A Fully-Lexicalized Probabilistic Model
for Japanese Zero Anaphora Resolution

Ryohei Sasano[*]
Graduate School of Information Science
and Technology, University of Tokyo
ryohei@nlp.kuee.kyoto-u.ac.jp

Daisuke Kawahara
National Institute of Information
and Communication Technology
dk@nict.go.jp

Sadao Kurohashi
Graduate School of Infomatics,
Kyoto University
kuro@i.kyoto-u.ac.jp

Abstract

This paper presents a probabilistic model for Japanese zero anaphora resolution. First, this model recognizes discourse entities and links all mentions to them. Zero pronouns are then detected by case structure analysis based on automatically constructed case frames. Their appropriate antecedents are selected from the entities with high salience scores, based on the case frames and several preferences on the relation between a zero pronoun and an antecedent. Case structure and zero anaphora relation are simultaneously determined based on probabilistic evaluation metrics.

1 Introduction

Anaphora resolution is one of the most important techniques in discourse analysis. In English, definite noun phrases such as *the company* and overt pronouns such as *he* are anaphors that refer to preceding entities (antecedents). On the other hand, in Japanese, anaphors are often omitted and these omissions are called *zero pronouns*. We focus on zero anaphora resolution of Japanese web corpus, in which anaphors are often omitted and zero anaphora resolution plays an important role in discourse analysis.

Zero anaphora resolution can be divided into two phases. The first phase is zero pronoun detection and the second phase is zero pronoun resolution. Zero pronoun resolution is similar to coreference resolution and pronoun resolution, which have been studied for many years (e.g. Soon et al. (2001); Mitkov (2002); Ng (2005)). Isozaki and Hirao (2003) and Iida et al. (2006) focused on zero pronoun resolution assuming perfect pre-detection of zero pronouns. However, we consider that zero pronoun detection and resolution have a tight relation and should not be handled independently. Our proposed model aims not only to resolve zero pronouns but to detect zero pronouns.

Zero pronouns are not expressed in a text and have to be detected prior to identifying their antecedents. Seki et al. (2002) proposed a probabilistic model for zero pronoun detection and resolution that uses hand-crafted case frames. In order to alleviate the sparseness of hand-crafted case frames, Kawahara and Kurohashi (2004) introduced wide-coverage case frames to zero pronoun detection that are automatically constructed from a large corpus. They use the case frames as selectional restriction for zero pronoun resolution, but do not utilize the frequency of each example of case slots. However, since the frequency is shown to be a good clue for syntactic and case structure analysis (Kawahara and Kurohashi, 2006), we consider the frequency also can benefit zero pronoun detection. Therefore we propose a probabilistic model for zero anaphora resolution that fully utilizes case frames. This model directly considers the frequency and estimates case assignments for overt case components and antecedents of zero pronoun simultaneously.

In addition, our model directly links each zero pronoun to an entity, while most existing models link it to a certain mention of an entity. In our model, mentions and zero pronouns are treated similarly and all of them are linked to corresponding entities. In this point, our model is similar to

* Research Fellow of the Japan Society for the Promotion of Science (JSPS)

Table 1: Examples of Constructed Case Frames.

	case slot	examples	generalized examples with rate
tsumu (1) (load)	*ga* (subjective)	he, driver, friend, ···	[CT:PERSON]:0.45, [NE:PERSON]:0.08, ···
	wo (objective)	baggage, luggage, hay, ···	[CT:ARTIFACT]:0.31, ···
	ni (dative)	car, truck, vessel, seat, ···	[CT:VEHICLE]:0.32, ···
tsumu (2) (accumulate)	*ga* (subjective)	player, children, party, ···	[CT:PERSON]:0.40, [NE:PERSON]:0.12, ···
	wo (objective)	experience, knowledge, ···	[CT:ABSTRACT]:0.47, ···
⋮	⋮		⋮
hanbai (1) (sell)	*ga* (subjective)	company, Microsoft, firm, ···	[NE:ORGANIZATION]:0.16, [CT:ORGANIZATION]:0.13, ···
	wo (objective)	goods, product, ticket, ···	[CT:ARTIFACT]:0.40, [CT:FOOD]:0.07, ···
	ni (dative)	customer, company, user, ···	[CT:PERSON]:0.28, ···
	de (locative)	shop, bookstore, site ···	[CT:FACILITY]:0.40, [CT:LOCATION]:0.39, ···
⋮	⋮		⋮

the coreference model proposed by Luo (2007) and that proposed by Yang et al. (2008). Due to this characteristic, our model can utilize information beyond a mention and easily consider salience (the importance of an entity).

2 Construction of Case Frames

Case frames describe what kinds of cases each predicate has and what kinds of nouns can fill these case slots. We construct case frames from a large raw corpus by using the method proposed by Kawahara and Kurohashi (2002), and use them for case structure analysis and zero anaphora resolution. This section shows how to construct the case frames.

2.1 Basic Method

After a large corpus is parsed by a Japanese parser, case frames are constructed from modifier-head examples in the resulting parses. The problems of case frame construction are syntactic and semantic ambiguities. That is to say, the parsing results inevitably contain errors and predicate senses are intrinsically ambiguous. To cope with these problems, case frames are gradually constructed from reliable modifier-head examples.

First, modifier-head examples that have no syntactic ambiguity are extracted, and they are disambiguated by coupling a predicate and its closest case component. Such couples are explicitly expressed on the surface of text, and can be considered to play an important role in sentence meanings. For instance, examples are distinguished not by predicates (e.g., "*tsumu* (load/accumulate))", but by couples (e.g., "*nimotsu-wo tsumu* (load baggage)" and "*keiken-wo tsumu* (accumulate experience))". Modifier-head examples are aggregated in this way, and yield basic case frames.

Thereafter, the basic case frames are clustered to merge similar case frames. For example, since "*nimotsu-wo tsumu* (load baggage)" and "*busshi-wo tsumu* (load supplies)" are similar, they are clustered. The similarity is measured using a thesaurus (The National Language Institute for Japanese Language, 2004). Using this gradual procedure, we constructed case frames from approximately 1.6 billion sentences extracted from the web. In Table 1, some examples of the resulting case frames are shown.

2.2 Generalization of Examples

By using case frames that are automatically constructed from a large corpus, sparseness problem is alleviated to some extent, but still remains. For instance, there are thousands of named entities (NEs), which cannot be covered intrinsically. To deal with this sparseness problem, we generalize the examples of case slots. Kawahara and Kurohashi also give generalized examples such as "agent" but only a few types. We generalize case slot examples based on categories of common nouns and NE classes.

First, we use the categories that Japanese morphological analyzer JUMAN[1] adds to common nouns. In JUMAN, about twenty categories are defined and tagged to common nouns. For example, "*ringo* (apple)," "*inu* (dog)" and "*byoin* (hospital)" are tagged as "FOOD," "ANIMAL" and "FACILITY," respectively. For each category, we calculate the rate of categorized example among all case slot examples, and add it to the case slot as "[CT:FOOD]:0.07."

We also generalize NEs. We use a common standard NE definition for Japanese provided by IREX workshop (1999). IREX defined eight NE classes as shown in Table 2. We first recognize NEs in the source corpus by using an NE recognizer (Sasano and Kurohashi, 2008), and then construct case frames from the NE-recognized corpus.

---

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

Table 2: Definition of NE in IREX.

NE class	Examples
ORGANIZATION	NHK Symphony Orchestra
PERSON	Kawasaki Kenjiro
LOCATION	Rome, Sinuiju
ARTIFACT	Nobel Prize
DATE	July 17, April this year
TIME	twelve o'clock noon
MONEY	sixty thousand dollars
PERCENT	20%, thirty percents

As well as categories, for each NE class, we calculate the NE rate among all case slot examples, and add it to the case slot as "[NE:PERSON]:0.12."

The generalized examples are also included in Table 1. This information is utilized to estimate the case assignment probability, which will be mentioned in Section 3.2.3.

3 Zero Anaphora Resolution Model

In this section, we propose a probabilistic model for Japanese zero anaphora resolution.

3.1 Overview

The outline of our model is as follows:

1. Parse an input text using the Japanese parser KNP[2] and recognize NEs.

2. Conduct coreference resolution and link each mention to an entity or create new entity.

3. For each sentence, from the end of the sentence, analyze each predicate by the following steps:

   (a) Select a case frame temporarily.

   (b) Consider all possible correspondence between each input case component and an case slot of the selected case frame.

   (c) Regard case slots that have no correspondence as zero pronoun candidates.

   (d) Consider all possible correspondence between each zero pronoun candidate and an existing entity.

   (e) For each possible case frame, estimate each correspondence probabilistically, and select the most likely case frame and correspondence.

In this paper, we concentrate on three case slots for zero anaphora resolution: "*ga* (subjective)," "*wo* (objective)" and "*ni* (dative)," which cover about 90% of zero anaphora.

Morphological analysis, NE recognition, syntactic analysis and coreference resolution are conducted as pre-processes for zero anaphora resolution. Therefore, the model has already recognized existing entities before zero anaphora resolution.

For example, let us consider the following text:

(i) Toyota-*wa* 1997-*nen* hybrid car Prius-*wo hatsubai*(launch). 2000-*nen-karaha kaigai* (overseas)-*demo hanbai*(sell)-shiteiru.

(Toyota launched the hybrid car Prius in 1997. ϕ_1 started selling ϕ_2 overseas in 2000.)

Figure 1 shows the analysis process for this text. There are three mentions[3] in the first sentence, and the two mentions, *hybrid car* and *Prius*, appear in apposition. Thus, after the pre-processes, two entities, {Toyota} and {hybrid-car, Prius}, are created.

Then, case structure analysis for the predicate *hatsubai* (launch) is conducted. First, one of the case frames of *hatsubai* (launch) is temporarily selected and each input case component is assigned to an appropriate case slot. For instance, case component *Toyota* is assigned to *ga* case slot and *Prius* is assigned to *wo* case slot[4]. In this case, though there is a mention *hybrid-car* that is not a case component of *hatsubai* (launch) by itself, it refers to the same entity as *Prius* refers. Thus, there is no entity that is not linked to *hatsubai* (launch), and no further analysis is conducted.

Now, let us consider the second sentence. A mention *kaigai* (overseas) appears and a new entity {kaigai} is created. Then, case structure analysis for the predicate *hanbai* (sell) is conducted. There is only one overt case component *kaigai* (overseas), and it is assigned to a case slot of the selected case frame of *hanbai* (sell). For instance, the case frame *hanbai*(1) in Table 1 is selected and *kaigai* (overseas) is assigned to *de* (locative) case slot. In this case, the remaining case slots *ga*, *wo* and *ni* are considered as zero pronouns, and all possible correspondences between zero pronouns and remaining entities are considered. As a result of probabilistic estimation, the entity {Toyota} is assigned to *ga* case, the entity {hybrid-car, Prius} is assigned to *wo* case and no entity is assigned to *ni* case.

Now, we show how to estimate the correspondence probabilistically in the next subsection.
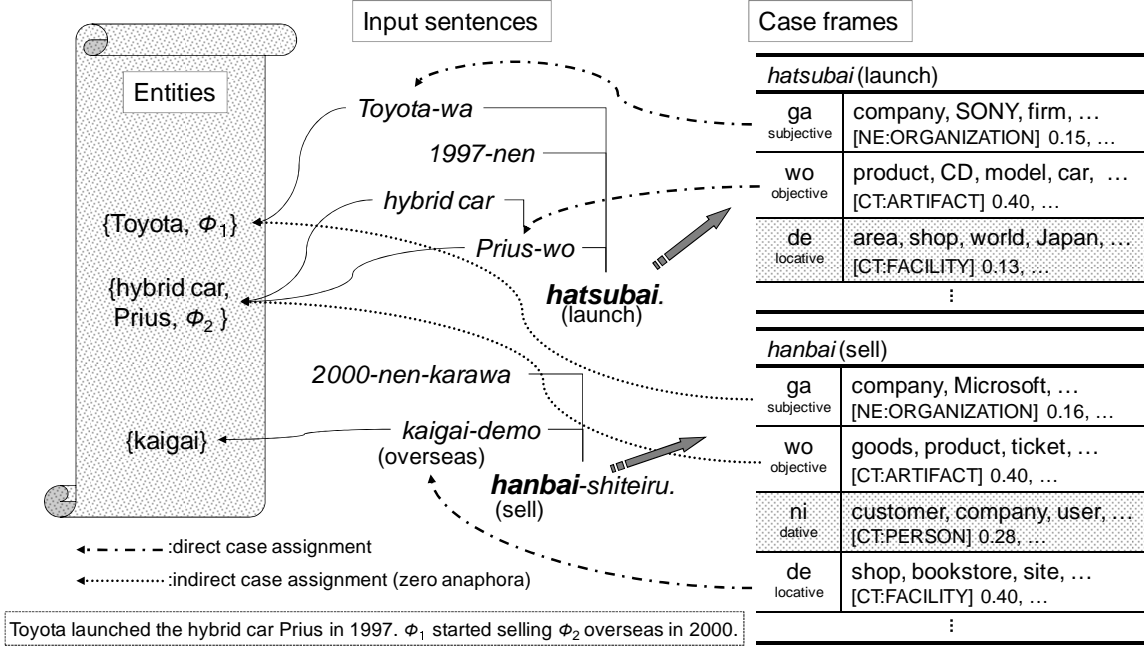
Figure 1: An Example of Case Assignment CA_k.

3.2 Probabilistic Model

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

$$(CF_{best}, CA_{best})$$
$$= \underset{CF, CA}{\operatorname{argmax}} P(CF, CA|v, ICC, ENT) \quad (1)$$

Though case assignment CA usually represents correspondences between input case components and case slots, in our model it also represents correspondences between antecedents of zero pronouns and case slots. Hereafter, we call the former direct case assignment (DCA) and the latter indirect case assignment (ICA). Then, we transform $P(CF_l, CA_k|v, ICC, ENT)$ as follows:

$$P(CF_l, CA_k|v, ICC, ENT)$$
$$= P(CF_l|v, ICC, ENT)$$
$$\quad \times P(DCA_k|v, ICC, ENT, CF_l)$$
$$\quad \times P(ICA_k|v, ICC, ENT, CF_l, DCA_k)$$
$$\approx P(CF_l|v, ICC) \times P(DCA_k|ICC, CF_l)$$
$$\quad \times P(ICA_k|ENT, CF_l, DCA_k) \quad (2)$$
$$= P(CF_l|v) \times P(DCA_k, ICC|CF_l)/P(ICC|v)$$
$$\quad \times P(ICA_k|ENT, CF_l, DCA_k) \quad (3)$$

$$\left(\because P(CF_l|v, ICC) = \frac{P(CF_l, ICC|v)}{P(ICC|v)} \right.$$
$$= \frac{P(ICC|CF_l, v) \cdot P(CF_l|v)}{P(ICC|v)}$$
$$= \frac{P(ICC|CF_l) \cdot P(CF_l|v)}{P(ICC|v)},$$

$(\because CF_l$ contains the information about v.)

$$P(DCA_k|ICC, CF_l)$$
$$\left. = \frac{P(DCA_k, ICC|CF_l)}{P(ICC|CF_l)} \right)$$

Equation (2) is derived because we assume that the case frame CF_l and direct case assignment DCA_k are independent of existing entities ENT, and indirect case assignment ICA_k is independent of input case components ICC.

Because $P(ICC|v)$ is constant, we can say that our model selects the case frame CF_{best} and the direct case assignment DCA_{best} and indirect case assignment ICA_{best} that maximize the probability $P(CF, DCA, ICA|v, ICC, ENT)$:

$$(CF_{best}, DCA_{best}, ICA_{best}) =$$
$$\underset{CF, DCA, ICA}{\operatorname{argmax}} \Big(P(CF|v) \times P(DCA, ICC|CF)$$
$$\times P(ICA|ENT, CF, DCA) \Big) \quad (4)$$

The probability $P(CF_l|v)$, called *generative probability of a case frame*, is estimated from case structure analysis of a large raw corpus. The following subsections illustrate how to calculate $P(DCA_k, ICC|CF_l)$ and $P(ICA_k|ENT, CF_l, DCA_k)$.

3.2.1 Generative Probability of Direct Case Assignment

For estimation of *generative probability of direct case assignment* $P(DCA_k, ICC|CF_l)$, we follow Kawahara and Kurohashi's (2006) method. They decompose $P(DCA_k, ICC|CF_l)$ into the following product depending on whether a case slot s_j is filled with an input case component or vacant:

$$
\begin{aligned}
P(DCA_k, & ICC|CF_l) = \\
& \prod_{s_j:A(s_j)=1} P(A(s_j) = 1, n_j, c_j|CF_l, s_j) \\
\times & \prod_{s_j:A(s_j)=0} P(A(s_j) = 0|CF_l, s_j) \\
= & \prod_{s_j:A(s_j)=1} \Big\{ P(A(s_j) = 1|CF_l, s_j) \\
& \qquad \times P(n_j, c_j|CF_l, s_j, A(s_j) = 1) \Big\} \\
\times & \prod_{s_j:A(s_j)=0} P(A(s_j) = 0|CF_l, s_j) \qquad (5)
\end{aligned}
$$

where the function $A(s_j)$ returns 1 if a case slot s_j is filled with an input case component; otherwise 0, n_j denotes the content part of the case component, and c_j denotes the surface case of the case component.

The probabilities $P(A(s_j) = 1|CF_l, s_j)$ and $P(A(s_j) = 0|CF_l, s_j)$ are called *generative probability of a case slot*, and estimated from case structure analysis of a large raw corpus as well as *generative probability of a case frame*.

The probability $P(n_j, c_j|CF_l, s_j, A(s_j) = 1)$ is called *generative probability of a case component* and estimated as follows:

$$
\begin{aligned}
& P(n_j, c_j|CF_l, s_j, A(s_j) = 1) \\
& \approx P(n_j|CF_l, s_j, A(s_j){=}1) \times P(c_j|s_j, A(s_j){=}1) \quad (6)
\end{aligned}
$$

$P(n_j|CF_l, s_j, A(s_j) = 1)$ means the generative probability of a content part n_j from a case slot s_j in a case frame CF_l, and estimated by using the frequency of a case slot example in the automatically constructed case frames. $P(c_j|s_j, A(s_j) = 1)$ is approximated by $P(c_j|case\_type\_of(s_j), A(s_j){=}1)$ and estimated from the web corpus in which the relationship between a surface case marker and a case slot is annotated by hand.

3.2.2 Probability of Indirect Case Assignment

To estimate *probability of indirect case assignment* $P(ICA_k|ENT, CF_l, DCA_k)$ we also decompose it into the following product depending

Table 3: Location Classes of Antecedents.

intra-sentence: case components of	
L_1	: parent predicate of V_z
L_2	: parent predicate of V_z" (parallel)
L_3	: child predicate of V_z
L_4	: child predicate of V_z (parallel)
L_5	: parent predicate of parent noun phrase of V_z
L_6	: parent predicate of parent predicate of V_z (parallel)
L_7	: other noun phrases following V_z
L_8	: other noun phrases preceding V_z
inter-sentence: noun phrases in	
L_9	: 1 sentence before
L_{10}	: 2 sentences before
L_{11}	: 3 sentences before
L_{12}	: more than 3 sentences before

on whether a case slot s_j is filled with an entity ent_j or vacant:

$$
\begin{aligned}
P(ICA_k & |ENT, CF_l, DCA_k) = \\
& \prod_{s_j:A'(s_j)=1} P(A'(s_j) = 1, ent_j|ENT, CF_l, s_j) \\
\times & \prod_{s_j:A'(s_j)=0} P(A'(s_j) = 0|ENT, CF_l, s_j) \quad (7)
\end{aligned}
$$

where the function $A'(s_j)$ returns 1 if a case slot s_j is filled with an entity ent_j; otherwise 0. Note that we only consider case slots *ga*, *wo* and *ni* that is not filled with an input case component. We approximate $P(A'(s_j) = 1, ent_j|ENT, CF_l, s_j)$ and $P(A'(s_j) = 0|ENT, CF_l, s_j)$ as follows:

$$
\begin{aligned}
& P(A'(s_j) = 1, ent_j|ENT, CF_l, s_j) \\
& \qquad \approx P(A'(s_j) = 1, ent_j|ent_j, CF_l, s_j) \\
& \qquad = P(A'(s_j) = 1|ent_j, CF_l, s_j) \qquad (8)
\end{aligned}
$$

$$
\begin{aligned}
& P(A'(s_j) = 0|ENT, CF_l, s_j) \\
& \qquad \approx P(A'(s_j) = 0|case\_type\_of(s_j)) \quad (9)
\end{aligned}
$$

Equation (8) is derived because we assume $P(A'(s_j) = 1|CF_l, s_j)$ is independent of existing entities that are not assigned to s_j. Equation (9) is derived because we assume $P(A'(s_j) = 0)$ is independent of ENT and CF_l, and only depends on the case type of s_j, such as *ga*, *wo* and *ni*. $P(A'(s_j){=}0|case\_type\_of(s_j))$ is the probability that a case slot has no correspondence after zero anaphora resolution and estimated from anaphoric relation tagged corpus.

Let us consider the probability $P(A'(s_j) = 1|ent_j, CF_l, s_j)$. We decompose ent_j into content part n_{j_m}, surface case c_{j_n} and location class l_{j_n}. Here, location classes denote the locational relations between zero pronouns and their antecedents. We defined twelve location classes as described in Table 3. In Table 3, V_z means a predicate that has a zero pronoun. Note that we also consider the

locations of zero pronouns that are linked to the target entity as location class candidates. Now we roughly approximate $P(A'(s_j)=1|ent_j, CF_l, s_j)$ as follows:

$$P(A'=1|ent_j, CF_l, s_j)$$

$$=P(A'=1|n_{j_m}, c_{j_n}, l_{j_n}, CF_l, s_j)$$

$$=\frac{P(n_{j_m}, c_{j_n}, l_{j_n}|CF_l, s_j, A'=1) \cdot P(A'=1|CF_l, s_j)}{P(n_{j_m}, c_{j_n}, l_{j_n}|CF_l, s_j)}$$

$$\approx \frac{P(n_{j_m}|CF_l, s_j, A'=1)}{P(n_{j_m}|CF_l, s_j)} \cdot \frac{P(c_{j_n}|CF_l, s_j, A'=1)}{P(c_{j_n}|CF_l, s_j)}$$

$$\cdot \frac{P(l_{j_n}|CF_l, s_j, A'=1)}{P(l_{j_n}|CF_l, s_j)} \cdot P(A'=1|CF_l, s_j) \quad (10)$$

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

$$\times \frac{P(c_{j_n}|case\_type\_of(s_j), A'=1)}{P(c_{j_n})}$$

$$\times P(A'=1|l_{j_n}, case\_type\_of(s_j)) \quad (11)$$

$$\left(\because \frac{P(l_{j_n}|CF_l, s_j, A'=1)}{P(l_{j_n}|CF_l, s_j)} \cdot P(A'=1|CF_l, s_j) \right.$$

$$\left. = \frac{P(A'=1, l_{j_n}|CF_l, s_j)}{P(l_{j_n}|CF_l, s_j)} = P(A'=1|CF_l, l_{j_n}, s_j) \right)$$

Note that because ent_j is often mentioned more than one time, there are several combinations of content part n_{j_m}, surface case c_{j_n} and location class l_{j_n} candidates. We select the pair of m and n with the highest probability.

Equation (10) is derived because we assume n_{j_m}, c_{j_n} and l_{j_n} are independent of each other. Equation (11) is derived because we approximate $P(A'=1|CF_l, l_{j_n}, s_j)$ as $P(A'=1|l_{j_n}, case\_type\_of(s_j))$, and assume $P(n_{j_m})$ and $P(c_{j_n})$ are independent of CF_l and s_j. Since these approximation is too rough, specifically, $P(n_{j_m})$ and $P(c_{j_n})$ tend to be somewhat smaller than $P(n_{j_m}|CF_l, s_j)$ and $P(c_{j_n}|CF_l, s_j)$ and equation (11) often becomes too large, we introduce a parameter $\alpha(\leq 1)$ and use the α-times value as $P(A'=1|ent_j, CF_l, s_j)$.

The first term of equation (11) represents how likely an entity that contains n_{j_m} as a content part is considered to be an antecedent, the second term represents how likely an entity that contains c_{j_n} as a surface case is considered to be an antecedent, and the third term gives the probability that an entity that appears in location class l_{j_n} is an antecedent.

The probabilities $P(n_{j_m})$ and $P(c_{j_n})$ are estimated from a large raw corpus. The probabili-

ties $P(c_{j_n}|case\_type\_of(s_j))$ and $P(A'=1|l_{j_n}, case\_type\_of(s_j))$ are estimated from the web corpus in which the relationship between an antecedent of a zero pronoun and a case slot, and the relationship between its surface case marker and a case slot are annotated by hand. Then, let us consider the probability $P(n_{j_m}|CF_l, s_j, A'(s_j)=1)$ in the next subsection.

3.2.3 Probability of Component Part of Zero Pronoun

$P(n_{j_m}|CF_l, s_j, A'=1)$ is similar to $P(n_j|CF_l, s_j, A=1)$ and can be estimated approximately from case frames using the frequencies of case slot examples. However, while $A'(s_j)=1$ means s_j is not filled with input case component but filled with an entity as the result of zero anaphora resolution, case frames are constructed by extracting only the input case component. Therefore, the content part of a zero anaphora antecedent n_{j_m} is often not included in the case slot examples. To cope with this problem, we utilize generalized examples.

When one mention of an entity is tagged any category or recognized as an NE, we also use the category or the NE class as the content part of the entity. For examples, if an entity {Prius} is recognized as an artifact name and assigned to *wo* case of the case frame *hanbai*(1) in Table 1, the system also calculates:

$$\frac{P(NE\!:\!ARTIFACT|hanbai(1), wo, A'(wo)=1)}{P(NE\!:\!ARTIFACT)}$$

besides:

$$\frac{P(Prius|hanbai(1), wo, A'(wo)=1)}{P(Prius)}$$

and uses the higher value.

3.3 Salience Score

Previous works reported the usefulness of salience for anaphora resolution (Lappin and Leass, 1994; Mitkov et al., 2002). In order to consider salience of an entity, we introduce salience score, which is calculated by the following set of simple rules:

- $+2$: mentioned with topical marker "*wa*".
- $+1$: mentioned without topical marker "*wa*".
- $+0.5$: assigned to a zero pronoun.
- $\times 0.7$: beginning of each sentence.

For examples, we consider the salience score of the entity {Toyota} in (i) in Section 3.1. In the first sentence, since {Toyota} is mentioned with topical marker "*wa*", the salience score is 2. At the beginning of the second sentence it becomes 1.4,

Table 4: Data for Parameter Estimation.

probability	data
$P(n_j)$	raw corpus
$P(c_j)$	raw corpus
$P(c_j|case\_type\_of(s_j), A(s_j)=1)$	tagged corpus
$P(c_j|case\_type\_of(s_j), A'(s_j)=1)$	tagged corpus
$P(n_j|CF_l, s_j, A(s_j)=1)$	case frames
$P(n_j|CF_l, s_j, A'(s_j)=1)$	case frames
$P(CF_l|v_i)$	case structure analysis
$P(A(s_j)=\{0,1\}|CF_l, s_j)$	case structure analysis
$P(A'(s_j)=0|case\_type\_of(s_j))$	tagged corpus
$P(A'(s_j)=1|l_j, case\_type\_of(s_j))$	tagged corpus

Table 5: Experimental Results.

	R	P	F
Kawahara & Kurohashi	.230 (28/122)	.173 (28/162)	.197
Proposed ($\alpha = 1$)	**.426** (52/122)	.271 (52/192)	.331
($\alpha = 1/2$)	.410 (50/122)	.373 (50/134)	**.391**
($\alpha = 1/4$)	.295 (36/122)	**.419** (36/86)	.346

and after assigned to the zero pronoun of "*hanbai*" it becomes 1.9. Note that we use the salience score not as a probabilistic clue but as a filter to consider the target entity as a possible antecedent. When we use the salience score, we only consider the entities that have the salience score no less than 1.

4 Experiments

4.1 Setting

We created an anaphoric relation-tagged corpus consisting of 186 web documents (979 sentences). We selected 20 documents for test and used the other 166 documents for calculating several probabilities. Since the anaphoric relations in some web documents were not so clear and too difficult to recognize, we did not select such documents for test. In the 20 test documents, 122 zero anaphora relations were tagged between one of the mentions of the antecedent and the target predicate that had the zero pronoun.

Each parameter for proposed model was estimated using maximum likelihood from the data described in Table 4. The case frames were automatically constructed from web corpus comprising 1.6 billion sentences. The case structure analysis was conducted on 80 million sentences in the web corpus, and $P(n_j)$ and $P(c_j)$ were calculated from the same 80 million sentences.

In order to concentrate on zero anaphora resolution, we used the correct morphemes, named entities, syntactic structures and coreferential relations that were annotated by hand. Since correct coreferential relations were given, the number of created entities was same between the gold standard and the system output because zero anaphora resolution did not create new entities.

4.2 Experimental Results

We conducted experiments of zero anaphora resolution. As the parameter α introduced in Section 3.2.2., we tested 3 values 1, 1/2, and 1/4. For comparison, we also tested Kawahara and Kurohashi's (2004) model. The experimental results are shown in Table 5, in which recall R, precision P and F-measure F were calculated by:

$$R = \frac{\text{\# of correctly recognized zero anaphora}}{\text{\# of zero anaphora tagged in corpus}},$$

$$P = \frac{\text{\# of correctly recognized zero anaphora}}{\text{\# of system outputted zero anaphora}},$$

$$F = \frac{2}{1/R + 1/P}.$$

Kawahara and Kurohashi's model achieved almost 50% as F-measure against newspaper articles. However, as a result of our experiment against web documents, it achieved only about 20% as F-measure. This may be because anaphoric relations in web documents were not so clear as those in newspaper articles and more difficult to recognize. As to the parameter α, the larger α tended to output more zero anaphora, and the highest F-measure was achieved against $\alpha = 1/2$.

When using $\alpha = 1/2$, there were 72 (=122−50) zero pronouns that were tagged in the corpus and not resolved correctly. Only 12 of them were correctly detected and assigned to a wrong entity, that is, 60 of them were not even detected. Therefore, we can say our recall errors were mainly caused by the low recall of zero pronoun detection.

In order to confirm the effectiveness of generalized examples of case slots and salience score, we also conducted experiments under several conditions. We set $\alpha = 1/2$ in these experiments. The results are shown in Table 6, in which CT means generalized categories, NE means generalized NEs and SS means salience score.

Without using any generalized examples, the F-measure is less than Kawahara and Kurohashi's method, which use similarity to deal with sparseness of case slot examples, and we can confirm the effectiveness of the generalized examples. While generalized categories much improved the F-measure, generalized NEs contribute little. This may be because the NE rate is smaller than common noun rate, and so the effect is limited.

We also confirmed that the salience score filter improved F-measure. Moreover, by using salience score filter, the zero anaphora resolution becomes about ten times faster. This is because the system

Table 6: Experiments under Several Conditions.

CT	NE	SS	R	P	F
		√	.131 (16/122)	.205 (16/78)	.160
	√	√	.164 (20/122)	.247 (20/81)	.197
√		√	.402 (49/122)	.368 (49/133)	.384
√	√		.385 (47/122)	.196 (47/240)	.260
√	√	√	**.410** (50/122)	**.373** (50/134)	**.391**

can avoid checking entities with low salience as antecedent candidates.

4.3 Comparison with Previous Works

We compare our accuracies with (Seki et al., 2002). They achieved 48.9% in precision, 88.2% in recall, and 62.9% in F-measure for zero pronoun detection, and 54.0% accuracy for antecedent estimation on 30 newspaper articles, that is, they achieved about 34% in F-measure for whole zero pronoun resolution. It is difficult to directly compare their results with ours due to the difference of the corpus, but our method achieved 39% in F-measure and we can confirm that our model achieves reasonable performance considering the task difficulty.

5 Conclusion

In this paper, we proposed a probabilistic model for Japanese zero anaphora resolution. By using automatically constructed wide-coverage case frames that include generalized examples and introducing salience score filter, our model achieves reasonable performance against web corpus. As future work, we plan to conduct large-scale experiments and integrate this model to a fully-lexicalized probabilistic model for Japanese syntactic and case structure analysis (Kawahara and Kurohashi, 2006).

References

Iida, Ryu, Kentaro Inui, and Yuji Matsumoto. 2006. Exploiting syntactic patterns as clues in zero-anaphora resolution. In *Proceedings of COLING/ACL 2006*, pages 625–632.

IREX Committee, editor. 1999. *Proc. of the IREX Workshop*.

Isozaki, Hideki and Tsutomu Hirao. 2003. Japanese zero pronoun resolution based on ranking rules and machine learning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 184–191.

Kawahara, Daisuke and Sadao Kurohashi. 2002. Fertilization of Case Frame Dictionary for Robust Japanese Case Analysis. In *Proceedings of the 19th International Conference on Computational Linguistics*, pages 425–431.

Kawahara, Daisuke and Sadao Kurohashi. 2004. Zero pronoun resolution based on automatically constructed case frames and structural preference of antecedents. In *Proceedings of the 1st International Joint Conference on Natural Language Processing (IJCNLP-04)*, pages 334–341.

Kawahara, Daisuke and Sadao Kurohashi. 2006. A fully-lexicalized probabilistic model for japanese syntactic and case structure analysis. In *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*, pages 176–183.

Lappin, Shalom and Herbert J. Leass. 1994. An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4):535–562.

Luo, Xiaoqiang. 2007. Coreference or not: A twin model for coreference resolution. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 73–80.

Mitkov, Ruslan, Richard Evans, and Constantin Orăsan. 2002. A new, fully automatic version of mitkov's knowledge-poor pronoun resolution method. In *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002)*.

Ng, Vincent. 2005. Machine learning for coreference resolution: From local classification to global ranking. In *Proceedings of the 43rd Annual Meeting of the Asssociation for Computational Linguistics*, pages 157–164.

Sasano, Ryohei and Sadao Kurohashi. 2008. Japanese named entity recognition using structural natural language processing. In *Proceedings of the 3rd International Joint Conference on Natural Language Processing (IJCNLP-08)*, pages 607–612.

Seki, Kazuhiro, Atsushi Fujii, and Tetsuya Ishikawa. 2002. A probabilistic method for analyzing Japanese anaphora integrating zero pronoun detection and resolution. In *Proceedings of the 19th International Conference on Computational Linguistics*, pages 911–917.

Soon, Wee Meng, Hwee Tou Ng, and Daniel Chung Yong Lim. 2001. A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27(4):521–544.

The National Language Institute for Japanese Language. 2004. *Bunruigoihyo*. Dainippon Tosho, (In Japanese).

Yang, Xiaofeng, Jian Su, Jun Lang, Ghew Lim Tan, Ting Liu, and Sheng Li. 2008. An entity-mention model for coreference resolution with inductive logic programming. In *Proceedings of ACL-08: HLT*, pages 843–851.