

A Probabilistic Model for Associative Anaphora Resolution

Ryohei Sasano and Sadao Kurohashi

Graduate School of Informatics, Kyoto University,

Yoshida-honmachi, Sakyo-ku, Kyoto

{sasano, kuro}@i.kyoto-u.ac.jp

Abstract

This paper proposes a probabilistic model for associative anaphora resolution in Japanese. Associative anaphora is a type of bridging anaphora, in which the anaphor and its antecedent are not coreferent. Our model regards associative anaphora as a kind of zero anaphora and resolves it in the same manner as zero anaphora resolution using automatically acquired lexical knowledge. Experimental results show that our model resolves associative anaphora with good performance and the performance is improved by resolving it simultaneously with zero anaphora.

1 Introduction

The correct interpretation of anaphora is vital for natural language understanding. Bridging anaphora (Clark, 1975) represents a special part of the general problem of anaphora resolution, which has been studied and discussed for various languages and domains (Hahn et al., 1996; Murata et al., 1999; Poesio et al., 2004; Gasperin and Vieira, 2004; Gasperin and Briscoe, 2008).

Usually bridging anaphora considers two types:¹ **associative anaphors** are noun phrases (NPs) that have an antecedent that is necessary to their interpretation but the relation between the anaphor and its antecedent is different from identity; and **indirect anaphors** are those that have an identity relation with their antecedents but the anaphor and its antecedent have different head

nouns. In this paper, we focus on associative anaphora in Japanese.

Associative anaphora resolution is decomposed into two steps: acquiring lexical knowledge for associative anaphora resolution, and resolving associative anaphora using the acquired knowledge.

Grammatical salience plays a lesser role for resolving anaphors with full lexical heads, than for pronominal anaphora (Strube and Hahn, 1999; Modjeska, 2002). Furthermore, since associative anaphors and their antecedents usually have different head nouns, string matching technique cannot be applied. Therefore, a large and diverse amount of lexical knowledge is essential to understand associative anaphora. For example, to recognize the meronymic relation between “a house” and “the roof” in (1), such knowledge as “a roof” is a part of a building or vehicle is required. To recognize the attributive relation between “Prius” and “the price” in (2), such knowledge as “price” is a price of some goods or service is required.

- (1) There was a house. **The roof** was white.
- (2) Toyota launched the hybrid car Prius in 1997. **The price** was 21.5 million yen.

To acquire such lexical knowledge, various studies have been carried out. Early studies used hand-crafted lexical knowledge such as Word-Net (Strube and Hahn, 1999; Vieira and Poesio, 2000; Meyer and Dale, 2002), but obtained poor or mediocre results. Hence, Poesio et al. (2002) proposed to exploit “ N_h of N_m ” phrases in large corpora to resolve associative anaphora in English; Murata et al. (1999) proposed to exploit “ N_m no N_h ” phrases to resolve associative anaphora in Japanese. Here, the Japanese postposition “*no*” roughly corresponds to “of,” but it has

¹The terminology that we use here is introduced by Hawkins (1978), which is also used in (Vieira et al., 2006).

much broader usage. These studies obtained reasonable results, but the coverage of the acquired knowledge was not sufficient. Recently, a number of researchers argued for using the Web as a source of lexical knowledge, and the Web has been shown to be a useful resource for anaphora resolution (Bunescu, 2003; Markert et al., 2003; Poesio et al., 2004).

Hence, in this study, we acquire the lexical knowledge for associative anaphora resolution from “ N_m no N_h ” phrases in the Web by using the method described in (Sasano et al., 2004). We proposed a method for acquiring such lexical knowledge, called *nominal case frames* (NCFs), using an ordinary language dictionary and “ N_m no N_h ” phrases, and constructed NCFs from newspaper corpora. In this study, we aim to acquire a sufficient amount of lexical knowledge by constructing NCFs from the Web.

As for associative anaphora resolution itself, we propose an integrated probabilistic model for zero anaphora and associative anaphora resolution, in which associative anaphora is regarded as a kind of zero anaphora and resolved by using the same model as zero anaphora. Our model assumes zero pronouns that represent indispensable entities of target noun phrases, which are called *zero adnominal* in (Yamura-Takei, 2003), and conducts zero pronoun resolution.

Let us consider the associative anaphoric relation between “*Prius*” and “*kakaku*” (price). Although “*kakaku*” itself is considered as the anaphor from a conventional point of view (3a), our model assumes a zero pronoun ϕ and considers it as the anaphor (3b).

- (3) a. *Prius* - *kakaku* (price)
 [antecedent: *Prius*, anaphor: *kakaku* (price)]
- b. *Prius* - (ϕ -no) *kakaku* (price (of ϕ))
 [antecedent: *Prius*, anaphor: ϕ]

The point of this study is three-fold: the acquisition of the lexical knowledge for associative anaphora resolution from the Web, the application of zero anaphora resolution model to associative anaphora resolution, and the integrated resolution of zero anaphora and associative anaphora.

2 Construction of Nominal Case Frames

Most nouns have their indispensable entities: “price” is a price of some goods or service, “roof”

is a roof of some building, and “coach” is a coach of some sports. The relation between a noun and its indispensable entities is parallel to that between a verb and its arguments or obligatory cases. In this paper, we call indispensable entities of nouns *obligatory cases*. Note that, *obligatory* does not mean grammatically obligatory but obligatory to interpret the meaning of the noun. Associative anaphora resolution needs comprehensive information of obligatory cases of nouns. *Nominal case frames* (NCFs) describe such information, and we construct them from the Web.

2.1 Automatic Construction of NCFs

First, we briefly introduce our method for constructing NCFs from raw corpora proposed in (Sasano et al., 2004).

Whereas verbal case frame construction uses arguments of each verb (Kawahara and Kurohashi, 2002), nominal case frame construction basically uses adnominal constituents of each noun. However, while the meaning of a verbal argument can be distinguished by the postposition, such as “*ga*” (nominative), “*wo*” (accusative), and “*ni*” (dative), the meaning of an adnominal constituent can not be distinguished easily, because most adnominal constituents appear with the same postposition “*no*” (of). Thus, we first conduct a semantic analysis of adnominal constituents, and then construct NCFs using the results as follows:

1. Collect syntactically unambiguous noun phrases “ N_m no N_h ” from the automatic resulting parses of large corpora.
2. Analyze the relation between N_m and N_h by Kurohashi and Sakai’s method (1999) that exploits an ordinary language dictionary.
3. Depending on the results, classify N_m , and obtain preliminary case slots for N_h .
4. Merge case slots if two preliminary case slots of N_h are similar.
5. Consider frequent case slots as obligatory cases of N_h . The frequency thresholds are varied according to semantic analyses.
6. For each meaning of N_h , collect case slots and construct case frames.

The point of this method is the integrated use of an ordinary dictionary and example phrases from

Table 1: Examples of constructed nominal case frames.

| | Case slot | Examples with freq | Generalized examples with rate |
|---------------------------------------|----------------|---|--|
| <i>kakaku</i> (1) (price) | | Definition: the amount of money you have to pay for something . | |
| | [something] | <i>shōhin</i> (goods):9289, <i>seihin</i> (product):2520, <i>buhin</i> (part):341, <i>yunyuhin</i> (importation):232, ... | [CT:ARTIFACT]:0.93, ... |
| <i>yane</i> (1) (roof) | | Definition: the structure that covers or forms the top of a building etc. | |
| | [building] | <i>ie</i> (house):2505, <i>kuruma</i> (car):1565, <i>koya</i> (hut):895, <i>tatemono</i> (building):883, <i>minka</i> (private house):679, ... | [CT:FACILITY]:0.44, [CT:VEHICLE]:0.13, ... |
| <i>shusho</i> (1) (prime minister) | | Definition: the elected leader of the government in a country that has a parliament. | |
| | [country] | <i>nihon</i> (Japan):2355, <i>kuni</i> (country):272, <i>doitsu</i> (Germany):157, <i>chūgoku</i> (China):130, ... | [NE:LOCATION]:0.82, [CT:VEHICLE]:0.13, ... |
| <i>imouto</i> (1) (sister) | | Definition: a girl or woman who has the same parents as you. | |
| | <relationship> | <i>watashi</i> (me):3385, <i>ore</i> (me):1188, <i>boku</i> (me):898, <i>jibun</i> (oneself):341, <i>tomodachi</i> (friend):537, ... | [CT:PERSON]:0.74, [NE:PERSON]:0.22, ... |
| <i>rebā</i> (1) (lever) | | Definition: a stick or handle on a machine . | |
| | [machine] | <i>buḡeki</i> (brake):122, <i>sokketo</i> (socket):67, <i>waipā</i> (wiper):54, <i>souchi</i> (device):52, ... | [CT:ARTIFACT]:0.61, [CT:VEHICLE]:0.04, ... |
| <i>rebā</i> (2) (liver) | | Definition: the liver of an animal , used as food. | |
| | [animal] | <i>niwatori</i> (chicken):153, <i>buta</i> (pig):153, <i>ushi</i> (cattle):62, <i>doubutsu</i> (animal):25, ... | [CT:ANIMAL]:0.98, ... |
| <i>senshu</i> (1) (player) | | Definition: someone who takes part in a sport . | |
| | [sport] | <i>yakyū</i> (baseball):1252, <i>rirē</i> (relay):736, <i>kyōgi</i> (competition):430, <i>sakkā</i> (soccer):394, ... | [CT:ABSTRACTION]:0.56, ... |
| | <affiliation> | <i>chīmu</i> (team):4409, <i>nihon</i> (Japan):3222, <i>reddu</i> (Reds):771, <i>kankoku</i> (Korea):644, <i>rīgu</i> (league) ... | [NE:LOCATION]:0.33, [CT:ORGANIZATION]:0.30, ... |

* “[]” and “<>” denote dictionary-based and semantic feature-based analysis respectively. For details see (Sasano et al., 2004).

large corpora. Dictionary definition sentences are an informative resource to recognize obligatory cases of nouns. However, it is difficult to resolve associative anaphora by using a dictionary as it is, because all nouns in a definition sentence are not an obligatory case, and only the frequency information of noun phrases tells us which is the obligatory case. On the other hand, a simple method that just collects and clusters “ N_m no N_h ” phrases based on some similarity measure of nouns cannot construct comprehensive nominal case frames, because of polysemy and multiple obligatory cases. For details see (Sasano et al., 2004).

It is desirable to use a probability distribution for deciding whether a case slot is obligatory or not. However, it is difficult to estimate a probability distribution, since we construct nominal case frames not by using the examples of associative anaphora itself but by using the examples of noun phrases “ N_m no N_h ” (N_h of N_m). We use such noun phrases because indispensable entities of noun “ N_h ” often appear as “ N_m .” However, we can say neither frequently appeared “ N_m ” is an indispensable entity of “ N_h ,” nor an indispensable entity frequently appears as “ N_m .” For example, the name of a country is considered as an indispensable entity of “*shusho*” (prime minister), but

does not frequently appear as “ N_m .”² Thus, it is difficult to estimate a probability distribution and we use a hard decision.

2.2 NCF Construction from the Web

We constructed nominal case frames from the Web Corpus (Kawahara and Kurohashi, 2006), which comprises 1.6 billion unique Japanese sentences. In this corpus, there were about 390 million noun phrases “ N_m no N_h ,” about 100 million unique noun phrases, and about 17 million unique head nouns “ N_h .” There were about 4.07 million head nouns that appeared more than 10 times in the corpus, and we used only such head nouns.

The resultant nominal case frames consisted of about 564,000 nouns including compound nouns. We show examples of constructed nominal case frames in Table 1. The average number of case frames for a noun that has case frames was 1.0031, and the average number of case slots for a case frame was 1.0101. However, these statistics differed with the frequency of the noun. Therefore, we investigated the statistics of constructed nominal case frames for each group classified by the frequency of the nouns. Table 2 shows the re-

²It is because “the prime minister of Japan” is often mentioned by simply “the prime minister” in Japanese.

Table 2: The statistics of constructed NCFs.

| Frequency ranking | Proportion of nouns with NCF | # of NCFs per noun with NCF | # of CSs per NCF | Coverage |
|-------------------|------------------------------|-----------------------------|------------------|----------|
| -100 | 56.0% | 1.34 | 1.07 | 17.3% |
| -1000 | 68.8% | 1.17 | 1.16 | 25.6% |
| -10000 | 51.7% | 1.11 | 1.17 | 27.0% |
| -100000 | 14.8% | 1.05 | 1.13 | 17.6% |
| 100001- | 13.7% | 1.0009 | 1.0053 | 12.5% |
| all | 13.9% | 1.0031 | 1.0101 | 100% |

Table 3: Evaluation of constructed NCFs.

| Precision | Recall | F-measure |
|--------------|--------------|-----------|
| 62/70 (0.89) | 62/84 (0.74) | 0.81 |

sult. As for the 10,000 most frequently appeared nouns, which occupied about 70% of all noun appearances, the average number of case frames for a noun was 1.11, and the average number of case slots for a case frame was 1.17.

For evaluating the resultant case frames, we randomly selected 100 nouns from the 10,000 most frequent nouns, and created gold standard case frames for these nouns by hand. For each noun, case frames were given if the noun was considered to have any indispensable entity, and for each case frame, obligatory case slots were given manually: 70 case frames were created that had 84 case slots; 56 case frames had only one case slot, the other 14 case frames had two case slots. 30 nouns had no case frames.

We then evaluated the automatically constructed case slots for these selected nouns. The evaluation result is shown in Table 3: the system output 70 case slots, and out of them, 62 case frames were judged as correct. The F-measure was 0.81. Since the boundary between indispensable cases and optional cases of a noun is not always obvious, this score is considered to be reasonable.

2.3 Generalization of Examples

By using nominal case frames constructed from the Web, sparseness problem was alleviated to some extent, but still remained. For instance, there were thousands of named entities (NEs), which could not be covered intrinsically. To deal with this sparseness problem, we generalized the examples of case slots.

First, we used the categories that Japanese mor-

phological analyzer JUMAN³ adds to common nouns. In JUMAN, about twenty categories are defined and tagged to common nouns. For example, “*kuruma* (car),” “*niwatori* (chicken),” and “*tatemono* (building)” are tagged as “VEHICLE,” “ANIMAL” and “FACILITY,” respectively. For each category, we calculated the rate of categorized examples among all case slot examples, and added it to the case slot as “[CT:VEHICLE]:0.13.”

We also generalized NEs. We used a common standard NE definition for Japanese provided by IREX workshop (1999). We first recognized NEs in the source corpus by using an NE recognizer (Sasano and Kurohashi, 2008), and then constructed NCFs from the NE-recognized corpus. As well as categories, for each NE class, we calculated the NE rate among all case slot examples, and added it to the case slot as “[NE:PERSON]:0.22.” The generalized examples are also included in Table 1.

3 Probabilistic Model

In this study, we apply a lexicalized probabilistic model for zero anaphora resolution proposed in (Sasano et al., 2008) to associative anaphora resolution.

3.1 A Lexicalized Probabilistic Model for Zero Anaphora Resolution

In English, overt pronouns such as “she” and definite noun phrases such as “the company” are anaphors that refer to preceding entities (antecedents). On the other hand, in Japanese, anaphors are often omitted, which are called *zero pronouns*, and zero anaphora resolution is one of the most important techniques for semantic analysis in Japanese.

Here, we introduce our model for zero anaphora resolution (Sasano et al., 2008). This model first resolves coreference and identifies discourse entities; then from the end of each sentence, analyzes each predicate by the following steps:

1. Select a case frame temporarily.
2. Consider all possible correspondences between each input argument and a case slot of the selected case frame.
3. Regard case slots that have no correspondence as zero pronoun candidates.

³<http://nlp.kuee.kyoto-u.ac.jp/nl-resource/juman-e.html>

4. Consider all possible correspondences between zero pronoun candidates and existing entities.
5. For each possible case frame, estimate each correspondence probabilistically, and select the most likely case frame and correspondence.

Figure 1 shows an example of correspondences between case frames and discourse entities.

The probabilistic model gives a probability to each possible case frame CF and case assignment CA when target predicate v , input arguments IA and existing discourse entities ENT are given, and outputs the case frame and case assignment that have the highest probability. That is to say, their model selects the case frame CF_{best} and the case assignment CA_{best} that maximize the probability $P(CF, CA|v, IA, ENT)$:

$$(CF_{best}, CA_{best}) = \operatorname{argmax}_{CF, CA} P(CF, CA|v, IA, ENT) \quad (i)$$

By decomposing case assignment (CA) into direct case assignment (DCA) and the indirect case assignment (ICA) and using several independence assumptions, Equation (i) is transformed into the following equation:⁴

$$(CF_{best}, DCA_{best}, ICA_{best}) = \operatorname{argmax}_{CF, DCA, ICA} \left(P(CF|v) \times P(DCA, IA|CF) \times P(ICA|ENT, CF, DCA) \right) \quad (ii)$$

Here, $P(CF_l|v)$ denotes the probability to select CF_l when target predicate v is given, and estimated by using case structure analysis of large raw corpora.

$P(DCA_k, IA|CF_l)$ denotes the probability to generate direct case assignment and input arguments when a case frame is given, and estimated by using case structure analysis of large raw corpora, the frequency of a case slot example in the automatically constructed verbal case frames, and the web corpus in which the relation between a surface case marker and a case slot is manually annotated.

$P(ICA_k|ENT, CF_l, DCA_k)$ denotes the probability to generate indirect case assignment when existing discourse entities, a case frame and

⁴For details see (Sasano et al., 2008).

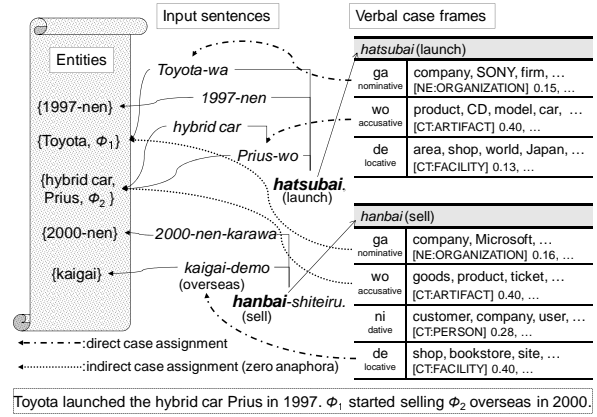


Figure 1: An example of correspondences between verbal case frames and discourse entities.

direct case assignments are given, and estimated by using several preferences on the relation between a zero pronoun and an antecedent, such as a lexical preference, a surface case preferences, and a locational preference.

For example, the lexical preference represents how likely an entity that contains n_{j_m} as a content part is considered to be an antecedent and is estimated by the following equation.

$$\frac{P(n_{j_m}|CF_l, s_j, A'(s_j) = 1)}{P(n_{j_m})} \quad (iii)$$

where, the function $A'(s_j)$ returns 1 if a case slot s_j is filled with an antecedent of a zero pronoun; otherwise 0. $P(n_j|CF_l, s_j, A'(s_j) = 1)$ is calculated by using case frames and denotes the probability of generating a content part n_j of a zero pronoun, when a case frame and a case slot are given and the case slot is filled with an antecedent of a zero pronoun.

3.2 Extension to Associative Anaphora Resolution

We then extend this probabilistic model to associative anaphora resolution. In this model, associative anaphora is regarded as a kind of zero anaphora, that is, the relation between a noun and its obligatory cases is considered to be parallel to that between a verb and its arguments. Omitted obligatory cases are considered to be zero pronouns and resolved by the same process as zero anaphora resolution.

We conduct associative anaphora resolution for only non-coreferent noun phrases. This is because most of the relationships between coreferent noun

conducted, zero pronouns that represent omitted cases of verbs are not considered.

For example, in case of {*hybrid-car*, *Prius*} with $\alpha = 0.6$, if zero anaphora resolution is not conducted, the salience score at the beginning of the third sentence becomes 0.72, because the zero anaphora resolution of “*hanbai*” is not considered; and thus {*hybrid-car*, *Prius*} is not considered as an antecedent candidate.

In order to recognize discourse structure more properly, our model basically resolves associative anaphora simultaneously with zero anaphora, and aims to consider zero pronouns that represent omitted cases of verbs.

3.4 Summary of Our model

Our model is summarized as follows:

1. Parse an input text using the Japanese parser KNP⁶ and recognize NEs.
2. Resolve coreference, and link each mention to an entity or create a new entity.
3. From the end of each sentence, zero anaphora and associative anaphora resolution is conducted for each verb and non-coreferent noun by the following steps:
 - (a) Select a case frame temporarily.
 - (b) Consider all possible correspondences between each input argument and a case slot of the selected case frame.
 - (c) Regard case slots that have no correspondence as zero pronoun candidates.
 - (d) Consider all possible correspondences between zero pronoun candidates and existing entities that has a salience score no less than 1.0.
 - (e) Estimate each correspondence probabilistically, and select the most likely case frame and a correspondence.

4 Experiments

4.1 Setting

We created an anaphoric relation-tagged corpus consisting of 186 web documents (979 sentences), in which all predicate-argument relations and relations between nouns were manually tagged. We show some examples:

- (6) *Toyota-wa 1997-nen Prius-wo hatsubai.*
year launch
2000-nen-kara-wa kaigai-demo hanbai.
year overseas sell
 (Toyota launched Prius in 1997.
 ϕ_1 started selling ϕ_2 overseas in 2000.)
 TAG: *hatsubai* \Leftarrow *ga:Toyota, wo:Prius,*
(NOM) (ACC)
hanbai \Leftarrow ***ga:Toyota, wo:Prius***
(NOM) (ACC)

For the predicate “*hatsubai*” (launch), “*Toyota*” is tagged as *ga* (nominative) case and “*Prius*” is tagged as *wo* (accusative) case. For the predicate “*hanbai*” (sell), “*Toyota*” is tagged as omitted *ga* (nominative) case and “*Prius*” is tagged as omitted *wo* (accusative) case, which are indicated in bold, and such omitted cases are the target of zero anaphora resolution.

As for relations between nouns, both overt and implicit relations are tagged with the Japanese case marker “*no*” (adnominal). In addition, relations between nouns are classified into three categories: indispensable, possible, and adjunct. Since it is not always obvious whether the relations are indispensable or not, borderline relations between indispensable and adjunct are tagged possible. We consider only the implicit relations that are tagged indispensable as the target of associative anaphora resolution.

- (7) *Ken-wa imouto-to yatte-kita.*
sister came.
 (Ken came with ϕ 's sister.)
 TAG: *imouto* \Leftarrow ***no:Ken*** (indispensable)
(ADN)
- (8) *Kôen-ni ikuto benchi-ga atta.*
park went bench was
 (I went to the park. There was a bench in ϕ .)
 TAG: *benchi* \Leftarrow *no:kôen* (possible)
(ADN)

We used 62 documents for testing and used the other 124 documents for calculating several probabilities. In the 62 test documents, 110 associative anaphoric relations were tagged. Each parameter for the proposed model was estimated using maximum likelihood from raw corpora, the tagged corpus, and case frames. As verbal case frames, we used the case frames constructed from the Web corpus comprising 1.6 billion sentences (Sasano et al., 2009).

In order to concentrate on associative anaphora resolution, we used the correct morphemes, named entities, syntactic structures, and coreference re-

⁶<http://nlp.kuee.kyoto-u.ac.jp/nl-resource/knp-e.html>

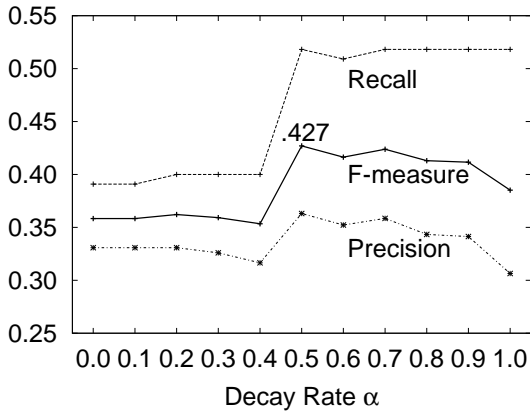


Figure 3: Experimental results of associative anaphora resolution on several salience decay rates α .

lations that were annotated by hand. Since correct coreference relations were given, the number of created entities was the same between the gold standard data and the system output because zero anaphora and associative anaphora resolution did not create new entities.

4.2 Results

Figure 3 shows the experimental results of associative anaphora resolution, in which we used generalized examples, resolved zero anaphora automatically, and varied the decay rate α introduced in Section 3.3 from 0 to 1. When we used the decay rates smaller than 0.5, the recall score worsened clearly. On the other hand, although we expected to obtain higher precision with small decay rate, the highest precision was achieved by the decay rate 0.5. Consequently, we obtained the highest F-measure of 0.427 with the decay rate 0.5. In the following experiments, we fixed the decay rate 0.5.

We utilized two baseline models for demonstrating the effectiveness of our approach: a random model and a salience-based model. The random model selects a case frame and its correspondence randomly from all possible case frames and correspondences. The salience-based model selects a case frame and its correspondence that assign a zero pronoun candidate the existing entity that have highest salience score. In addition, in order to confirm the effectiveness of generalized examples of NCFs, we conducted experiments without using generalized examples. Table 4 shows the experimental results. We can confirm that our proposed model outperforms two baseline models. Without using any generalized examples, the

Table 4: Experimental results of associative anaphora resolution with two baseline models and our model with/without generalized examples.

| Model | Recall | Precision | F-measure |
|----------------|---------------------|---------------------------------|---------------------------------|
| Random* | 0.148 (16.3/110) | 0.035 (16.3/467.5) | 0.056 |
| Salience-based | 0.400 (44/110) | 0.135 (44/325) | 0.202 |
| ----- | | | |
| Proposed | | | |
| CT | NE | | |
| | | 0.318 (35/110) | 0.257 (35/136) |
| | ✓ | 0.345 (38/110) | 0.268 (38/142) |
| ✓ | | 0.464 (51/110) | 0.333 (51/153) |
| ✓ | ✓ | 0.518 (57/110) | 0.363 (57/157) |

CT: Using examples generalized by categories.

NE: Using examples generalized by named entities.

* The average of 10 trials is shown.

F-measure was about 0.14 lower than the method using generalized examples, and we can also confirm the effectiveness of the generalized examples. While generalization of categories much improved the F-measure, generalization of NEs contributed little. This is because the NE rate was smaller than the common noun rate, and so the effect was limited. This tendency was also seen in zero anaphora resolution (Sasano et al., 2008).

In order to investigate the effects of zero anaphora resolution, we tested our model under three conditions: without zero anaphora resolution (no resolution), with zero anaphora resolution (automatically resolved), and with using correct zero anaphora relations that are manually tagged (manually identified). The performance of automatic zero anaphora resolution resulted in a recall of 0.353, a precision of 0.375, and an F-measure of 0.364. Table 5 shows the experimental results. To resolve associative anaphora simultaneously with zero anaphora improved F-measure by 0.072; using correct zero anaphora relations improved F-measure by 0.103. We can confirm that the performance of associative anaphora resolution is improved by considering zero anaphora.

Note that, strictly speaking, these comparisons are not fair because we set the decay rate α to maximize the performance when using generalized examples and resolving zero anaphora automatically. However, these tendencies described above were also seen with other decay rates.

Table 5: The effects of zero anaphora resolution.

| Zero anaphora | Recall | Precision | F-measure |
|------------------------|-------------------|-------------------|-----------|
| No resolution | 0.373 (41/110) | 0.339 (41/121) | 0.355 |
| Automatically resolved | 0.518 (57/110) | 0.363 (57/157) | 0.427 |
| Manually identified | 0.573 (63/110) | 0.382 (63/165) | 0.458 |

4.3 Discussion

By using generalized examples and resolving simultaneously with zero anaphora, our model achieved a recall of 0.518 (57/110), but there were still 53 associative anaphoric relations that were not recognized. Table 6 shows the causes of them.

22 false negatives were caused by salience score filtering. Note that, it does not mean that these 22 associative anaphoric relations were always recognized correctly if the correct antecedents were not filtered by salience score.

Case frame sparseness caused only 5 false negatives. Considering that the recall of nominal case frames was 74% as shown in Table 3, this seems to be too few. This is because we do not considered the relations that tagged possible, and only considered obviously indispensable relations. From this result, we can say that coverage of nominal case frames for nouns that have obviously indispensable entities is much higher than 74%, which is considered to achieve a coverage of about 95% (105/110).

4.4 Comparison with previous work

Murata et al. (1999) proposed a method of utilizing “ N_m no N_h ” phrases for associative anaphora resolution.⁷ They basically used all “ N_m no N_h ” phrases from corpora as a lexical knowledge, and used rule-based approach. They obtained a recall of 0.63 and a precision of 0.68 by using examples of “ X no Y ” (Y of X), a recall of 0.71 and a precision of 0.82 by assuming ideal nominal case frames. One reason of such high performance may be that they considered referential properties of noun phrases, such as generic, indefinite, and definite, while our model does not. We can also say that their experiments were conducted on small and supposedly easy corpora. Half of their corpora

⁷Murata et al. (1999) and we (Sasano et al., 2004) used the terminology *indirect anaphora*, but concerned with the same phenomena as we concerned with in this paper.

Table 6: Causes of false negatives.

| Causes | Num |
|----------------------------|---------|
| Filtered by salience score | 22 (15) |
| Judge as non-anaphoric | 13 (14) |
| Select false antecedents | 13 (13) |
| Case frame sparseness | 5 (5) |
| Total | 53 (47) |

*“()” denotes the number of causes when using correct zero anaphora tags.

were occupied by fairy tale, against which domain specific rules are considered to be effective.

We proposed a rule-based approach for associative anaphora resolution based on automatically acquired nominal case frames (Sasano et al., 2004).⁷ We obtained a recall of 0.633 and a precision of 0.508 against news paper articles. However, we regarded some additional relations that can be interpreted by considering coreference relations as associative anaphoric relations.

- (9) *Chechen Kyôwakoku-no shuto-ni ...*
 Chechen Republic capital
 ... *shuto seiatsu-no saishu dankai-ni ...*
 capital conquer last stage

(... to **the capital** of Chechen Republic ... in the last stage to conquer **the capital** ...)

For example, although the second mention of “*shuto*” (capital) in example (9) means “*Chechen Kyôwakoku-no shuto*” (the capital of Chechen Republic), it can be interpreted by recognizing the coreference relation between the first and second mentions of “*shuto*” (capital). Therefore, as mentioned in Section 3.2, we do not consider such relations as associative anaphora in this study; we included such relations as associative anaphora in (Sasano et al., 2004). The relatively high score is caused by this criterion.

5 Conclusion

In this paper, we proposed a probabilistic model for associative anaphora resolution. Our model regards associative anaphora as a kind of zero anaphora and resolves it in the same manner as zero anaphora resolution that uses automatically acquired case frames. We also showed that the performance of associative anaphora resolution can be improved by resolving it simultaneously with zero anaphora. As future work, we plan to consider referential properties of noun phrases in associative anaphora resolution.

References

- Razvan Bunescu. 2003. Associative anaphora resolution: A web-based approach. In *Proc. of EACL'03: Workshop on The Computational Treatment of Anaphora*, pages 47–52.
- Herbert H Clark. 1975. Bridging. In *Proc. of the Conference on Theoretical Issues in Natural Language Processing*, pages 169–174.
- Caroline Gasperin and Ted Briscoe. 2008. Statistical anaphora resolution in biomedical texts. In *Proc. of COLING'08*, pages 257–264.
- Caroline Gasperin and Renata Vieira. 2004. Using word similarity lists for resolving indirect anaphora. In *Proc. of ACL'04: Workshop on Reference Resolution and its Applications*, pages 40–46.
- Udo Hahn, Michael Strube, and Katja Markert. 1996. Bridging textual ellipsis. In *Proc. of COLING'96*, pages 496–501.
- John A. Hawkins. 1978. *Definiteness and indefiniteness: a study in reference and grammaticality prediction*. Croom Helm Ltd.
- IREX Committee, editor. 1999. *Proc. of the IREX Workshop*.
- Daisuke Kawahara and Sadao Kurohashi. 2002. Fertilization of case frame dictionary for robust Japanese case analysis. In *Proc. of COLING'02*, pages 425–431.
- Daisuke Kawahara and Sadao Kurohashi. 2006. Case frame compilation from the web using high-performance computing. In *Proc. of LREC'06*, pages 1344–1347.
- Sadao Kurohashi and Yasuyuki Sakai. 1999. Semantic analysis of Japanese noun phrases: A new approach to dictionary-based understanding. In *Proc. of ACL'99*, pages 481–488.
- Shalom Lappin and Herbert J. Leass. 1994. An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4):535–562.
- Katja Markert, Malvina Nissim, and Natalia N Modjeska. 2003. Using the web for nominal anaphora resolution. In *Proc. of EACL'03: Workshop on the Computational Treatment of Anaphora*, pages 39–46.
- Josef Meyer and Robert Dale. 2002. Using the WordNet hierarchy for associative anaphora resolution. In *Proc. of SemaNet'02: Building and Using Semantic Networks*.
- Ruslan Mitkov, Richard Evans, and Constantin Orăsan. 2002. A new, fully automatic version of Mitkov's knowledge-poor pronoun resolution method. In *Proc. of CICLing'02*.
- Natalia N Modjeska. 2002. Lexical and grammatical role constraints in resolution other-anaphora. In *Proc. of DAARC'02*.
- Masaki Murata, Hitoshi Isahara, and Makoto Nagao. 1999. Resolution of indirect anaphora in Japanese sentences using examples “X no Y”(Y of X). In *Proc. of ACL'99: Workshop on Coreference and Its Applications*.
- Massimo Poesio, Tomonori Ishikawa, Sabine Schulte im Walde, and Renata Vieira. 2002. Acquiring lexical knowledge for anaphora resolution. In *Proc. of LREC'02*, pages 1220–1224.
- Massimo Poesio, Pahul Mehta, Axel Maroudas, and Janet Hitzeman. 2004. Learning to Resolve Bridging References. In *Proc. of ACL'04*, pages 143–150.
- Ryohei Sasano and Sadao Kurohashi. 2008. Japanese named entity recognition using structural natural language processing. In *Proc. of IJCNLP'08*, pages 607–612.
- Ryohei Sasano, Daisuke Kawahara, and Sadao Kurohashi. 2004. Automatic construction of nominal case frames and its application to indirect anaphora resolution. In *Proc. of COLING'04*, pages 1201–1207.
- Ryohei Sasano, Daisuke Kawahara, and Sadao Kurohashi. 2008. A fully-lexicalized probabilistic model for japanese zero anaphora resolution. In *Proc. of COLING'08*, pages 769–776.
- Ryohei Sasano, Daisuke Kawahara, and Sadao Kurohashi. 2009. The effect of corpus size on case frame acquisition for discourse analysis. In *Proc. of NAACL-HLT'09*, pages 521–529.
- Michael Strube and Udo Hahn. 1999. Functional centering – grounding referential coherence in information structure. *Computational Linguistics*, 25(3):309–344.
- Renata Vieira and Massimo Poesio. 2000. An empirically based system for processing definite descriptions. *Computational Linguistics*, 26(4):539–592.
- Renata Vieira, Eckhard Bick, Jorge Coelho, Vinicius Muller, Sandra Collovini, Jose Souza, and Lucia Rino. 2006. Semantic tagging for resolution of indirect anaphora. In *Proc. of the 7th SIGdial Workshop on Discourse and Dialogue*, pages 76–79.
- Mitsuko Yamura-Takei. 2003. Approaches to zero adnominal recognition. In *Proc. of ACL'03: Student Research Workshop*, pages 87–94.