LUO, X. AND ZITOUNI, I. 2005. Multi-lingual coreference resolution with syntactic features. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*. 660–667.

MCCALLUM, A. AND WELLNER, B. 2003. Object consolidation by graph partitioning with a conditionally trained distance metric. In *Proceedings of the KDD Workshop on Data Cleaning, Record Linkage, and Object Consolidation*. 19–24.

MCCARTHY, J. F. AND LEHNERT, W. G. 1995. Using decision trees for coreference resolution. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (*IJCAI*). 1050–1055.

MINAMI, F. 1974. *Gendai Nihongo no Kouzou* (in Japanese). Taishukan.

MITKOV, R. 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*. 14–21.

MURATA, M. AND NAGAO, M. 1997. Resolution of verb ellipsis in Japanese sentences using surface expressions and examples. In *Proceedings of 4th Natural Language Processing Pacific Rim Symposium*. 75–80.

NAKAIWA, H. AND SHIRAI, S. 1996. Anaphora resolution of Japanese zero pronouns with deictic reference. In *Proceedings of the 16th International Conference on Computational Linguistics* (*COLING*). 812–817.

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

NG, V. 2004. Learning noun phrase anaphoricity to improve coreference resolution: Issues in representation and optimization. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics* (*ACL*). 152–159.

NG, V. AND CARDIE, C. 2002. Improving machine learning approaches to coreference resolution. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (*ACL*). 104–111.

OKUMURA, M. AND TAMURA, K. 1996. Zero pronoun resolution in Japanese discourse based on centering theory. In *Proceedings of the 16th International Conference on Computational Linguistics* (*COLING*). 871–876.

POESIO, M., URYUPINA, O., VIEIRA, R., ALEXANDROV-KABADJOV, M., AND GOULART, R. 2004. Discourse-new detectors for definite description resolution: A survey and a preliminary proposal. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics* (*ACL*) *Workshop on Reference Resolution and its Applications*. 47–54.

RESNIK, P. 1993. Selection and information: A class-based approach to lexical relationships. Ph.D. thesis, Department of Computer and Information Science, University of Pennsylvania.

SEKI, K., FUJII, A., AND ISHIKAWA, T. 2002. A probabilistic method for analyzing Japanese anaphora integrating zero pronoun detection and resolution. In *Proceedings of the 19th International Conference on Computational Linguistics* (*COLING*). 911–917.

SOON, W. M., NG, H. T., AND LIM, D. C. Y. 2001. A machine learning approach to coreference resolution of noun phrases. *Computa. Linguist. 27,* 4, 521–544.

STRUBE, M. AND MÜLLER, C. 2003. A machine learning approach to pronoun resolution in spoken dialogue. In *Proceedings of the 41st ACL*. 168–175.

SUZUKI, J., HIRAO, T., SASAKI, Y., AND MAEDA, E. 2003. Hierarchical directed acyclic graph kernel: Methods for structured natural language data. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics* (*ACL*). 32–39.

UTSURO, T. AND MATSUMOTO, Y. 1997. Learning probabilistic subcategorization preference by identifying case dependencies and optimal noun class generation level. In *Proceedings of the 5th Conference on Applied Natural Language Processing*. 364–371.

VAPNIK, V. N. 1998. *Statistical Learning Theory. Adaptive and Learning Systems for Signal Processing Communications, and Control*. John Wiley & Sons.

WALKER, M., IIDA, M., AND COTE, S. 1994. Japanese discourse and the process of centering. *Computa. Linguist. 20,* 2, 193–233.

YANG, X., SU, J., AND TAN, C. L. 2005. Improving pronoun resolution using statistics-based semantic compatibility information. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics* (*ACL*). 165–172.

YANG, X., SU, J., AND TAN, C. L. 2006. Kernel-based pronoun resolution with structured syntactic knowledge. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics* (*COLING-ACL*). 41–48.

YANG, X., ZHOU, G., SU, J., AND TAN, C. L. 2003. Coreference resolution using competition learning approach. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics* (*ACL*). 176–183.

Received February 2007; revised August 2007; accepted August 2007