Utilizing Features of Verbs in Statistical Zero Pronoun Resolution for Japanese Speech

Sen Yoshida and Masaaki Nagata

NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corp.
2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
yoshida@cslab.kecl.ntt.co.jp, nagata.masaaki@lab.ntt.co.jp

Abstract. This paper proposes a statistical zero pronoun resolution method that utilizes features of verbs. In Japanese speech, the subject is often omitted, especially when it is the first person. To resolve such zero pronouns, features related to the verbs such as functional expressions play important roles. However, recent state-of-the-art zero-pronoun resolution systems lack these features because they are mainly designed for written texts such as newspaper articles, in which first person subjects are rare. We show that a set of verbal features has the ability to distinguish first persons from others in monologue transcriptions, and this improves the accuracy of zero pronoun resolution with statistical machine learning.

Keywords: anaphora resolution, zero pronoun, spoken language processing, Japanese

1 Introduction

Anaphora is a phenomenon whereby an expression (anaphor) can be clarified by binding it with an entity in its context. A typical example of an anaphor is a pronoun. In languages where the subject is often omitted, e.g. Japanese, the omitted subject can be regarded as an anaphor, and this is called a zero pronoun.

For systems that need deep text processing, such as an automatic text summarization system, it is helpful to resolve such anaphora. The task of identifying the referent of an anaphor is called anaphora resolution and this has been the subject of many studies, including work on zero anaphora resolution (Isozaki and Hirao, 2003; Iida et al., 2007a). Those state-of-the-art systems utilize statistical machine learning to achieve good levels of performance. They learn the weights of features extracted from corpora such as annotated newspaper articles to obtain rules for anaphora resolution.

Although most such systems target written texts, anaphora resolution is also valuable for spoken text processing, such as speech summarization (Steinbergera et al., 2007) and interactive QA systems (van Schooten and op den Akker, 2005; Fukumoto, 2006). Recently, some projects have produced corpora of anaphora on spoken texts and used them to build anaphora resolution systems based on statistical machine learning. For example, Müller (2007) tagged anaphora on the ICSI Meeting Corpus and performed pronoun resolution using an empirical method that utilizes a logistic regression classifier.

As regards zero pronouns in Japanese spoken texts, however, there have been few studies that adopt a state-of-the-art statistical machine learning approach with large corpora. Dohsaka (1990) proposed a zero pronoun resolution system with heuristic constraints, which is tested only on a small set of typed dialogues. Fukumoto (2006) introduced zero pronoun resolution into an interactive question answering system. However, the resolution method is only applicable to QA dialogues. It is necessary to develop an anaphora resolution system for Japanese spoken texts based on recent statistical learning methods to obtain adequate quality.

One issue we must consider when dealing with spoken texts is the usage of the first and second person. Most corpora of written texts used for statistical anaphora resolution consist mainly of newspaper articles, where the subjects and objects are rarely first / second persons. On the other hand, in spoken texts such as transcriptions of presentations or meetings, the subjects and objects are often the first person (the speaker) or the second person (the addressee), and they are often omitted in Japanese. Therefore, to distinguish the person of a zero pronoun can be a substantial cue for zero pronoun resolution.

In Japanese, promising cues for determining the person of the subject or the object are features related to the verb, such as auxiliary verbs and other functional expressions. Nakaiwa and Ikehara (1992) used semantic attributes of verbs in their rule-based zero pronoun resolution system. Yamamoto et al. (1997) showed that features related to verbs contribute to the accuracy of their person resolution system based on decision trees learned from a dialogue corpus.

For these reasons, in this paper we propose a method for utilizing verb related features for zero pronoun resolution based on statistical machine learning from a corpus of spoken texts. The system learns the weights of features from annotation data that have anaphora tags with the attributes of persons. This paper is organized as follows: In Section 2 we describe our anaphora data. In Section 3 we present a set of verbal features and show an experimental result that employs it to distinguish first persons from others. In Section 4 we propose our anaphora resolution method and show that the resolution accuracy improves by adding the verbal features. In Section 5 we discuss our method. Section 6 provides our conclusion.

2 Annotation of anaphora in CSJ

In this section, we describe how we tagged anaphora on spoken texts.

As a base corpus of spoken texts, we adopt the Corpus of Spontaneous Japanese (CSJ) (Maekawa et al., 2000). The main part of CSJ consists of monologues and they are divided into two categories; academic presentations and simulated public speech. Each talk is about 10 - 15 minutes long. For both categories, CSJ has a core collection of monologues, in which morphemes and dependency between bunsetsus (base phrases in Japanese) are manually annotated. We tagged anaphora on twelve talks from the core collection, consisting of six academic presentation transcriptions and six simulated public speech transcriptions.

Because the talks are monologues, they include many first-person pronouns and zero pronouns, but few second persons. Therefore, in this paper we focus on the first person.

2.1 Tag types

As a guideline for tagging anaphora, we referred to the NAIST Text Corpus (Iida et al., 2007b), which is a corpus of predicate-argument structures and coreference relations on Japanese newspaper articles. In their paper, they define three types of tags; **predicate**, **event noun**, and **coreference**. In addition, they define two more types in their annotation manual[1]; **bridging reference** and **case alternation**. We annotated all five types of tags for our corpus.

Predicate tags are tagged on verbs, adjectives, and the adverb *da* with its preceding noun. To avoid annotating subsidiary verbs, we use the "longer words" as the base unit for annotation. They are defined by CSJ as well as the shorter words.

The definition of coreference is modified from the NAIST Text Corpus. According to the original definition, all nouns that appear repeatedly in a document are tagged as coreferences. We found that although tagging such repeats manually is laborious, most such repeats in a speech can be detected by the simple matching of strings. Therefore, in our annotation policy, the target of the coreference tag is limited to pronouns and referential expressions. Here, referential expressions consist of referential pronouns *kore* (this), *sore* (it), *are* (that), and nouns preceded by referential

[1] http://cl.naist.jp/řyu-i/coreference_tag.html

adnominals *kono* (this), *sono* (the, that), *ano* (that), *onaji* (the same). Because the corpus is drawn from speech, such anaphora include deictic exophora (reference to an entity in the real world).

If a pronoun or a referential expression indirectly refers to its antecedent via a bridging relation, i.e. the antecedent and the anaphor can be bound together with the particle *no* as "X *no* Y" (Y of X), it is tagged as a bridging reference.

A case alternation tag is attached to adverbs that alternate the case of the sentence arguments; e.g. *reru* (passive).

2.2 Specifying the tag range and the referent

The target range of each annotation tag is a shorter word (or a beginning word and an ending word, if the target consists of more than one shorter word) that is defined in CSJ.

For the purpose of identifying a shorter word within a talk in CSJ, we use a combination of unit IDs on the hierarchy of CSJ's data structure; the sentence ID, the bunsetsu ID, the longer word ID, and the shorter word ID.[2] For example, the ID 23_7_2_1 represents the first shorter word in the second longer word in the seventh bunsetsu in the 23rd sentence.

Sometimes an anaphor refers not to a word or a phrase in the preceding utterances, but to a clause or a sentence. In such cases, the range of the referent tends to be ambiguous and specifying it is rather difficult, especially in spontaneous speech (Poesio and Artstein, 2008; Müller, 2007). According to the NAIST Text Corpus manual, when an anaphor is regarded as a reference to a clause or a sentence, the referent is not identified but a label is attached simply to indicate that the anaphor refers to a clause or a sentence. We follow this policy with our corpus. Similarly, when the referent is an entity in the real world and does not appear in the text, a label is attached to indicate that it is an exophora.

The NAIST Text Corpus manual separates exophoras into three types; **first-person**, **second-person**, and **general**. In other words, the person of a referent is specified only when it is an exophora. When a referent is an endophora, information regarding the person is not available. In our annotation design, the label of the person can be attached regardless of whether the referents are exophoras or endophoras. We attached person labels for all first-person endophoras and exophoras as well as other exophoras for our experiment. This approach is a partial adoption of the entity-mention model (Yang et al., 2008). In terms of that model, all first-person zero pronouns are regarded as mentions of one first-person entity, which is actually the presenter of the talk.

We also added an attribute label for deixis. Typically, a **deictic** label is attached when the presenter says *kore* (this) to point to something such as a graph in a presentation slide.

2.3 Annotation Result

We selected six talks from both academic presentations and simulated public speech, and so annotated a total of twelve talks. The numbers of annotations in each type of anaphora are given in Table 1.

Table 1: Numbers of annotations

type	number
predicate	3,032
event noun	433
coreference	477
bridging reference	58
case alternation	165

[2] In CSJ XML, these are ClauseUnitID, Dep_BunsetsuUnitID, LUWID, and SUWID, respectively.

Table 2: Result of first-person classification

precision	533/683 = 78.0%
recall	533/661 = 80.6%
F measure	79.3%

Table 3: Feature weights for first-person classification

weight	feature
0.814	VSA=physical action
0.814	VSA=emotive action
0.787	VSA=thinking action
0.704	aux. verb=*masu* (polite)
0.533	aux. verb=*tai* (want to)
...	...
-0.531	VSA=natural phenomena
-0.566	VSA=generation
-0.840	aux. verb=*reru* (passive or respect)

3 Distinguishing the first person

In this section, we conduct an experiment to show the ability of verbal features to distinguish the first person from the zero pronouns in CSJ.

We built a binary classifier that determines whether or not each predicate's *ga* element is the first person, using a Support Vector Machine (SVM)[3] with a linear kernel. The features used for this task include the following:

verbal semantic attribute (VSA) The classification of verbs by Nakaiwa and Ikehara (1992), such as 'give and take' expressions and transfer expressions.

auxiliary verb The lemma of the auxiliary verb that is attached to the predicate.

semantic category of the predicate We adopt the category definition of Goi-Taikei Japanese lexicon (Ikehara et al., 1997).

adnominal predicate The value of this feature is 1 if the predicate is an adnominal for any noun phrase.

The experiment was performed for the predicate tags in our anaphora annotation data of twelve CSJ talks. From the predicates that have zero pronoun *ga* elements we eliminated exophoras and endophoras whose antecedents were not words but phrases or sentences. As a result we obtained 1,416 samples, and of these 661 (46.7%) were the first person. We conducted a leave-one-out cross validation. Namely, we ran twelve tests by choosing one talk as the test set and using the remaining talks as the training set.

The resulting accuracy is shown in Table 2, and the learned weights of the features are shown in Table 3. It is shown that the features described in this section have the ability to distinguish the first-person zero pronouns from others.

4 Resolution of zero pronouns

In this section we propose a zero pronoun resolution method. The method can utilize the verbal features provided in the previous section.

Because anaphora resolution must cope with many features and rules to achieve good performance, many state-of-the-art anaphora resolution systems adopt a statistical machine learning

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

approach. There have been certain zero pronoun resolution studies that made use of statistical machine learning with corpora of newspaper articles (Isozaki and Hirao, 2003; Iida et al., 2007a). Our approach is similar to these. As a learning algorithm we adopt a ranking SVM (Joachims, 2002), which is an instance of preference learning.

From a practical point of view, a zero pronoun resolution process can be divided into two subtasks; zero pronoun detection and referent identification. This paper deals only with the referent identification. Here we resolve the *ga* (subject) argument of each predicate, which can be assumed to be required for any predicate. We can extend this system to other types of arguments, such as objects, by introducing a case frame dictionary of verbs and adjectives, which specifies the case required for each verb or adjective.

4.1 Resolution algorithm

Zero pronouns are resolved in the following steps:

1. For each predicate p in both the training set and the test set, extract all nouns from the preceding 20 sentences. Here, the "sentence" unit is not clear in spontaneous speech. We adopted the unit proposed by Takanashi et al. (2003).[4] Each extracted noun is regarded as the antecedent candidate c^p of p.
2. Add the special candidate for the first person if it is needed. If no first-person candidate is extracted in the previous step, a special candidate that represents a first-person referent is added to the set of candidates $\{c^p\}$.
3. For each predicate p and each of its antecedent candidates c_i^p, calculate the feature vector x_i^p.
4. For each predicate p and each of its antecedent candidates c_i^p, calculate the preference value. The preference value is 2 (preferred to the default) if c_i^p is annotated as the antecedent of p in the annotation data or c_i^p is a candidate for the first person and the person attribute of p's *ga* element in the annotation data is **first**. Otherwise, the preference value is 1.
5. Execute the preference learning. Namely, run the learning command of the ranking SVM on the training data $\{x\}$ so that the rank of any candidate c_i^p within p's candidates $\{c^p\}$ can be predicted.
6. Run the rank prediction command of the ranking SVM on the test data. Obtain the highest ranked candidate c_i^p for each predicate p.

In Step 3 of the resolution process, the following features are extracted, in addition to the verbal features provided in Section 3.

case marker The case marker that follows an antecedent candidate, such as *ga* (subject) or *wa* (topic).

adnominal The value of this feature is 1 if the predicate is an adnominal for the candidate. For example, in the clause *sugureta gijutsu* (excellent technique), the predicate *sugureta* is adnominal for *gijutsu*.

distance The number of sentences between the antecedent candidate and the predicate. The value is normalized to [0, 1] so that the magnitude is adjusted to that of the other features.

semantic category of the antecedent candidate The category defined in the Goi-Taikei Japanese lexicon. (See the "semantic category of the predicate" feature above.)

syntax pattern acceptance The Goi-Taikei Japanese lexicon also includes a set of predicate-argument patterns. If a pair consisting of the antecedent candidate and the predicate is accepted by any pattern, the system sets the value at 1. This is almost the same as the "semantic constraints" (Isozaki and Hirao, 2003).

[4] It is labeled as "ClauseUnitID" in CSJ.

Table 4: Resolution accuracy

resolution w/ the special cand. w/o verbal features	502/1416 = 35.5%
resolution w/ the special cand. w/ verbal features	563/1416 = 39.8%
first-person determination + resolution w/o verbal features	846/1416 = 59.7%

predicate lemma The string of the predicate. Because spontaneous speech has disfluency, we do not use a transcription. Instead, we use the lemma of the longer word.[5]

For the first-person special candite, only the semantic category and the predicate lemma are extracted as the feature.

The value of the predicate lemma feature and the features described in Section 3 do not change from candidate to candidate. Therefore, those features have no effect on the preference of the candidates. However, if they are used in combination with the features of the candidates or the relation between the candidate and the predicate, they become important clues. We can handle combinatorial features by making use of a polynomial kernel for the SVM.

4.2 Evaluation

We have implemented our system, which reads CSJ's clause unit XML files and the anaphora tag files, extracts features from them, and runs the proposed resolution algorithm, on top of Unstructured Information Management Architecture (UIMA) (Ferrucci and Lally, 2004).

We undertook an experiment using the same data set as that used in the first person distinguishing experiment described in Section 3. Again, we performed leave-one-out cross validation using our annotation data. As the learning machine we used the ranking mode of SVM^{light} with a 2-dimensional polynomial kernel. The constant C (trade-off between training error and margin) was 0.1.

We compared two feature sets; one including the verbal features described in Section 3 as well as the features described in the preceding subsection, and the other excluding the verbal features.

In addition, we also implemented another version of the resolution algorithm, which is a sequential combination of the first-person determination described in Section 3 and the antecedent resolution described in this section, but without the special candidates. Namely, the system

1. determines whether or not each predicate's *ga* element is the first person by the algorithm described in Section 3, and then
2. for each predicate whose *ga* element is determined not as the first person by the previous step, resolves its antecedent by the algorithm described in Section 4.1 except adding the special candidates for the first person in Step 2 of the resolution process.

We used the feature sets excluding the verbal features in the experiment for this algorithm.

The results are shown in Table 4. The accuracy for the first-person determination + resolution w/o verbal features is calculated against the total combination. Namely, a resolution is regarded as a success iff both the system result and the actual indicates the first person or the selected antecedents coincide between the system result and the actual.

For the resolution utilizing the first-person special candidates, the accuracy is improved by adding the verbal features. Moreover, by sequentially combining the first-person determination with the antecedent resolution, the accuracy is largely improved.

5 Discussion

As described in the previous section, the accuracy in antecedent resolution is largely improved by adding the first-person determination as the preprocessing. This result indicates that, for spoken

[5] Labeled as LUWLemma in CSJ XML.

monologue texts, it is reasonable to resolve zero anaphoras not only by binding them with their antecedents, as most state-of-the-art zero anaphora resolvers do (Isozaki and Hirao, 2003; Iida et al., 2007a), but also by indentifing the persons.

The accuracy of antecedent resolution with the proposed method utilizing the special candidate is not so well compared to the "first-person determination + resolution" method. A major reason for this might be the deficiency of feature values for the special candidate. In the proposed method, only the semantic categories of the first person and the predicate's lemma are extracted as the feature values. On the other hand, other normal candidates have feature values for positional and syntactic features.

Previous studies on Japanese anaphora resolution (Isozaki and Hirao, 2003; Iida et al., 2007a) have achieved an F measure of around 70% for newspaper articles. Aside from spontaneity and other factors of speech, it might be possible to improve our resolution accuracy by, for example, adding some more features.

As the system deals with spoken texts, we can think of utilizing physical quantities of speech as well as syntactic and semantic features. Such features include F0 and the power of antecedent candidates and case markers. We can also make use of the intonation labels that are provided by the CSJ corpus. Moreover, the length of time between the antecedent candidate and the predicate could be used in addition to the number of sentences between them.

6 Conclusions

In this paper, we presented a set of verbal features for statistical machine learning that had the ability to distinguish first persons from others in monologues, and proposed a method to utilize it for the resolution of zero pronouns in Japanese spoken texts. Our experiment showed that the resolution accuracy was improved by adding the verbal features to antecedent features. Especially, the sequential combination of the first-person determination and the antecedent resolution obtained high accuracy.

We may be able to improve the accuracy of the zero pronoun resolution by adding more features.

As our future work, we will expand the target for anaphora resolution from zero pronouns in monologues to other types of anaphors. Those anaphors include demonstrative noun phrases such as the phrase "this figure" uttered by a speaker pointing at a figure in a slide, and second-person zero pronouns in meeting conversations.

References

Dohsaka, K. 1990. Identifying the Referents of Zero-Pronouns in Japanese Based on Pragmatic Constraint Interpretation. In *Proceedings of the Ninth European Conference on Artificial Intelligence*, pages 240–245.

Ferrucci, D. and A. Lally. 2004. UIMA: An architectural approach to unstructured information processing in the corporate research environment. *Journal of Natural Language Engineering*, 10(3-4):327–348.

Fukumoto, J. 2006. Answering Questions of Information Access Dialogue (IAD) Task Using Ellipsis Handling of Follow-Up Questions. In *Proceedings of the HLT-NAACL 2006 Workshop on Interactive Question Answering*, pages 41–48.

Iida, R., K. Inui, and Y. Matsumoto. 2007a. Zero-Anaphora Resolution by Learning Rich Syntactic Pattern Features. *ACM Transactions on Asian Language Information Processing*, 6(4).

Iida, R., M. Komachi, K. Inui, and Y. Matsumoto. 2007b. Annotating a Japanese Text Corpus with Predicate-Argument and Coreference Relations. In *Proceedings of the ACL 2007 Linguistic Annotation Workshop*, pages 132–139.

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

Isozaki, H. and T. Hirao. 2003. Japanese Zero Pronoun Resolution based on Ranking Rules and Machine Learning. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pages 184–191.

Joachims, T. 2002. Optimizing Search Engines Using Clickthrough Data. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 133–142.

Maekawa, K., H. Koiso, S. Furui, and H. Isahara. 2000. Spontaneous Speech Corpus of Japanese. In *Proceedings of the Second International Conference of Language Resources and Evaluation*, pages 947–952.

Müller, C. 2007. Resolving It, This, and That in Unrestricted Multi-Party Dialog. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 816–823.

Nakaiwa, H. and S. Ikehara. 1992. Zero Pronoun Resolution in a Japanese to English Machine Translation System by using Verbal Semantic Attributes. In *Proceedings of the Third Conference on Applied Natural Language Processing*, pages 201–208.

Poesio, M. and R. Artstein. 2008. Anaphoric Annotation in the ARRAU Corpus. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation*.

Steinbergera, J., M. Poesiob, M. A. Kabadjovb, and K. Jezeka. 2007. Two uses of anaphora resolution in summarization. *Information Processing and Management*, 43(6):1663–1680.

Takanashi, K., T. Maruyama, K. Uchimoto, and H. Isahara. 2003. Identification of "Sentences" in Spontaneous Japanese: Detection and Modification of Clause Boundaries. In *Proceedings of the ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pages 183–186.

van Schooten, B. W. and R. op den Akker. 2005. Follow-up utterances in QA dialogue. *Traitement Automatique des Langues*, 46(3):181–206.

Yamamoto, K., E. Sumita, O. Furuse, and H. Iida. 1997. Ellipsis Resolution in Dialogues via Decision-Tree Learning. In *Proceedings of the Fourth Natural Language Processing Pacific Rim Symposium*, pages 423–428.

Yang, X., J. Su, J. Lang, C. Tan, T. Liu, and S. Li. 2008. An Entity-Mention Model for Coreference Resolution with Inductive Logic Programming. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, pages 843–851.