

# Resolving Direct and Indirect Anaphora for Japanese Definite Noun Phrases

NAOYA INOUE<sup>†</sup>, RYU IIDA<sup>††</sup>, KENTARO INUI<sup>†</sup> and YUJI MATSUMOTO<sup>†</sup>

An anaphoric relation can be either direct or indirect. In some cases, the antecedent being referred to lies outside of the discourse its anaphor belongs to. Therefore, an anaphora resolution model needs to consider the following two decisions in parallel: *antecedent selection*—selecting the antecedent itself, and *anaphora type classification*—classifying an anaphor into direct anaphora, indirect anaphora or exophora. However, there are non-trivial issues for taking these decisions into account in anaphora resolution models since the anaphora type classification has received little attention in the literature. In this paper, we address three non-trivial issues: (i) how the antecedent selection model should be designed, (ii) what information helps with anaphora type classification, (iii) how the antecedent selection and anaphora type classification should be carried out, taking Japanese as our target language. Our findings are: first, an antecedent selection model should be trained separately for each anaphora type using the information useful for identifying its antecedent. Second, the best candidate antecedent selected by an antecedent selection model provides contextual information useful for anaphora type classification. Finally, the antecedent selection should be carried out before anaphora type classification.

**Key Words:** Anaphora resolution, antecedent selection, anaphora type classification, direct anaphora, indirect anaphora, exophora

## 1 Introduction

Anaphora resolution has been studied intensively in recent years because of its significance in many natural language processing (NLP) applications such as information extraction and machine translation. In nominal anaphora, an anaphor (typically a definite noun phrase) and its antecedent in the preceding discourse holds either a *direct* anaphoric relation (e.g. coreference) or an *indirect* relation (e.g., bridging reference (Clark 1977)). *Direct anaphoric relation* refers to a link in which an anaphor and an antecedent are in such a relation as *synonymy* and *hyponymy/hyponymy*, as in *house–building*. *Indirect anaphoric relation*, on the other hand, refers to a link in which an anaphor and an antecedent have such relations as *meronymy/holonymy* and *attribute/value* as in *ticket–price*. For the other case, a noun phrase occasionally holds an

<sup>†</sup>Graduate School of Information Science, Nara Institute of Science and Technology

<sup>††</sup>Department of Computer Science, Tokyo Institute of Technology

*exophoric* relation to an antecedent that lies outside the discourse that the noun phrase presents. Recent studies in anaphora resolution have proposed the resolution frameworks for each direct and indirect anaphoric case respectively (Soon et al. 2001; Iida et al. 2005; Poesio et al. 2004), placing the main focus on the *direct* anaphoric case. The identification of exophoric relations, in contrast, has been paid little attention in the literature. Anaphoricity determination, which is the task of determining whether an anaphor has an antecedent in the preceding discourse or not, is related to identifying exophoric relations, but the methods for anaphoricity determination are not designed to explicitly capture exophoric relations because they are tuned for finding NP coreference chains in discourse.

However, for the practical use of anaphora resolution, we need to solve the following non-trivial problem: in a real text, anaphors such as definite noun phrases can occur as either direct anaphoric, indirect anaphoric or exophoric relations, which is not easy to disambiguate from its surface expression. That is, in anaphora resolution, it is necessary to judge what kind of anaphoric relation is used to tie an anaphor and its (potential) antecedent (henceforth, we call this task *anaphora type classification*). In fact, our corpus analysis (detailed in Section 5) shows that more than 50% of noun phrases modified by a definiteness modifier have non-trivial ambiguity in terms of the anaphora types that have to be classified for each given text. Given these issues, we decompose the task of nominal anaphora resolution as a combination of two distinct but arguably interdependent subtasks.

- *Antecedent selection*: the task of identifying the antecedent of a given anaphor, and
- *Anaphora type classification*: the task of judging what kind of *anaphora type* is used for a given anaphor, i.e., classifying a given anaphor into *direct anaphoric*, *indirect anaphoric* or *exophoric*.

Given this task decomposition, three unexplored issues immediately come up:

**Issue 1.** Whether the model for antecedent selection should be designed and trained separately for direct anaphora and indirect anaphora or whether it can be trained as a single common model;

**Issue 2.** What contextual information is useful for determining each anaphora type;

**Issue 3.** How the two subtasks can be best combined (e.g. which subtask should be carried out first).

In this paper, we explore these issues taking Japanese as our target language. Specifically, we focus on anaphora resolution for noun phrases modified by a definiteness modifier, as detailed in the next section.

This paper is organized as follows. In the next section, we describe our motivation for this work

more specifically. In Section 3, we review previous work of antecedent selection and anaphora type classification. In Section 4, we give a detailed explanation of our investigations. In Section 5, the dataset for our experiments is described. We then show the experimental setup and results of our investigations and discussion in Section 6. Finally, the conclusion is presented along with future work.

## 2 Motivation for our approach

As mentioned, an anaphor can hold a direct or indirect relation with its antecedent. Occasionally, an anaphor refers to an antecedent that is not in the same discourse. The terms *direct anaphora* and *indirect anaphora* have been used to denote some different anaphoric phenomena in previous works, e.g. *direct anaphora* in (Vieira and Poesio 2000) indicates only the reference that an anaphor and its antecedent have identical head words, whereas *direct anaphora* in (Mitkov et al. 2000) includes a synonymous or generalization/specialization link of an anaphor and its antecedent. As a result, we redefine the following three *anaphora types* to denote the use of anaphoric expressions in our classification task:

- *direct anaphora*: An anaphor refers to its antecedent directly. In example (1), *The CD* refers to *the new album* directly.

(1) Her new album was released yesterday. I want to get the CD as soon as possible.

- *indirect anaphora*: An anaphor has an antecedent related with the anaphor rather than referred to, as in example (2).

(2) The artist announced her new song. I want to get the CD as soon as possible.

*The CD* refers to *her new song* indirectly. The discourse entity that directly corresponds to *the CD* is not in the preceding sentence; instead *her new song* is considered as an antecedent of *the CD* because it is associated with *the CD*.

- *exophora*: An anaphor that has no antecedent in a text is regarded as exophoric. An exophoric expression is typically used in newspaper articles; for instance, *the day* refers to the date of the post.

For our target language, Japanese, noun phrases (NP) behave similarly to those in English; that is, a definite NP may bear a direct anaphoric relation but may also bear an indirect anaphoric relation to its antecedent as shown in examples (3) and (4).

(3) 新しいミニバン<sub>(i')</sub>が発売された。この新型車<sub>(i)</sub>は燃費が非常によい。

A new minivan<sub>(i')</sub> was released. The vehicle<sub>(i)</sub> has good gas mileage.

- (4) 家具屋で机<sub>(i')</sub>を見た。そのデザイン<sub>(i)</sub>は素晴らしかった。

*I saw a desk<sub>(i')</sub> in a furniture shop. The design<sub>(i)</sub> was marvelous.*

“この新型車 (*the vehicle*)” refers to “新しいミニバン (*a new minivan*)” directly in (3), while “そのデザイン (*the design*)” refers to “机 (*a desk*)” indirectly in (4). As seen from the above examples (1), (2) and reported in Section 1, the anaphora type can be different for a unique expression. In other words, the anaphora type has to be disambiguated taking its appearing context into account.

In Japanese, however, the problem can be even more complex because a definite NP is not always marked by a definiteness modifier, such as *this*(この), *the*(その), or *that*(あの). For example, bare NP 大統領 (*president*) refers to 韓国大統領 (*Korean President*) in text (5).

- (5) 今月4日、韓国大統領<sub>(i')</sub>が来日した。大統領<sub>(i)</sub>は翌日の記者会見で新プランの詳細を語った。

*Korean President<sub>(i')</sub> visited Japan on the 4th this month. (The) president<sub>(i)</sub> talked about the details of his new plan at the news conference next day.*

For this reason, it is sometimes difficult even for human annotators to determine the definiteness of a bare NP. As the first step toward complete understanding of Japanese NP anaphora, we focus on anaphora resolution for NPs marked with either *this NP* (この+NP), *the NP* (その+NP) or *that NP* (あの+NP), which account for a large proportion of occurrences of nominal anaphora in Japanese texts.

### 3 Related work

In this section, we review previous research on anaphora resolution for antecedent selection and anaphora type classification respectively. In Section 3.1, we look over how the previous work had taken the approaches to antecedent selection for direct anaphora and indirect anaphora. In Section 3.2, we discuss Vieira and Poesio’s work and Nakaiwa’s work on anaphora type classification.

#### 3.1 Antecedent selection

A wide range of approaches to anaphora resolution has been proposed in earlier work. There exist two main approaches: rule-based approaches and machine learning-based approaches. In contrast to the rule-based approaches such as (Brennan et al. 1987; Shalom and J. 1994; Baldwin 1995; Nakaiwa et al. 1995a; Okumura and Tamura 1996; Mitkov 1997), empirical, or machine learning-based approaches have been shown to be a cost-efficient solution achieving performance

that is comparable to the best performing rule-based systems (McCarthy and Lehnert 1995; Ge, Hale, and Charniak 1998; Soon et al. 2001; Ng and Cardie 2001; Strube and Muller 2003; Iida et al. 2005; Yang, Zhou, Su, and Tan 2003, etc.). Most of these studies focus only on the coreference resolution task, particularly in the context of evaluation-oriented research programs such as Message Understanding Conference (MUC)<sup>1</sup> and Automatic Content Extraction (ACE)<sup>2</sup>. To the contrary, the methods for indirect anaphora resolution have been relatively unexplored compared with direct anaphora. Those works are implemented by rule-based approaches (Poesio et al. 1997; Murata et al. 1999; Bunescu 2003, etc.) and learning-based approaches (Poesio et al. 2004), encoding the centering theory (Grosz et al. 1995), lexical resources such as WordNet (Fellbaum 1998) and web-based knowledge. In comparison to direct anaphora, the resolution of indirect anaphora is still a much more difficult task because it is required to capture the wide variety of semantic relations (e.g. *store-the discount*, *drilling-the activity*). For example, (Poesio et al. 2002) proposed acquiring the lexical knowledge of the meronymy relations for resolving bridging descriptions by using syntactic patterns such as *the NP of NP* and *NP's NP*.

Recall that these works are based on the assumption that the system knows that the given anaphor is direct anaphora or indirect anaphora, which motivates us to explore the design of the antecedent selection model.

### 3.2 Anaphora type classification

As mentioned in Section 1, there has been little attention paid to the issue of anaphora type classification. Exceptions can be seen in (Nakaiwa et al. 1995b) and (Vieira and Poesio 2000). Nakaiwa's work focuses on the extra-sentential resolution of Japanese zero pronouns in machine translation. They identify zero pronouns whose referent is the extra-sentential element such as *I*, *we* and *you* by using the semantic constraints such as modal expressions, verbal semantic attributes. In their classification, the verbs depended on by pronouns are important clues, whereas the contextual information is important in anaphora type classification as mentioned in Section 2.

Vieira and Poesio's work (2000) is motivated by corpus study for the use of definite descriptions<sup>3</sup>. Their system does not only find an antecedent but classifies a given definite description into the following three categories.

- *direct anaphora*: subsequent-mention definite descriptions that refer to an antecedent with

<sup>1</sup>[http://www-nlpir.nist.gov/related\\_projects/muc/index.html](http://www-nlpir.nist.gov/related_projects/muc/index.html)

<sup>2</sup><http://www.nist.gov/speech/tests/ace/>

<sup>3</sup>Noun phrases with the definite article *the*.

the same head noun as the description;

- *bridging descriptions*: definite descriptions that either (i) have an antecedent denoting the same discourse entity, but using a different head noun (as in *house ... building*), or (ii) are related by a relation other than identity to an entity already introduced in the discourse;
- *discourse-new*: first-mention definite descriptions that denote objects not related by shared associative knowledge to entities already introduced in the discourse.

Compared with our taxonomy, their definition of *direct anaphora* is restricted to the case where an anaphor and its antecedent have an identical head. Therefore, the other cases (e.g. a pair of *new album* and *the CD*) are not regarded as direct anaphora but such cases are classified into bridging descriptions. The definition of *discourse-new*, on the other hand, refers to the same notion as our definition of *exophora* except that the generic use of the definite article *the* as in *play the piano* is classified into *discourse-new*. Note that Japanese definiteness modifiers are not used in such a way.

In their work, the system chooses the correct anaphora type of a given definite NP and if possible, finds its antecedent following a set of hand-coded rules on the basis of the lexical and syntactic features. The process can be regarded as four notable steps.

1. The system applies some heuristics exploiting lexical and syntactic features based on (Hawkins 1978) to detect non-anaphoric cases ('unfamiliar use' or 'larger situation use' in Hawkins's work) to an anaphor. If the test succeeds, it interprets the anaphor as discourse-new.
2. The system tries to find a same-head antecedent (i.e., an antecedent as direct anaphora) from a set of potential candidates appearing in the preceding discourse. If a suitable candidate is found, the system classifies an anaphor as *direct anaphora* and returns the candidate as its antecedent.
3. The rules to recognize discourse-new, such as 'pre-modifier use' and 'proper noun use' (e.g. *the United States*), are applied to an anaphor. If the test succeeds, the anaphor is classified as discourse-new.
4. The system tries to find an NP associated with an anaphor (which is called *an anchor* in their work) in the preceding discourse. If such an NP is found, the anaphor is classified as bridging description and judges the NP as its anchor. Otherwise, the system does not output anymore.

The heuristics to detect non-anaphoric or discourse-new anaphors are based on the syntactic and lexical features, while the rules for direct anaphora and bridging descriptions simply try to find an antecedent. Consequently, their work can be said to focus on detecting discourse-new descriptions compared to our work. They reported their system achieved 57% recall and 70%

precision in their empirical evaluation.

Note that their system carries out anaphora type classification before antecedent selection. However, it remains unexplored how to integrate antecedent identification and anaphora type classification into anaphora resolution, which is to be investigated as *issue 2* and *issue 3*, which we addressed in Section 1.

## 4 Model

The purpose of our work is to investigate the three unexplored issues shown in Section 1. First of all, we explain our learning-based antecedent selection models and anaphora type classification models.

### 4.1 Antecedent selection

One issue to explore in antecedent selection is whether a single common model should be built for both direct and indirect anaphora or a separate model should be built for each. In this section, in order to explore *issue 1*, we design two different models for selecting antecedents.

From the point of view in which we consider both anaphora types in parallel in an antecedent identification, we can consider the following two strategies.

- *Single model*: Designing the model for the resolution of both direct and indirect anaphora. The information to capture an direct-anaphoric antecedent and indirect-anaphoric antecedent is jointly incorporated into a single common model. The model is trained with labeled examples of both direct and indirect anaphora.
- *Separate model*: Preparing two distinct models for each anaphora type separately; i.e., the selection model for direct anaphora and the model for indirect anaphora. Unlike the *single model*, each model incorporates the information to capture an antecedent for each anaphora type separately. In the *direct antecedent selection model*, only the information that captures a direct-anaphoric antecedent is used. In the *indirect antecedent selection model*, on the other hand, only the information for the indirect-anaphoric antecedent is used. For the training, labeled examples of direct anaphora are only used in the *direct antecedent selection model* and labeled examples of indirect anaphora are only used in the *indirect antecedent selection model*.

The separate model approach is expected to be advantageous because useful information for detecting direct-anaphoric antecedents is different from one for indirect-anaphoric antecedents. For example, synonymous relations between anaphor and antecedent are important for selecting

Table 1 Summary of the information used in each anaphora type classifier.

Contextual Information	aC/S	cC/S	sS/C	dS/C	iS/C	pS/C
Use an anaphor?	✓	✓	✓	✓	✓	✓
Use all potential antecedents?		✓				
Use an antecedent selected by <i>single</i> -ASM?			✓			
Use an antecedent selected by <i>direct</i> -ASM?				✓		✓
Use an antecedent selected by <i>indirect</i> -ASM?					✓	✓

ASM indicates antecedent selection model. The aC/S, cC/S, sS/C, dS/C iS/C and pS/C denote anaphora type classification models described in Section 4.2.

direct-anaphoric antecedents. In example (1), an antecedent selection model has to know that *CD* and *album* are synonymous. For indirect anaphora, on the other hand, it is required to recognize such semantic relations as *part-whole* and *attribute-value* as shown in example (2), where it is essential that *CD* is semantically related with *song*.

There are a variety of existing machine learning-based methods designed for coreference resolution ranging from classification-based models (Soon et al. 2001, etc.) and preference-based models (Ng and Cardie 2001, etc.) to comparison-based models (Iida et al. 2005; Yang et al. 2003, etc.). Among them, we adopt a state-of-the-art model for coreference resolution in Japanese (Iida et al. 2005), called the *tournament model* because it achieved the best performance for coreference resolution in Japanese. The tournament model selects the best candidate antecedent by conducting one-on-one games in a step-ladder tournament. More specifically, the model conducts a tournament consisting of a series of games in which candidate antecedents compete with each other and selects the winner of the tournament as the best candidate antecedent. The model is trained with instances, each created from an antecedent paired with one other competing candidate.

## 4.2 Anaphora type classification

In this section, we elaborate *issue 2* and *issue 3* for anaphora type classification. An interesting question for this subtask is whether anaphora type classification should be carried out before antecedent selection or after because the available information differs depending on the order of those two subtasks. To reflect this, we consider two kinds of configurations: *Classify-then-Select* and *Select-then-Classify* as follows. The difference between the clues that each classifier uses is summarized in Table 1. The classifiers are trained in a supervised fashion.

**Classify-then-Select (C/S) model.** Given an anaphor, an anaphora type classifier first determines whether the anaphor bears either direct anaphora, indirect anaphora or exophora. If

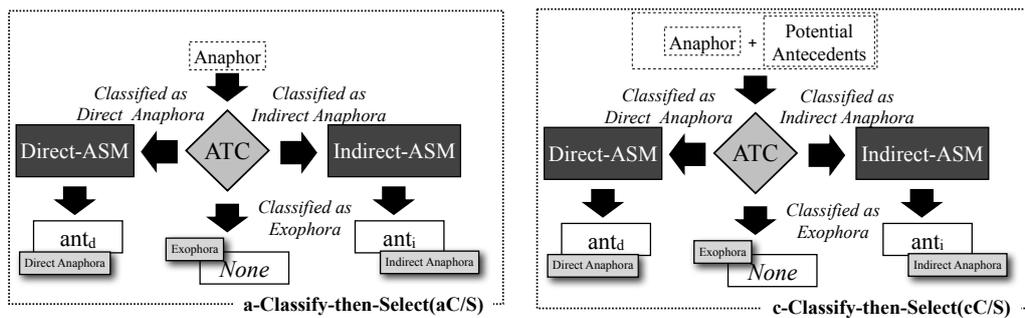


Fig. 1 Classify-then-Select Anaphora Resolution Models. *ASM* denotes Antecedent Selection Model and *ATC* denotes Anaphora Type Classifier.  $ant_d$  and  $ant_i$  denote an antecedent selected by Direct-ASM and Indirect-ASM respectively.

the anaphora type is judged as direct anaphora, then the direct antecedent selection model is called. If the anaphora type is judged as indirect anaphora, on the other hand, then the indirect antecedent selection model is called. There is no antecedent selection model called if exophora is selected.

By altering the choice of information used in anaphora type classification, the following two alternative models are available for the Classify-then-Select configuration, each of which is illustrated in Figure 1.

- *a-Classify-then-Select (aC/S) Model*: Classify anaphora type of a given anaphor by using the anaphor and its properties before selecting the antecedent.
- *c-Classify-then-Select (cC/S) Model*: Classify anaphora type of a given anaphor by using the anaphor, its properties and the lexical and syntactic information from all potential antecedents before selecting the antecedent.

By comparing the cC/S model with the aC/S model, we can see the effect of using contextual information in anaphora type classification. The feature set used in the models is detailed in Section 6.1.

**Select-then-Classify (S/C) model.** Given an anaphor, an antecedent selection model first selects the most likely antecedent and an anaphora type classifier determines the anaphora type by utilizing information from both the anaphor and the selected candidate antecedent(s). This way of configuration has an advantage over the Classify-then-Select models in that it determines the anaphora type of a given anaphor taking into account the information of its most likely candidate antecedent. The candidate antecedent selected in the first step can be expected to

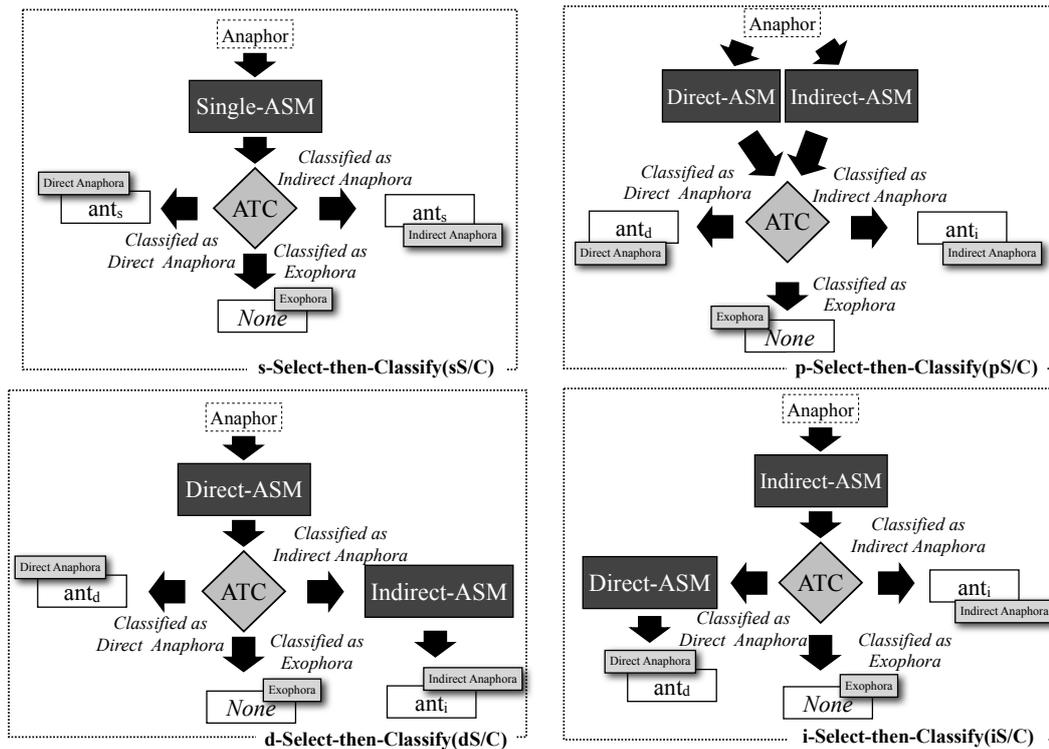


Fig. 2 Select-then-Classify Anaphora Resolution Models. *ASM* denotes Antecedent Selection Model and *ATC* denotes Anaphora Type Classifier.  $ant_s$ ,  $ant_d$  and  $ant_i$  denote an antecedent selected by Single-ASM, Direct-ASM and Indirect-ASM respectively.

provide contextual information useful for anaphora type classification: for example, if *her new song* is selected as the best candidate antecedent in example (6), the anaphora type will be easily identified by using the lexical knowledge that *CD* is the semantically related object of *song*.

(6) *The artist announced her new song. I want to get the CD as soon as possible.*

Since we have two choices of antecedent selection models (i.e., the single and separate models) as shown in 4.1, finally at least the following four models are available for anaphora type classification, each of which is illustrated in Figure 2.

- *s-Select-then-Classify (sS/C) Model*: Select the best candidate antecedent with the single model and then classify the anaphora type.

- *d-Select-then-Classify (dS/C) Model*: Select the best candidate antecedent by the direct anaphora model and then classify the anaphora type. If the candidate is classified as indirect anaphora, search for the antecedent with the indirect anaphora model.
- *i-Select-then-Classify (iS/C) Model*: Analogous to the d-Select-then-Classify (dS/C) model with the steps reversed.
- *p-Select-then-Classify (pS/C) Model*: Call the direct anaphora and indirect anaphora models in parallel to select the best candidate antecedent for each case and then classify the anaphora type referring to both candidates.

The pS/C configuration provides richer contextual information for classifying anaphora type than any other configuration because it can always refer to the most likely candidate antecedents of direct anaphora and indirect anaphora, which may be useful for determining anaphora type.

We adopt the *one-versus-rest* method for the three-way classification in our experiments. In other words, we recast the multi-class classification problem as combinations of a binary classification. Given an anaphor, each anaphora type classifier outputs a score that represents the likelihood of its anaphora type. According to these three scores, we select the anaphora type that achieves the maximum score.

The training procedure of each model depends on which kinds of information is needed. To exemplify how to create training instances, assume that we have the following text to create training instances.

(7) Mariah Carey<sub>(3')</sub> is an artist who comes from the USA<sub>(1)</sub>.

She announced her new song<sub>(2')</sub> yesterday.

I'm looking forward to hearing the new song<sub>(2)</sub>.

The beautiful voice<sub>(3)</sub> will attract me.

with annotated as the following:

the USA<sub>(1)</sub> :: exophora

the new song<sub>(2)</sub> – her new song<sub>(2')</sub> :: direct anaphora

the beautiful voice<sub>(3)</sub> – Mariah Carey<sub>(3')</sub> :: indirect anaphora

In the aC/S model, the information of an anaphor is needed to determine anaphora type. Since there are three instances in (7), the classifier is trained with the USA<sub>(1)</sub> as exophoric instance, the new song<sub>(2)</sub> as direct-anaphoric instance and the beautiful voice<sub>(3)</sub> as indirect-anaphoric instance. For the cC/S configuration, in addition to the anaphor information, the classifier takes all the potential antecedents. More specifically, the classifier is trained with the pair of the anaphor and its potential antecedents, hence, {the USA<sub>(1)</sub>, (Mariah Carey, an artist)} as exophoric instance, {the new song<sub>(1)</sub>, (Mariah Carey, an artist, the USA, she, her new

song, yesterday, I)} as direct-anaphoric instance and {the beautiful voice<sub>(3)</sub>, (Mariah Carey, an artist, the USA, she, her new song, yesterday, I, the new song)} as indirect-anaphoric instance<sup>4</sup>.

For the S/C configuration, we use the pair of an anaphor and annotated antecedent or *pseudo*-antecedent as a training instance. It depends on the anaphora type of interested anaphor and the type of antecedent selection model that the classifier utilizes to determine whether it is annotated antecedent or *pseudo*-antecedent. At first, in the sS/C configuration, the classifier selects *pseudo*-antecedent of the USA<sub>(1)</sub> using the single model since there is no annotated antecedent in the training set. Suppose an artist is selected; we create a training instance of exophora from the USA<sub>(1)</sub> paired with an artist. For direct-anaphoric and indirect-anaphoric instances, we simply take the anaphor and the annotated antecedent, i.e. ⟨the new song<sub>(2)</sub>, her new song<sub>(2′)</sub>⟩ and ⟨the beautiful voice<sub>(3)</sub>, Mariah Carey<sub>(3′)</sub>⟩ as each training instance.

Second, the case of the dS/C configuration is slightly more complex. Analogously to the sS/C model, for the USA<sub>(1)</sub>, we obtain the *pseudo*-antecedent using the direct antecedent selection model. Suppose an artist is selected; the pair ⟨the USA<sub>(1)</sub>, an artist⟩ is used as an exophoric instance. For the new song<sub>(2)</sub>, we use the pair ⟨the new song<sub>(2)</sub>, her new song<sub>(2′)</sub>⟩ as direct-anaphoric instance. For the beautiful voice<sub>(3)</sub>, however, since the annotated antecedent Mariah Carey<sub>(3′)</sub> is unlikely to be selected as the best candidate of the beautiful voice<sub>(3)</sub> by the *direct* antecedent selection model, the classifier does not use the annotated example. In dS/C anaphora type classification fashion, it is required to classify the beautiful voice<sub>(3)</sub> paired with the *pseudo*-best candidate selected by the *direct* antecedent selection model as the *indirect* anaphora. We therefore run the *direct* antecedent selection model to select the *pseudo*-best candidate. Suppose yesterday is selected; we create a training instance of indirect anaphora from the beautiful voice<sub>(3)</sub> paired with yesterday.

An analogous method applies also to the iS/C configuration; that is, we run the *indirect* antecedent selection model to select the *pseudo*-best candidate except for the indirect-anaphoric instance the beautiful voice<sub>(3)</sub>. We assume that Mariah Carey is selected as the *pseudo*-best candidate for the USA<sub>(1)</sub> and an artist for the new song<sub>(2)</sub>. We create the following training instances: ⟨the USA<sub>(1)</sub>, Mariah Carey⟩ as exophora, ⟨the new song<sub>(2)</sub> an artist⟩ as direct anaphora, ⟨the beautiful voice<sub>(3)</sub>, Mariah Carey<sub>(3′)</sub>⟩ as indirect anaphora.

---

<sup>4</sup>We enumerated only noun phrases as the potential antecedents for convenience. In our evaluations, we include verbal predicates in the list of potential antecedents for such cases as ...we calculate the value in advance. – The precomputation ....

Table 2 Distribution of anaphoric relations in the annotated corpus

Syntax	Broadcast				Editorial			
	Direct	Indirect	Exophora	Ambiguous	Direct	Indirect	Exophora	Ambiguous
Noun	530	466	-	0	550	561	-	0
Predicate	70	435	-	8	114	883	-	2
Overall	600	901	248	8	664	1,444	222	2

‘Noun’ and ‘Predicate’ denote the syntactic category of an antecedent. ‘Ambiguous’ was annotated to an anaphor which holds both direct and indirect anaphoric relations. In our evaluations, we discarded such instances.

Finally, in the pS/C configuration, we need *triplets* (an anaphor, a direct-anaphoric antecedent, an indirect-anaphoric antecedent). For **the USA<sub>(1)</sub>**, since we have no annotated antecedents for direct and indirect anaphora, we run both the direct and indirect antecedent selection model to select *pseudo*-best candidates of **the USA<sub>(1)</sub>**. Supposing that **an artist** is selected by the *direct* antecedent selection model and **Mariah Carey** is selected by the *indirect* model, we create a training instance of exophora from (the USA<sub>(1)</sub>, an artist, Mariah Carey). For **the new song<sub>(2)</sub>**, since we have no annotated antecedent for *indirect* anaphora, the indirect antecedent selection model is chosen to select the *pseudo*-best candidate. Suppose **an artist** is selected; we create a direct-anaphoric training instance (the new song<sub>(2)</sub>, her new song<sub>(2')</sub>, an artist). For **the beautiful voice<sub>(3)</sub>**, analogous to **the new song<sub>(2)</sub>**, supposing **yesterday** is selected by the direct antecedent selection model, we create an indirect-anaphoric training instance (the beautiful voice<sub>(3)</sub>, yesterday, Mariah Carey<sub>(3')</sub>).

## 5 Dataset

For training and testing our models, we created an annotated corpus that contains 2,929 newspaper articles consisting of 19,669 sentences for 2,320 broadcasts, 18,714 sentences for 609 editorials, which is the same articles as in the NAIST Text Corpus (Iida et al. 2007). The NAIST Text Corpus also contains anaphoric relations of noun phrases, but they are strictly restricted as coreference relations (i.e. two NPs must refer to the same entity in the world). For this reason, most NPs marked with a definiteness modifier that we need are not annotated even when two NPs have a direct-anaphoric relation. Therefore, we re-annotated (i) direct anaphoric relations, (ii) indirect anaphoric relations and (iii) exophoric noun phrases of noun phrases marked by one of the three definiteness modifiers, that is *this* (この), *the* (その), and *that* (あの). In the specification of our corpus, not only noun phrases but verb phrases are chosen as antecedents. For example,

the verbal predicate *calculates in advance* is selected as an antecedent of *the precomputation* in example (8).

- (8) システムは前もって値を 計算する<sub>(i')</sub>。 その前計算<sub>(i)</sub> はシステムの性能を大幅に向上させている。

*The system calculates<sub>(i')</sub> the value in advance. The precomputation<sub>(i) significantly improves its performance.</sub>*

We also annotated anaphoric relations in the case where an anaphor is anaphoric with more than two antecedents. For example, we label anaphoric relations for the two pairs of NPs *mouse devices*–*the other items* and *keyboards*–*the other items* as seen in example (9).

- (9) ABC コンピュータは マウス<sub>(i')</sub> と キーボード<sub>(j')</sub> の値下げを発表した。 その他の商品<sub>(i,j)</sub> については値下げをしないと主張した。

*ABC computer announced that they reduced the price of mouse devices<sub>(i')</sub> and keyboards<sub>(j')</sub>. They claimed that they would not cut the price of the other items<sub>(i,j)</sub>.*

Finally, we obtained 1,264 instances of direct anaphora, 2,345 instances of indirect anaphora, and 470 instances of exophora. The detailed statistics are shown in Table 2. To assess the reliability of the annotation, we estimated its agreement rate with the two annotators from 418 examples<sup>5</sup> in terms of *K* statistics (Sidney and Castellan 1988). It resulted in  $K = 0.73$ , which indicates good reliability. For measuring the agreement ratio of antecedent selection, we used 322 examples (109 for direct anaphora and 213 for indirect anaphora) whose anaphora types are identically identified by both two annotators. The agreement ratio was calculated<sup>6</sup> according to the following equation:

$$\text{Agreement} = \frac{\# \text{ of instances which both two annotators identified the same antecedent}}{\# \text{ of all instances}}.$$

The agreement ratio for annotating direct-anaphoric relation obtained 80.7% (88/109). However, for 21 examples whose antecedents are not identically selected by the annotators, our analysis revealed that 52.4% (11/21) of these examples are cases where the antecedents annotated by the two annotators are different but in anaphoric relation, which should be regarded as an agreement. Therefore, the inter-annotator agreement ratio of direct-anaphoric relation achieves 90.8% (99/109), which indicates good reliability but it is required to consider anaphoric chains in the annotation procedure. The agreement ratio of indirect-anaphoric relation, on the other hand, obtained a comparatively lower ratio of 62.9% (134/213). One of the typically non-matching

<sup>5</sup>These examples are randomly sampled from our corpus, and account for 10% of all the examples.

<sup>6</sup>We regarded the matching of the rightmost offset as the agreement. When multiple antecedents are annotated, the criterion of matching is that one of the antecedents is at least identical with one of the antecedents annotated by the other annotator.

cases is shown in example (10).

- (10) 政府<sub>(i)</sub> は明日までに 委員<sub>(j)</sub> を決める方針だ。その人選<sub>(k)</sub> は我々にも影響が及ぶだろう。  
*The government<sub>(i)</sub> is going to determine the member of the committee<sub>(j)</sub> by tomorrow.*  
*Probably the election<sub>(k)</sub> will also affect us.*

In this example, both *the government* and *the member of the committee* are considered to be associated objects of *the election*, which indicates that multiple discourse elements are often associated with one anaphor in various semantic relations in indirect anaphora. We should reflect on such problems when the annotation scheme and task definition of indirect anaphora resolution are argued, including bridging reference resolution.

## 6 Evaluation

We conduct empirical evaluations in order to investigate the three issues shown in Section 1. First, we compare two antecedent selection models, the single and separate models described in Section 4.1 in order to find out *issue 1*, i.e., whether an antecedent selection model should be trained separately for direct anaphora and indirect anaphora. Second, the anaphora type classification models described in Section 4.2 are evaluated to explore what information helps with the anaphora type classification (*issue 2*). Finally, we evaluate the overall accuracy of the entire anaphora resolution task to explore how the models can be best configured (*issue 3*).

In our experiments, we used anaphors whose antecedent is a head of NP that appears in the preceding context of the anaphor (i.e., cataphora is ignored), only taking articles in the broadcast domain into account. Therefore, we used 572 instances of direct anaphora, 878 instances of indirect anaphora and 248 instances of exophora. The evaluation was carried out by 10-fold cross-validation. In our evaluation of antecedent selection, if a selected antecedent is in the same direct-anaphoric chain as the labeled antecedent, this selected antecedent is evaluated as correct<sup>7</sup>.

For creating binary classifiers used in antecedent selection and anaphora type classification, we adopted Support Vector Machines (Vapnik 1995)<sup>8</sup>, with a polynomial kernel of degree 2 and its default parameters.

### 6.1 Feature set

<sup>7</sup>We manually checked our results because of the lack of annotation of anaphoric chains as noted in Section 5. Due to the cost of this manual checking, we took only the broadcast articles into account in our experiments, leaving the editorials out.

<sup>8</sup>*SVM<sup>light</sup>* <http://svmlight.joachims.org/>

Table 3 Feature set for antecedent selection and the S/C models.

Feature	Description
DEFINITIVE	1 if $C_p$ is definite noun phrase; else 0.
DEPEND_CLASS*	$POS \in \{NOUN, PREDICATE\}$ * of word which $C_p$ depends.
DEPENDED_CLASS*	$POS \in \{NOUN, PREDICATE\}$ * of word depending $C_p$ .
ANAPHOR_DM_TYPE	Type of definiteness modifier of $ANA$ .
ANAPHOR_HEAD	Head morpheme of $ANA$ .
ANAPHOR_POS	POS of $ANA$ .
ANAPHOR_CASE	Case particle of $ANA$ .
CANDIDATE_HEAD	Head morpheme of $C_p$ .
CANDIDATE_POS	POS of $C_p$ .
CANDIDATE_NE	Proper noun-type of $C_p$ .
CANDIDATE_CASE	Case particle of $C_p$ .
CANDIDATE_BGH_ID*	The semantic class ID of $C_p$ at the level of a middle grain size defined in Bunrui Goi Hyo.
WN_SEMANTIC_RELATION	The semantic relation between $ANA$ and $C_p$ found in WordNet.
STRING_MATCH_TYPE*	The string match type $\in \{HEAD, PART, COMPLETE\}$ if the string of $C_p$ matches the string of $ANA$ ; else empty.
SENTENCE_DISTANCE	The number of sentences intervening between $C_p$ and $ANA$
SIMILARITY*	Distributional similarity between $ANA$ and $C_p$
PMI**	Point-wise mutual information between $ANA$ and $C_p$
BGH_COMMON_ANCE*	The depth of lowest common ancestor of $C_p$ and $ANA$ in BGH
SYNONYMOUS	1 if $C_p$ and $ANA$ are synonymous; else 0.
IS_HYPONYM_OF_ANAPHOR	1 if $C_p$ is a hyponym of $ANA$ ; else 0.
DEPEND_RELATION	Function word when $C_L$ depends on $C_R$ if $C_L$ depends on $C_R$ ; else empty.
SENTENCE_DISTANCE	The number of sentences intervening between $C_L$ and $C_R$
DEPENDED_COUNT_DIFF*	Difference between the count of bunsetsus depending $C_L$ and $C_R$ .

$ANA$  denotes an anaphor.  $C_{p \in \{L, R\}}$  denotes either of the two compared candidate antecedents ( $C_L$  and  $C_R$  denote the left and right candidate, respectively). ‘\*’ denotes features used only in the direct antecedent selection model (ASM), the single model or the dS/C model, and ‘\*\*’ only in the indirect-ASM, the single model or the iS/C model. In the pS/C model, the feature set extracted from *direct*-ASM is distinguished from the one extracted from *indirect*-ASM.

The feature set for antecedent selection is designed based on the literature of coreference resolution (Iida et al. 2005; Ng and Cardie 2001; Soon et al. 2001; Denis and Baldridge 2008; Yang et al. 2003, etc) as summarized in Table 3. In addition, we introduce the following lexical

Table 4 Feature set for the C/S models

Feature	Description
ANAPHOR_DM_TYPE	Type of definiteness modifier of <i>ANA</i> .
ANAPHOR_HEAD	Head morpheme of <i>ANA</i> .
ANAPHOR_POS	POS of <i>ANA</i> .
ANAPHOR_CASE	Case particle of <i>ANA</i> .
HOLDING_POS*	POS of all the candidates in the preceding sentences.
HAS_SYNONYM_OF_ANAPHOR*	1 if there exists a synonym of <i>ANA</i> in the preceding sentences; else 0.
HAS_HYPONYM_OF_ANAPHOR*	1 if there exists a hyponym of <i>ANA</i> in the preceding sentences; else 0.
HAS_STRING_MATCHED*	1 if there exists NP whose string matches the last string of (head of) <i>ANA</i> in the preceding sentences; else 0.
MAX_PMI*	Maximum PMI between <i>ANA</i> and each candidates in the preceding sentences.
MAX_NOUN_SIM*	Maximum noun-noun similarity between <i>ANA</i> and each candidates in the preceding sentences.

*ANA* denotes an anaphor. ‘\*’ denotes the features that capture the contextual information, which is only used for cC/S model. S/C models use the feature set of related antecedent selection model described in Table 3.

semantic features:

- **WN\_SEMANTIC\_RELATION**: In order to capture various semantic relations between an anaphor and its antecedent, we incorporate the binary features that represent the semantic relation found in the Japanese WordNet 0.9 (Isahara et al. 2008)<sup>9</sup>.
- **SYNONYMOUS** and **IS\_HYPONYM\_OF\_ANAPHOR**: We recognize synonymous and hyper-hyponym relations by using a very large amount of synonym and hypernym-hyponym relations (about three million hypernymy relations and two hundred thousand synonymy relations) automatically created from Web texts and Wikipedia (Sumida et al. 2008).
- **BGH\_ID**, **BGH\_COMMON\_ANCESTOR**: We incorporate the lexical information obtained from the *Bunrui Goi Hyo* thesaurus (NLRI 1964). We encode the information as two types: (i) binary features that represent the semantic class ID, and (ii) a real-valued feature that indicates the depth of the lowest common ancestor of an anaphor and its candidate.
- **SIMILARITY**: To robustly estimate semantic similarities between an anaphor and its candidate antecedent, we adopt the cosine similarity between an anaphor and candidate

<sup>9</sup><http://nlpwww.nict.go.jp/wn-ja/>

antecedent, which is calculated from a cooccurrence matrix of  $(n, \langle c, v \rangle)$ , where  $n$  is a noun phrase appearing in an argument position of a verb  $v$  marked by a case particle  $c$ . The cooccurrences are counted from two decades worth of news paper articles, and their distribution  $P(n, \langle c, v \rangle)$  is estimated by pLSI (Hofmann 1999) with 1,000 hidden topic classes to overcome the data sparseness problem.

- PMI: The degree of indirect-anaphoric association between an anaphor *ANA* and candidate *CND* is calculated differently depending on whether *CND* is a noun or predicate. For the case of a noun, we follow the literature of indirect anaphora resolution (Poesio et al. 2004; Murata et al. 1999, etc.) to capture such semantic relations as *part-whole*. The associativeness is calculated from the cooccurrences of *ANA* and *CND* in the pattern of “*CND* の *ANA* (*ANA* of *CND*)”. Frequencies of cooccurrence counts are obtained from the Web Japanese N-gram Version 1 (Kudo and Kazawa 2007). For the case of a predicate, on the other hand, the associativeness is calculated from the cooccurrences of *ANA* and *CND* in the pattern where *CND* syntactically depends on (i.e. modifies) *ANA* (in English, the pattern like “*ANA* that (*subj*) *CND*”). If we find many occurrences of, for example, “闘う (*to fight*)” modifying “夢 (*a dream*)” in a corpus, then “夢 (*a dream*)” is likely to refer to an event referred to by “闘う (*to fight*)” as in (11).

(11) チャンピオンと闘い<sub>(i')</sub>たい。その夢<sub>(i)</sub>は実現すると信じている。

*I want to fight<sub>(i')</sub> the champion. I believe the dream<sub>(i)</sub> will come true.*

For anaphora type classification, we use a different feature set depending on the configuration described in 4.2. For the Classify-then-Select configuration, as summarized in Table 4, it includes such features as HAS\_SYNONYM\_OF\_ANAPHOR and HAS\_STRING\_MATCHED, which capture contextual information encoded from all potential antecedents, based on the literature (Vieira and Poesio 2000, etc.). For the Select-then-Classify configurations, on the other hand, an anaphora type classifier uses the best candidate(s) selected in antecedent selection phase as its contextual information, instead of the information encoded from all the potential antecedents. This sort of information is encoded as features analogous to that for antecedent selection as summarized in Table 3.

## 6.2 Results of antecedent selection

The results of antecedent selection are shown in Table 5. The results<sup>10</sup> indicate that the Separate Model outperforms the Single Model on two anaphora types. As for *issue 1*, we conclude

<sup>10</sup>The accuracy of the separate model is better than the single model with statistical significance ( $p < 0.01$ , McNemar test).

Table 5 Results of antecedent selection

Anaphora Type	Single Model	Separate Model
Direct anaphora	63.3% (362/572)	65.4% (374/572)
Indirect anaphora	50.5% (443/878)	53.2% (467/878)
Overall	55.2% (801/1,450)	<b>58.0% (841/1,450)</b>

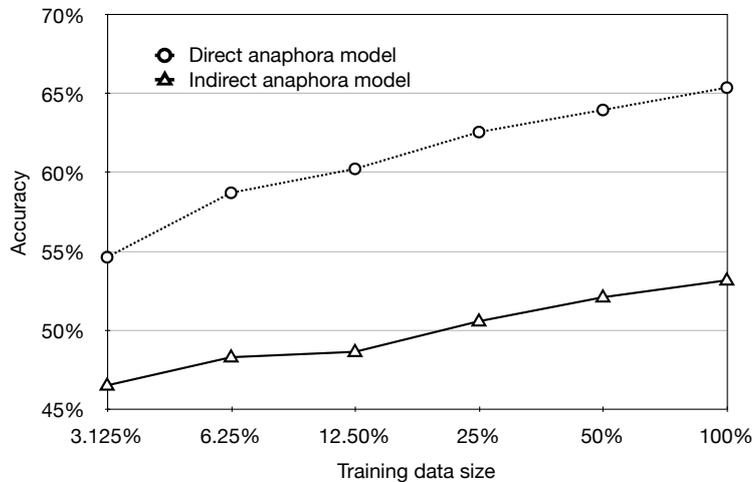


Fig. 3 Learning curve for Separate Models

that the information used for antecedent selection should be separated for each anaphora type and the selection models should be trained for each anaphora type. We therefore discard the Single Model for the further experiments (i.e. discarding sS/C model).

We also illustrate the learning curves of each model, shown in Figure 3. Reducing the training data to 50%, 25%, 12.5%, 6.25% and 3.13%, we conducted the evaluation over three random trials for each size and averaged the accuracies. Figure 3 indicates that in the direct antecedent selection model the accuracy becomes better as the training data increase, whereas the increase of the indirect one looks difficult to improve although our data set included more instances for indirect anaphora than for the direct one. These results support the finding in previous work that an indirect anaphora is harder to resolve than direct anaphora and suggest that we need a more sophisticated antecedent selection model for indirect anaphora.

Our error analysis revealed that a majority (about 60%) of errors in direct anaphora were caused by the fact that both correct and incorrect candidates belong to the same semantic category. Example (12) shows a typical selection error:

- (12) 私は映画<sub>(j)</sub>の知識がないが、『フランケンシュタイン』<sub>(i')</sub>ぐらいは知っている。  
この映画<sub>(i)</sub>は、本当に名作だ。

Table 6 Results of anaphora type classification

Model	Direct Anaphora			Indirect Anaphora			Exophora			Accuracy
	P	R	F	P	R	F	P	R	F	
aC/S	67.7%	74.5%	70.9%	80.6%	87.1%	83.7%	75.0%	36.3%	48.9%	75.4%
cC/S	69.4%	73.4%	71.4%	74.9%	87.5%	80.7%	92.5%	25.0%	39.4%	73.6%
dS/C	70.9%	84.6%	<b>77.1%</b>	83.2%	85.6%	<b>84.4%</b>	90.1%	40.3%	55.7%	<b>78.7%</b>
iS/C	67.7%	74.8%	71.1%	78.1%	88.3%	82.9%	93.2%	27.8%	42.9%	74.9%
pS/C	71.2%	82.0%	76.1%	82.1%	86.7%	84.3%	91.9%	41.1%	<b>57.2%</b>	78.4%

*I don't have good knowledge of movies<sub>(j)</sub> but still know of "Frankenstein"<sub>(i)</sub>.*

*I think this movie<sub>(i)</sub> is indeed a great masterpiece.*

where the wrong candidate “映画<sub>(j)</sub> (*movies*<sub>(j)</sub>)” was selected as the antecedent of “この映画<sub>(i)</sub> (*this movie*<sub>(i)</sub>)”.<sup>11</sup> As can be imagined from this example, there is still room for improvement by carefully taking into account this kind of error using other clues such as information from salience. For indirect anaphora, we analyzed our resource to capture the associativeness between an anaphor and its antecedent, encoded as PMI in the feature set. Our analysis indicated that about half of the pattern ‘ANT of ANA’, which occurred in the test data, had been assigned a minus value, i.e., no positive association found between an anaphor and its antecedent for the resource when applying PMI. To evaluate the contribution to our model, we conducted an evaluation where the PMI feature set was disabled. As a result of this additional evaluation, the model obtained 51.4% (451/878), which is no significant difference compared with the original accuracy. We need to find more useful clues to capture the associativeness between an anaphor and the related object in indirect anaphora. The low quality of our annotating data of indirect-anaphoric relation, as mentioned in Section 5, might be also one of the reasons for the low accuracy of indirect anaphora resolution.

### 6.3 Results of anaphora type classification

Now, we move on to *issue 2* and *issue 3*. The results of anaphora type classification are shown in Table 6. The cC/S model obtained the lowest accuracy of 73.6%, which indicates that contextual information features proposed in the literature (Vieira and Poesio 2000, etc.), such as HAS.STRING.MATCHED, were not actually informative. Note that the performance of the cC/S model is lower than the aC/S model<sup>12</sup>, which identifies an anaphora type by using only the information of an anaphor. On the other hand, the dS/C model successfully improved

<sup>11</sup>In Japanese, the plural form of a noun is not morphologically distinguished from its singular form.

<sup>12</sup>The difference is statistically significant ( $p < 0.06$ , McNemar test).

Table 7 The majority of misclassified-exophoric instances

NP of an anaphor	Occurrences in our corpus		
	Direct anaphora	Indirect anaphora	Exophora
年 ( <i>year</i> )	42.9% (9/21)	9.5% (2/21)	47.6% (10/21)
日 ( <i>day</i> )	68.3% (82/120)	0.9% (1/120)	30.8% (37/120)
時 ( <i>time</i> )	8.9% (5/56)	82.1% (46/56)	8.9% (5/56)
時期 ( <i>period</i> )	25.0% (5/20)	35.0% (7/20)	40.0% (8/20)

its performance by using the information of selected candidate antecedent as the contextual information. The dS/C model achieved the best accuracy of 78.7%, which indicates that the selected best candidate antecedent provides useful contextual information for anaphora type classification<sup>13</sup>. The iS/C and pS/C models, however, do not improve their performance as well as the dS/C model although it uses the selected best candidate(s) information. It is considered that the fundamental reason is the poor performance of the *indirect* antecedent selection model as shown in Table 5, i.e., the *indirect* antecedent selection model does not provide correct contextual information to anaphora type classification. It is expected that all the S/C models get better performance when the antecedent selection model improves.

The identification of exophora is a more difficult task than the other anaphora types as shown in the low F-measure and recall in Table 6. Our analysis for the exophoric instances misclassified by the dS/C model revealed that the typical errors were temporal expressions such as 年 (*year*), 日 (*day*) and 時期 (*period*). We observed that such expressions occurred as not only exophora but also as the other anaphora types many times, as summarized in Table 7, which indicates that the interpretation of temporal expression is also important for identifying the other anaphora types. In our current framework, however, it is hard to recognize such expressions accurately since the precise recognition of temporal expressions is required to identify a relation between an event specified by the expression and the other events. We consider integrating the framework of temporal relation identification, which has been proposed in the evaluation-oriented studies such as TempEval<sup>14</sup>, with anaphora type classification framework, which will be our future work.

## 6.4 Results of overall anaphora resolution

Finally, we evaluated the overall accuracy of the entire anaphora resolution task given by:

$$Accuracy = \frac{\# \text{ of instances whose antecedent and anaphora type is identified correctly}}{\# \text{ of all instances}}.$$

<sup>13</sup>The dS/C model outperformed the aC/S, cC/S models with statistical significance using  $p < 0.03$ ,  $p < 0.01$ , as McNemar test parameters respectively.

<sup>14</sup><http://www.timeml.org/tempeval/>

Table 8 Overall results of anaphora resolution

Model	Accuracy
a-Classify-then-Select	47.3% (803/1,698)
c-Classify-then-Select	46.3% (787/1,698)
d-Select-then-Classify	<b>50.6% (859/1,698)</b>
i-Select-then-Classify	46.3% (787/1,648)
p-Select-then-Classify	50.4% (855/1,698)

The results are shown in Table 8. Again, the dS/C model achieved the best accuracy, which is significantly better than the Classify-then-Select models.

## 7 Conclusion

We have addressed the three issues of nominal anaphora resolution for Japanese NPs marked by a definiteness modifier under two subtasks, i.e., *antecedent selection* and *anaphora type classification*. The issues we addressed were: (i) how the antecedent selection model should be designed, (ii) what information helps anaphora type classification, and (iii) how the antecedent selection and anaphora type classification should be carried out. Our empirical evaluations showed that the separate model achieved better accuracy than the single model, and the d-Select-then-Classify and p-Select-then-Classify models give the best results. We have made several findings through the evaluations: (i) an antecedent selection model should be trained separately for each anaphora type using the information useful for identifying its antecedent, (ii) the best candidate antecedent selected by an antecedent selection model provides contextual information useful for anaphora type classification, and (iii) the antecedent selection should be carried out before anaphora type classification.

However, there is still considerable room for improvement in both subtasks. Our error analysis for antecedent selection reveals that the wrong antecedent, which belongs to the same semantic category as correct antecedent, is likely to be selected while selecting direct-anaphoric antecedent, and the association measure of indirect-anaphoric relatedness does not contribute to selecting the indirect-anaphoric antecedent. We will incorporate the information that captures salience and various noun-noun relatedness into antecedent selection in future work. For anaphora type classification, our analysis reveals that temporal expressions typically cause error in the identification of exophora. To recognize such expressions precisely, we will consider integrating temporal relation identification with anaphora type classification. Our future work also includes taking general noun phrases into account in anaphora resolution.

## Acknowledgment

I would like to thank the reviewers of this paper for their helpful comments.

## Reference

- Baldwin, F. B. (1995). *Cogniac: a discourse processing engine*. Ph.D. thesis.
- Brennan, S. E., Friedman, M. W., and Pollard, C. J. (1987). “A centering approach to pronouns.” In *Proceedings of the 25th annual meeting on Association for Computational Linguistics*, pp. 155–162.
- Bunescu, R. (2003). “Associative anaphora resolution: A web-based approach.” In *Proceedings of the EACL 2003 Workshop on The Computational Treatment of Anaphora*, pp. 47–52.
- Clark, H. H. (1977). “Bridging.” In *Thinking: Readings in Cognitive Science*.
- Denis, P. and Baldridge, J. (2008). “Specialized Models and Ranking for Coreference Resolution.” In *Proceedings of Empirical Methods in Natural Language Processing*, pp. 660–669 Honolulu, Hawaii. Association for Computational Linguistics.
- Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. MIT Press.
- Ge, N., Hale, J., and Charniak, E. (1998). “A statistical approach to anaphora resolution.” In *Proceedings of the 6th Workshop on Very Large Corpora*, pp. 161–170.
- Grosz, B. J., Weinstein, S., and Joshi, A. K. (1995). “Centering: a framework for modeling the local coherence of discourse.” *Computational Linguistics*, **21** (2), pp. 203–226.
- Hawkins, J. A. (1978). *Definiteness and Indefiniteness: A Study in Reference and Grammaticality Prediction*. Croom Helm Linguistic Series. Taylor & Francis.
- Hofmann, T. (1999). “Probabilistic Latent Semantic Indexing.” In *Proceedings of the Twenty Second Annual International SIGIR Conference on Research and Development in Information Retrieval*, pp. 50–57.
- Iida, R., Inui, K., Matsumoto, Y., and Sekine, S. (2005). “Noun Phrase Coreference Resolution in Japanese based on Most Likely Candidate Antecedents Noun Phrase Coreference Resolution in Japanese Based on Most Likely Antecedent Candidates.” In *Journal of Information Processing Society of Japan*, pp. 831–844.
- 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 ACL 2007 Workshop on Linguistic Annotation Workshop*, pp. 132–139.
- Isahara, H., Bond, F., Uchimoto, K., Utiyama, M., and Kanzaki, K. (2008). “Development of

- the Japanese WordNet.” In *Proceedings of the 6th International Language Resources and Evaluation*.
- Kudo, T. and Kazawa, H. (2007). *Web Japanese N-gram Version 1*. Gengo Shigen Kyokai.
- Mccarthy, J. F. and Lehnert, W. (1995). “Using decision trees for coreference resolution.” In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pp. 1050–1555.
- 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 36th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the Association for Computational Linguistics Workshop on Operational Factors in Practical*.
- Mitkov, R., Evans, R., Orasan, C., Barbu, C., Jones, L., and Sotirova, V. (2000). “Coreference and anaphora: developing annotating tools, annotated resources and annotation strategies.” In *Proceedings of the Discourse Anaphora and Anaphora Resolution Colloquium*.
- Murata, M., Isahara, H., and Nagao, M. (1999). “Resolution of Indirect Anaphora in Japanese Sentences Using Examples “XnoY(Yof X)”.” In *Proceedings of the ACL 1999 Workshop on Coreference and Its Applications*.
- Nakaiwa, H., Shirai, S., Ikehara, S., and Kawaoka, T. (1995a). “Extrasentential Resolution of Japanese Zero Pronouns using Semantic and Pragmatic Constraints.” In *Proceedings of AAAI 1995 Spring Symposium Series, Empirical Methods in Discourse Interpretation and Generation*. AAAI.
- Nakaiwa, H., Shirai, S., Ikehara, S., and Kawaoka, T. (1995b). “Extrasentential Resolution of Japanese Zero Pronouns using Semantic and Pragmatic Constraints.” In *Proc. of AAAI 1995 Spring Symposium Series, Empirical Methods in Discourse Interpretation and Generation*. AAAI.
- Ng, V. and Cardie, C. (2001). “Improving machine learning approaches to coreference resolution.” In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 104–111.
- NLRI (Ed.) (1964). *Bunrui Goi Hyo (in Japanese)*. Shuei Shuppan.
- Okumura, M. and Tamura, K. (1996). “Zero pronoun resolution in Japanese discourse based on centering theory.” In *Proceedings of the 16th conference on Computational linguistics*, pp. 871–876.
- Poesio, M., Ishikawa, T., Walde, S. S. I., Vieira, R., and Sinos, R. (2002). “Acquiring lexical knowledge for anaphora resolution.” In *Proceedings of the 3rd Conference on Language Resources and Evaluation*, pp. 1220–1224.

- Poesio, M., Mehta, R., Maroudas, A., and Hitzeman, J. (2004). “Learning to resolve bridging references.” In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, pp. 144–151.
- Poesio, M., Vieira, R., and Teufel, S. (1997). “Resolving bridging references in unrestricted text.” In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, pp. 1–6.
- Shalom, L. and J., L. H. (1994). “An algorithm for pronominal anaphora resolution.” *Computational Linguistics*, **20** (4), pp. 535–561.
- Sidney, S. and Castellan, N. J. (1988). *Nonparametric statistics for the Behavioral Sciences*. McGraw Hill.
- Soon, W. M., Ng, H. T., and Lim, C. Y. (2001). “A Machine Learning Approach to Coreference Resolution of Noun Phrases.” *Computational Linguistics*, **27** (4), pp. 521–544.
- Strube, M. and Muller, C. (2003). “A machine learning approach to pronoun resolution in spoken dialogue.” In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pp. 168–175.
- Sumida, A., Yoshinaga, N., and Torisawa, K. (2008). “Boosting Precision and Recall of Hyponymy Relation Acquisition from Hierarchical Layouts in Wikipedia.” In *Proceedings of the 6th Language Resources and Evaluation Conference*.
- Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Wiley.
- Vieira, R. and Poesio, M. (2000). “An Empirically Based System for Processing Definite Descriptions.” *Computational Linguistics*, **26** (4), pp. 539–593.
- Yang, X., Zhou, G., Su, J., and Tan, C. L. (2003). “Coreference resolution using competition learning approach.” In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pp. 176–183.

**Naoya Inoue:** He received his bachelor’s in economics from Musashi University in 2008. He is currently a student in the Graduate School of Information Science at Nara Institute of Science and Technology. His research interests are in anaphora resolution, discourse understanding and spoken dialogue systems.

**Ryu Iida:** He received his master and Ph.D degrees of engineering from Nara Institute of Science and Technology in 2004 and in 2007. He is currently an assistant professor in the School of Information Science and Engineering at Tokyo Institute of Technology. His research interests include anaphora resolution and

information extraction.

**Kentaro Inui:** He received his doctoral degree of engineering from Tokyo Institute of Technology in 1995. Having experienced Assistant Professor at Tokyo Institute of Technology and Associate Professor at Kyushu Institute of Technology, he has been Associate Professor of Graduate School of Information Science at Nara Institute of Science and Technology since 2002. His research interests include natural language understanding and knowledge engineering.

**Yuji Matsumoto:** He is currently a Professor of Information Science, Nara Institute of Science and Technology. He received his M.S. and Ph.D. degrees in information science from Kyoto University in 1979 and in 1989. He joined Machine Inference Section of Electrotechnical Laboratory in 1979. He has then experienced an academic visitor at Imperial College of Science and Technology, a deputy chief of First Laboratory at ICOT, and an associate professor at Kyoto University. His main research interests are natural language understanding and machine learning.