Resolving Direct and Indirect Anaphora for Japanese Definite Noun Phrases

Naoya Inoue
NAIST, Japan
naoya-i@is.naist.jp

Ryu Iida
Tokyo Institute of Technology, Japan
ryu-i@cl.cs.titech.ac.jp

Kentaro Inui
NAIST, Japan
inui@is.naist.jp

Yuji Matsumoto
NAIST, Japan
matsu@is.naist.jp

Abstract

An anaphoric relation can be either direct or indirect. In some cases, the antecedent being referred to lies outside the discourse to which its anaphor belongs. Therefore, an anaphora resolution model needs to consider the following two decisions in parallel: (i) antecedent selection–selecting the antecedent itself, and (ii) anaphora type classification–classifying exophora, direct anaphora, or indirect anaphora. However, the anaphora type classification has received little attention in the literature. In this paper, we address this issue taking Japanese as our target language. Our findings were: first, an antecedent selection model should be trained separately for each anaphora type. Second, the best candidate antecedent selected by an antecedent selection model provides contextual information useful for anaphora type classification. In consequence, antecedent selection should be carried out before anaphora type classification.

1 Introduction

Anaphora resolution has been studied intensively in recent years because of its significance in many NLP applications such as information extraction and machine translation. In nominal anaphora resolution, an anaphor (typically a definite noun phrase) and its antecedent hold either a direct anaphoric relation (e.g., coreference) or an indirect relation (e.g., bridging reference (Clark, 1977)). In example (1), the anaphor *the CD* refers to *Her new album* directly.

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

In contrast, *the CD* refers to *her new song* indirectly in example (2).

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

In cases such as example (3), an antecedent referred to is not in the same discourse.

(3) *I want to get the CD as soon as possible.*

We call the reference in example (1) *direct anaphora*; in example (2) *indirect anaphora*; and in example (3) *exophora*. In anaphora resolution, besides identifying the antecedent itself, it is therefore necessary to consider these three possibilities in parallel and classify the anaphora type. However, these issues have received little attention in the literature.

In this paper, we regard 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 input definite NP, and
- *Anaphora type classification*: the task of judging whether the anaphor is direct anaphoric, indirect anaphoric or exophoric.

Given this task decomposition, the following unexplored issues immediately arise:

- Whether the model for antecedent selection should be designed and trained separately for direct anaphora and indirect anaphora or if it can be trained as a single common model,
- How the two subtasks can be best combined (e.g. which subtask should be carried out first), and
- Which context information is useful for anaphora type classification.

In this paper, we explore these issues taking Japanese as our target language.

Definite NPs in Japanese 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 (4) and (5).

(4) 新型車 が発売された。このミニバン は燃費が非常によい。
A new car was released. The minivan has good gas mileage.

(5) 家具屋で 机 を見た。そのデザイン は素晴らしかった。
I saw a desk in a furniture shop. The design was marvelous.

"このミニバン (*the minivan*)" refers to "新型車 (*a new car*)" directly in (4), while "そのデザイン (*the design*)" refers to "机 (*a desk*)" indirectly in (5). In Japanese, however, the problem can be even more complex because a definite NP is not always marked by a definiteness modifier. In an example (6), a bare NP 大統領 (literally, *President*) refers to 韓国大統領(*Korean President*).

(6) 今月 4 日、韓国大統領 が来日した。大統領 は翌日の記者会見で新プランの詳細を語った。
Korean President visited Japan on the 4th this month. (The) President talked about the details of his new plan at the news conference next day.

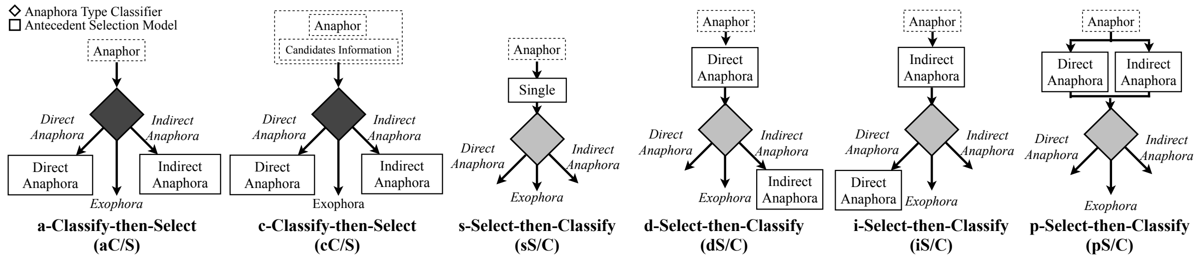

Figure 1: Anaphora resolution models

Furthermore, it is sometimes difficult even for human annotators to determine the definiteness of a bare NP. Therefore, our study for now deals with the resolution of noun phrases modified by a definiteness modifier, i.e., この (*this*) + NP, その (*the*) + NP and あの (*that*) + NP. This problem alone is large enough, because noun phrases modified by a definiteness modifier account for a large proportion of occurrences of nominal anaphora in Japanese texts.

This paper is organized as follows. In the next section, we give a detailed explanation of our investigations. In section 3, we introduce related work. In section 4, the experimental setup is described. In section 5, the results of our investigations are shown and discussed. Finally, the conclusion is given and our future work is described.

2 Anaphora Resolution Models

To explore how our two subtasks can be best configured, we consider several possible models for each subtask.

2.1 Antecedent selection

For antecedent selection, we can employ 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.). In this paper, we adopt Iida et al.'s *tournament model*, a state-of-the-art method for coreference resolution in Japanese (Iida et al., 2005).

One issue to explore for this subtask is whether a single common model should be trained for both direct and indirect anaphora or whether a separate model should be trained for each. So we consider the following two approaches:

- *Single model*: Train a single model with labeled examples of both direct and indirect anaphora.
- *Separate model*: Train two distinct models separately for direct and indirect cases; i.e., a *direct antecedent selection model* is trained only with labeled examples of direct anaphora and an *indirect antecedent selection model* only with labeled examples of indirect anaphora.

The separate model approach is expected to be advantageous because the semantic relation between an anaphor and its antecedent tends to be different in direct and indirect cases. It is crucial to identify synonymous relations between an anaphor and the antecedent to resolve direct anaphora. In example (1), the resolution model has to know that *CD* and *album* are synonymous. For indirect anaphora, on the other hand, the model is required to recognize such semantic relations as *part-whole* and *attribute-of* as in example (2), where knowing that *CD* is an object related to *song* is crucial. This expectation is supported empirically by our experiments as reported in section 5.

2.2 Anaphora type classification

For anaphora type classification, an interesting question is whether this subtask should be carried out before antecedent selection or after. Accordingly, we consider two kinds of configurations: Classify-then-Select and Select-then-Classify as summarized in Figure 1.

2.2.1 Classify-then-Select (C/S) model

Given an anaphor, this model first determines whether the anaphor bears direct anaphora, indirect anaphora or exophora. If the anaphora type is classified as direct, then the model calls for the direct antecedent selection model. If classified as indirect, on the other hand, then the model calls for the indirect antecedent selection model. No antecedent selection model is called for if the anaphora type is classified as exophora. The model for anaphora type classification is trained in a supervised fashion.

According to the information used for the classification, we can consider the following two configurations:

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

To verify the effects of using additional information (i.e., contextual information) we can choose the aC/S model as a baseline. We give a detail of the feature set for these models in Section 4.2.

2.2.2 Select-then-Classify (S/C) model

Given an anaphor, this model first selects the most likely antecedent and then calls for the anaphora type classifier to determine the anaphora type referring to 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 provide contextual information useful for anaphora type classification: for example, if *her new song* is selected as the candidate antecedent in example (2) in section 1, the anaphora type will be easily identified based on the lexical knowledge that *CD* is potentially an object related to *song*.

Since we have two options for antecedent selection, the single and separate models, as described in 2.1, we can consider at least the following four configurations (see Figure 1):

- *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 using the direct anaphora model and then determine 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*: Reverse the steps in the d-Select-then-Classify (dS/C) model.
- *p-Select-then-Classify (pS/C) Model*: Call for the direct anaphora and indirect anaphora models in parallel to select the best candidate antecedent for each case and then determine the anaphora type referring to both candidates.

The pS/C configuration provides the anaphora type classifier with richer contextual information than any other configuration while it requires more computational cost because it always searches for the best candidate, simultaneously searching for direct anaphora and indirect anaphora.

The training of the classifier depends on which configuration is chosen. Assume that we have the following examples in our training set:

(7) *Her new* album$_i$ *was released* yesterday.
I want to get the CD$_i$ *as soon as possible.*
- anaphor: *the CD$_i$*
- antecedent: *album$_i$*
- anaphora type: *direct* anaphora

(8) *The* artist *announced her new* song$_{j'}$.
I want to get the CD$_j$ *as soon as possible.*
- anaphor: *the CD$_j$*
- antecedent: *song$_j$*
- anaphora type: *indirect* anaphora

(9) *I want to get* the CD$_j$ *as soon as possible.*
- anaphor: *the CD$_j$*
- antecedent: None
- anaphora type: *exophora*

Given these annotated examples, sS/C configuration simply takes *the CD$_i$* paired with *album$_i$* from (7) as an instance of *direct* anaphora and *the CD$_j$* paired with *song$_{j'}$* from (8) as an instance of *indirect* anaphora. For (9), all the configurations including sS/C take *the CD* and run its related antecedent selection model to select *pseudo* antecedent. The antecedent is paired with *the CD* as an instance of *exophora*.

The case of the dS/C configuration is slightly more complex. Analogous to the sS/C model, for (7), the classifier takes *the CD$_i$* paired with its direct antecedent *album$_i$* as an instance of *direct* anaphora. For (8), however, since the true *indirect song* is unlikely to be selected as the best candidate by the *direct* anaphora model, we run the *direct* anaphora model to select the *pseudo*-best candidate. Suppose *artist* is selected; then, we create a training instance of indirect anaphora from the *the CD$_j$* paired with *artist*.

An analogous method applies also to the iS/C configuration. Assuming that *yesterday* is selected as the *pseudo*-best candidate indirect antecedent, we label *the CD$_i$* paired with *yesterday* as direct anaphora and *the CD$_j$* paired with *artist* as indirect anaphora.

Finally, for the pS/C configuration, we create a training instance of direct anaphora from ⟨*the CD$_i$*, *album$_i$*, *yesterday*⟩ and a training instance of indirect anaphora from ⟨*the CD$_j$*, *artist*, *song$_{j'}$*⟩.

3 Related Work

Anaphora resolution is an important process for various NLP applications. In contrast with rule-based approaches, such as (Brennan et al., 1987; Lappin and Leass, 1994; Baldwin, 1995; Nakaiwa and Shirai, 1996; Okumura and Tamura, 1996; Mitkov, 1997, etc.), empirical, or machine learning-based approaches to this problem have shown to be a cost-efficient solution achieving performance that is comparable to the best performing rule-based systems (McCarthy and Lehnert, 1995; N.Ge et al., 1998; Soon et al., 2001; Ng and Cardie, 2001; Strube and Muller, 2003; Iida et al., 2005; Yang et al., 2003, etc.). Most of these studies are focused only on the coreference resolution task, particularly in the context of evaluation-oriented research programs such as Message Understanding Conference (MUC)[1] and Automatic Content Extraction (ACE)[2], leaving the issues regarding indirect anaphora relatively unexplored.

As mentioned in Section 1, there has been little attention paid to the issue of anaphora type classification. An exception can be seen in Vieira and Poesio (2000)'s work, which proposes a method of anaphora resolution for definite noun phrases. Their model has two folds. In the first phase, it classifies a given input definite NP into one of three categories, which (very loosely) correspond to our anaphora types: *direct anaphora* (reference to an antecedent with the same head noun), *bridging description* (reference to the same discourse entity already introduced in the discourse using a different head noun or reference to an antecedent which is a distinct but related object of the anaphor), and *discourse-new* (reference to a discourse entity not already introduced in the discourse). This classification is done by a set of hand-coded rules. In

[1] http://www-nlpir.nist.gov/related_projects/muc/index.html
[2] http://www.nist.gov/speech/tests/ace/

the second phase, the antecedent is selected from a set of potential candidates appearing in the preceding discourse in a fashion depending on the category determined in the first phase. So Vieira and Poesio's model carries out anaphora type classification before antecedent selection. In this paper, however, we show through empirical evaluations that this way of configuring the two subtasks is not necessarily the best choice.

4 Evaluation

Our evaluation consists of three steps. First, in order to find out whether an antecedent selection model should be designed and trained separately for direct anaphora and indirect anaphora, we evaluate two antecedent selection models, the single model and separate model described in Section 2.1. Second, the anaphora type classification models described in Section 2.2 are evaluated to explore what information helps the anaphora type classification. Finally, we evaluate the overall accuracy of the entire anaphora resolution task to explore how the models can best be configured. The evaluation is carried out by 10-fold cross-validation.[3]

For the three-way classification for anaphora type classification, we adopt *one-versus-rest* method in our experiments. For creating binary classifiers used in antecedent selection and anaphora type classification, we adopt Support Vector Machines (Vapnik, 1995) [4], with a polynomial kernel of degree 2 and its default parameters.

4.1 Data

For training and testing our models, we created the annotated corpus using the NAIST Text Corpus (Iida et al., 2007), which is publicly available from the Web site of the NAIST NLP group. The NAIST Text Corpus also contains anaphoric relations of noun phrases, but they are strictly limited to coreference relations (i.e. two NP must refer to the same entity in the world). For this reason, we re-annotated (i) direct anaphoric relations, (ii) indirect anaphoric relations and (iii) exophoric NPs marked by three definiteness modifiers, i.e., この (*this*), その (*the*) and あの (*that*). In the specification of our corpus, not only noun phrases but verb phrases too are chosen as antecedents. Consequently, we obtained 600 instances of direct anaphora, 901 instances of indirect anaphora, and 248 instances of exophora. The detailed distribution is shown in Table 1.

In our experiments, we used anaphors whose antecedent is the head of an NP which appears in the preceding context of the anaphor (i.e., cataphora is ignored). Therefore, we used 572 instances of direct anaphora, 878 instances of indirect anaphora and 248 instances of exophora.

[3] In our evaluation of antecedent selection, we also manually checked for cases in which a selected antecedent is not labeled in our corpus but can be evaluated as correct (e.g., as when a discourse entity located in the same anaphoric chain as the labeled antecedent is selected instead).

[4] SVM^{light} http://svmlight.joachims.org/

Table 1: Distribution of anaphoric relations in the annotated corpus

Anaphora Type	Noun	Predicate	Overall
Direct anaphora	530	70	600
Indirect anaphora	466	435	901
Ambiguous	0	8	8
Exophora	-	-	248
Overall	996	513	1,757

'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 'Ambiguous' instances.

4.2 Feature Set

The feature set for antecedent selection was designed based on the literature on coreference resolution (Iida et al., 2005; Ng and Cardie, 2001; Soon et al., 2001; Denis and Baldridge, 2008; Yang et al., 2003, etc). Our feature set captures lexical characters of compared candidates and an anaphor, semantic agreements between them, and so on. See (Iida et al., 2005) in further detail.

- SYNONYM_OF and HYPONYM_OF: The synonymous and hyponymous relationships between an anaphor and a candidate antecedent were identified by the Japanese WordNet (Isahara et al., 2008), the *Bunrui Goi Hyo* thesaurus (NIJL, 2004), and a very large hypernymy hierarchy (about three million hypernymy relations) automatically created from Web texts and Wikipedia (Sumida et al., 2008).
- NOUN-NOUN_SIMILARITY: The distributional similarity between an anaphor and its candidate antecedent was 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 were counted using twenty years of newspaper articles and their distribution $P(n, \langle c, v\rangle)$ was estimated by pLSI (Hofmann, 1999) with 1,000 hidden topic classes.
- ASSOCIATIVENESS: To resolve indirect anaphora, the degree of associativeness between an anaphor *ANA* and candidate *CND* was calculated differently depending on whether *CND* is a noun or predicate. In the case of a noun, the associativeness is calculated from the cooccurrences of *ANA* and *CND* in the pattern "*CND* の *ANA* (*ANA of CND*)", following the literature of bridging reference resolution (Poesio et al., 2004, etc.). Cooccurrence counts were obtained from Web Japanese N-gram Version 1 (Kudo and Kazawa, 2007). In 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 (10).

(10) チャンピオンと闘い$_i$たい。その夢$_i$は実現すると信じている。
I want to fight$_i$ the champion. I'm sure that that dream$_i$ will come true.

For anaphora type classification, we use a different feature set depending on the configuration described in 2.2. For the Classify-then-Select configurations, we designed our feature set based on the literature (Poesio et al., 2004, etc.). Our feature set captures contextual information and lexical syntactic properties (i.e., NP head, POS, case particle and type of definiteness modifier) of an anaphor. The contextual information is encoded by using (i) part-of-speech of the candidate antecedents that the context holds, (ii) the information on whether the context has candidate antecedent(s) with the same head as a head of the anaphor or whether the candidate antecedent(s) is a hyponym or synonym of an anaphor, and (iii) maximum noun-noun similarity and associativeness (see above) for an anaphor and each candidate in the context.

For the Select-then-Classify configurations, on the other hand, the anaphora type classifier uses the information of the best candidate(s) selected in the first step. This sort of information is encoded as a feature set analogous to the feature set for antecedent selection.

5 Results

5.1 Antecedent selection

The results are shown in Table 2, which indicates that the selection models should be designed for each anaphora type separately.[5] We therefore discarded the Single Model for the subsequent experiments.

We also illustrate the learning curves of each model in Figure 2. 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. The curves indicate that in the direct antecedent selection model the accuracy improves as the training data increase, whereas the increase in the accuracy of the indirect antecedent selection model is minimal, even though our data set included more instances for indirect than direct anaphora. These results support the findings in research showing that indirect anaphora is harder to resolve than direct anaphora and suggesting 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 a confusion between candidates belonging to the same semantic category. Here is a typical example of this sort of error:

(11) 私は映画$_j$の知識がないが、『フランケンシュタイン』$_i$ぐらいは知っている。この映画$_i$は、本当に名作だ。
I don't have good knowledge of movies$_j$ but still know of "Frankenstein"$_i$. I think the movie$_i$ is indeed a great masterpiece.

[5]The difference between the Single Model and the Separate Model have statistical significant in the McNemar test ($p < 0.01$).

Table 2: 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)	**58.0% (841/1,450)**

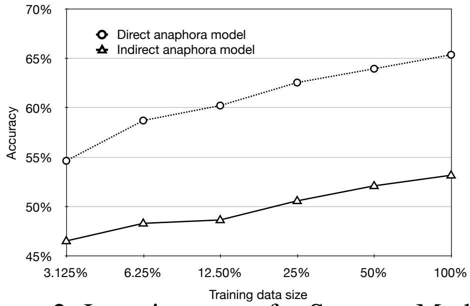

Figure 2: Learning curve for Separate Models

where "映画$_j$ (*movies$_j$*)" was wrongly selected as the antecedent of "この映画$_i$ (*the movie$_i$*)".[6] 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 conducted an evaluation where the ASSOCIATIVENESS feature set was disabled in order to evaluate our lexical resource as described in section 4.2. The results of this additional evaluation showed that the model obtained 51.4% (451/878) accuracy, which is no significant difference compared with the original accuracy. We need to find more, and more useful, clues to capture the associativeness between an anaphor and the related object in indirect anaphora.

5.2 Anaphora type classification

The results of anaphora type classification are shown in Table 3. The aC/S model obtained an accuracy of 75.4%, better than the cC/S model (73.6%), which indicates that contextual information proposed in the literature (Poesio et al., 2004, etc.) was not actually informative. The dS/C model achieved the best accuracy (78.7%) and the improvement indicates that the selected best candidate antecedent provides useful contextual information for anaphora type classification.

As shown in Table 3, identifying exophoric cases tends to be more difficult than identifying anaphoric cases. This difference largely arises from poor performance of our baseline model (aC/S). While our aC/S for indirect anaphora achieved an F-measure of 83.7%, the aC/S model for exophora achieved only an F-measure of 48.9%. Indirect anaphors can sometimes be identified even only by their head words, such as 結果 (result) and 他 (other). However, our analysis could not find such good indicators for exophora. Nevertheless, the pS/C model successfully

[6]In Japanese, in general there is no singular/plural distinction in morphological form.

Table 3: Results of anaphora type classification

	Direct Anaphora			Indirect Anaphora			Exophora			Accuracy	
Model	P	R	F	P	R	F	P	R	F	ATC	Overall
aC/S	67.7%	74.5%	70.9%	80.6%	87.1%	83.7%	75.0%	36.3%	48.9%	75.4%	47.3%
cC/S	69.4%	73.4%	71.4%	74.9%	87.5%	80.7%	92.5%	25.0%	39.4%	73.6%	46.3%
dS/C	70.9%	84.6%	**77.1%**	83.2%	85.6%	**84.4%**	90.1%	40.3%	55.7%	**78.7%**	**50.6%**
iS/C	67.7%	74.8%	71.1%	78.1%	88.3%	82.9%	93.2%	27.8%	42.9%	74.9%	46.3%
pS/C	71.2%	82.0%	76.1%	82.1%	86.7%	84.3%	91.9%	41.1%	**57.2%**	78.4%	50.4%

ATC in the accuracy column indicates the accuracy of anaphora type classification.

improved its performance by using the information of selected candidate antecedents.

5.3 Overall anaphora resolution

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

$$Accuracy = \frac{|C_{antecedent} \cap C_{anaphora\_type}|}{\text{\# of all instances}},$$

where $C_{antecedent}$ is the set of instances whose antecedent is correctly selected and $C_{anaphora\_type}$ is the set of instances whose anaphora type is correctly identified. The results are shown in Table 3. Again, the dS/C model achieved the best accuracy (50.6%), which is significantly better than the Classify-then-Select models.

6 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) how the antecedent selection and anaphora type classification should be carried out, (iii) what information helps anaphora type classification. Our empirical evaluations showed that the separate model achieved better accuracy than the single model and that the d-Select-then-Classify model gave the best results. We have made two findings through the evaluations: (i) an antecedent selection model should be designed separately for each anaphora type using the information useful for identifying its antecedent separately, and (ii) the candidate antecedent selected by an antecedent selection model provides contextual information useful for anaphora type classification. In consequence, we concluded that antecedent selection should be carried out before anaphora type classification.

However, there is still considerable room for improvement in each of two subtasks. Our error analysis has suggested that we need saliency information and various noun-noun relatedness information in antecedent selection. To this aim, we need more feature engineering as well as more extensive empirical evaluations.

References

F. B. Baldwin. 1995. *Cogniac: a discourse processing engine*. Ph.D. thesis.

S. E. Brennan, M. W. Friedman, and C. J. Pollard. 1987. A centering approach to pronouns. In *ACL*, pages 155–162.

H. H. Clark. 1977. Bridging. In *Thinking: Readings in Cognitive Science*.

P. Denis and J. Baldridge. 2008. Specialized models and ranking for coreference resolution. In *Proc. of the 2008 Conference on EMNLP*, pages 660–669.

National Institute for Japanese Language, editor. 2004. *Bunrui Goi Hyo (in Japanese)*. Shuei Shuppan.

T. Hofmann. 1999. Probabilistic latent semantic indexing. In *SIGIR*, pages 50–57.

R. Iida, K. Inui, Y. Matsumoto, and S. Sekine. 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*, pages 831–844.

R. Iida, M. Komachi, K. Inui, and Y. Matsumoto. 2007. Annotating a Japanese text corpus with predicate-argument and coreference relations. In *ACL Workshop on 'Linguistic Annotation Workshop'*, pages 132–139.

H. Isahara, F. Bond, K. Uchimoto, M. Utiyama, and K. Kanzaki. 2008. Development of the Japanese wordnet. In *LREC*.

T. Kudo and H. Kazawa. 2007. *Web Japanese N-gram Version 1*. Gengo Shigen Kyokai.

S. Lappin and H. J. Leass. 1994. An algorithm for pronominal anaphora resolution. *Computational Linguistics.*, 20(4):535–561.

J. F. McCarthy and W. G. Lehnert. 1995. Using decision trees for coreference resolution. In *IJCAI*, pages 1050–1555.

R. Mitkov. 1997. Factors in anaphora resolution: they are not the only things that matter. a case study based on two different approaches. In *ACL and EACL Workshop*.

H. Nakaiwa and S. Shirai. 1996. Anaphora resolution of Japanese zero pronouns with deictic reference. In *COLING*, pages 812–817.

V. Ng and C. Cardie. 2001. Improving machine learning approaches to coreference resolution. In *ACL*, pages 104–111.

N.Ge, J.Hale, and E.Charniak. 1998. A statistical approach to anaphora resolution. In *The 6th Workshop on Very Large Corpora*, pages 161–170.

M. Okumura and K. Tamura. 1996. Zero pronoun resolution in Japmaanese discourse based on centering theory. In *COLING*, pages 871–876.

M. Poesio, R. Mehta, A. Maroudas, and J. Hitzeman. 2004. Learning to resolve bridging references. In *ACL*, pages 144–151.

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

M. Strube and C. Muller. 2003. A machine learning approach to pronoun resolution in spoken dialogue. In *ACL*, pages 168–175.

A. Sumida, N. Yoshinaga, and K. Torisawa. 2008. Boosting precision and recall of hyponymy relation acquisition from hierarchical layouts in wikipedia. In *LREC*.

V. N. Vapnik. 1995. *The Nature of Statistical Learning Theory*.

X. Yang, G. Zhou, J. Su, and C. L. Tan. 2003. Coreference resolution using competition learning approach. In *ACL*, pages 176–183.