Resolution of Indirect Anaphora in Japanese Sentences Using Examples "X *no* Y (Y of X)"

Masaki Murata **Hitoshi Isahara**
Communications Research Laboratory
588-2, Iwaoka, Nishi-ku, Kobe, 651-2401, Japan
{murata,isahara}@crl.go.jp
TEL: +81-78-969-2181, FAX: +81-78-969-2189

Makoto Nagao
Kyoto University
Sakyo, Kyoto 606-8501, Japan
nagao@pine.kuee.kyoto-u.ac.jp

Abstract

A noun phrase can indirectly refer to an entity that has already been mentioned. For example, "*I went into an old house last night. The roof was leaking badly and ...*" indicates that "*the roof*" is associated with "*an old house*", which was mentioned in the previous sentence. This kind of reference (indirect anaphora) has not been studied well in natural language processing, but is important for coherence resolution, language understanding, and machine translation. In order to analyze indirect anaphora, we need a case frame dictionary for nouns that contains knowledge of the relationships between two nouns but no such dictionary presently exists. Therefore, we are forced to use examples of "X *no* Y" (Y of X) and a verb case frame dictionary instead. We tried estimating indirect anaphora using this information and obtained a recall rate of 63% and a precision rate of 68% on test sentences. This indicates that the information of "X *no* Y" is useful to a certain extent when we cannot make use of a noun case frame dictionary. We estimated the results that would be given by a noun case frame dictionary, and obtained recall and precision rates of 71% and 82% respectively. Finally, we proposed a way to construct a noun case frame dictionary by using examples of "X *no* Y."

1 Introduction

A noun phrase can indirectly refer to an entity that has already been mentioned. For example, "*I went into an old house last night. The roof was leaking badly and ...*" indicates that "*The roof*" is associated with "*an old house*," which has already been mentioned. This kind of reference (indirect anaphora) has not been thoroughly studied in natural language processing, but is important for coherence resolution, language understanding, and machine translation. We propose a method that will resolve the indirect anaphora in Japanese nouns by using the relationship between two nouns.

When we analyze indirect anaphora, we need a case frame dictionary for nouns that contains information about the relationship between two nouns. For instance, in the above example, the knowledge that "roof" is a part of a "house" is required to analyze the indirect anaphora. But no such noun case frame dictionary exists at present. We considered using the example-based method to solve this problem. In this case, the knowledge that "roof" is a part of "house" is analogous to "house of roof." Therefore, we use examples of the form "X of Y" instead. In the above example, we use linguistic data such as "the roof of a house." In the case of verbal nouns, we do not use "X of Y" but a verb case frame dictionary. This is because a noun case frame is similar to a verb case frame and a verb case frame dictionary does exist.

The next section describes a method for resolving indirect anaphora.

2 How to Resolve Indirect Anaphora

Anaphors and antecedents in indirect anaphora have a certain relationship. For example, "*yane* (roof)" and "*hurui ie* (old house)" are in an indirect anaphoric relationship which is a part-of relationship.

sakuban	*aru*	*hurui*	*ie-ni*	*itta.*
(last night)	(a certain)	(old)	(house)	(go)

(I went into an old house last night.)

<u>*yane-wa*</u>	*hidoi*	*amamoride* ...
<u>(roof)</u>	(badly)	(be leaking)

(<u>The roof</u> was leaking badly and ...) (1)

When we analyze indirect anaphora, we need a dictionary containing information about relationships between anaphors and their antecedents.

We show examples of the relationships between anaphors and antecedents in Table 1. The form of Table 1 is similar to the form of a verb case frame dictionary. We would call a dictionary containing the relationships between two nouns *a noun case frame dictionary* but no noun case frame dictionary has yet been created. Therefore, we substitute it with examples of "X *no* Y" (Y of X) and with a verb case frame dictionary. "X *no* Y" is a Japanese expression. It means "Y of X," "Y in X," "Y for X," etc.

Table 1: Relationships between anaphors and their antecedents

Anaphor	Possible antecedents	Relationship
kazoku (family)	*hito* (human)	belong
kokumin (nation)	*kuni* (country)	belong
genshu (the head of state)	*kuni* (country)	belong
yane (roof)	*tatemono* (building)	part of
mokei (model)	*seisanbutsu* (product) [ex. *hikouki* (air plain), *hune* (ship)]	object
gyouji (event)	*soshiki* (organization)	agent
jinkaku (personality)	*hito* (human)	possessive
kyouiku (education)	*hito* (human) *hito* (human) *nouryoku* (ability) [ex. *suugaku* (mathematics)]	agent recipient object
kenkyuu (research)	*hito* (human), *soshiki* (organization) *gakumon bun'ya* (field of study)	agent object
kaiseki (analysis)	*hito* (human), *kikai* (machine) *de-ta* (data)	agent object

Table 2: Case frame of verb "*kaiseki-suru* (analyze)"

Surface case	Semantic constraint	Examples
ga-case (subject)	human	*seito* (student), *kare* (he)
wo-case (object)	abstract, product	*atai* (value), *de-ta* (data)

We resolve the indirect anaphora using the following steps:

1. We detect some elements which could be analyzed by indirect anaphora resolution using "X *no* Y" and a verb case frame dictionary. When a noun was a verbal noun, we use a verb case frame dictionary. Otherwise, we use examples of "X *no* Y."

   For example, in the following example sentences *kaiseki* (analysis) is a verbal noun, and we use a case frame of a verb *kaiseki-suru* (analyze) for the indirect anaphora resolution of *kaiseki* (analysis). The case frame is shown in Table 2. In this table there are two case components, the *ga*-case (subject) and the *wo*-case (object). These two case components are elements which will be analyzed in indirect anaphora resolution.

   denkishingou-no *riyouni-ni yotte*
   (electronic detectors) (use) (by)
   (By using electronic detectors.)

   Butsurigakusha-wa tairyou-no *deeta-wo* (2)
   (physicist) (a large amount) (data)
   shuushuudekiru-youni-natta.
   (collect)
   (physicists had been able to collect large amounts of data.)

   sokode subayai <u>kaiseki</u>-no houhou-ga hitsuyouni-natta.
   (then) (quick) (analysis) (method) (require)
   (Then, they required a method of quick <u>analysis</u>.)

2. We take possible antecedents from topics or foci in the previous sentences. We assign them a certain weight based on the plausibility that they are antecedents. The topics/foci and their weights are defined in Table 3 and Table 4.

   For example, in the case of "*I went into an old house last night. The roof was leaking badly and ...*," "an old house" becomes a candidate of the desired antecedent. In the case of "analysis" in example sentence 2, "electronic detectors," "physicists," and "large amounts of data" become candidates of the two desired antecedents of "analysis." In Table 3 and Table 4 such candidates are given certain weights which indicate preference.

3. We determine the antecedent by combining the weight of topics and foci mentioned in step 2, the weight of semantic similarity in "X *no* Y" or in a verb case frame dictionary, and the weight of the distance between an anaphor and its possible antecedent.

   For example, when we want to clarify the antecedent of *yane* (roof) in example sentence 1, we gather examples of "Noun X *no yane* (roof)" (roof of Noun X), and select a possi-

Table 3: The weight as topic

Surface expression	Example	Weight
Pronoun/zero-pronoun *ga*/*wa*	[John *ga* (subject)] *shita* (done).	21
Noun *wa*/*niwa*	John *wa* (subject) *shita* (done).	20

Table 4: The weight as focus

Surface expression (Not including "*wa*")	Example	Weight
Pronoun/zero-pronoun *wo* (object)/*ni* (to) /*kara* (from)	[John *ni* (to)] *shita* (done).	16
Noun *ga* (subject)/*mo*/*da*/*nara*/*koso*	John *ga* (subject) *shita* (done).	15
Noun *wo* (object)/*ni*/, /.	John *ni* (object) *shita* (done).	14
Noun *he* (to)/*de* (in)/*kara* (from)/*yori*	*gakkou* (school) *he* (to) *iku* (go).	13

ble noun which is semantically similar to Noun X as its antecedent. In example sentence 2, when we want to have an antecedent of *kaiseki* (analysis) we select as its antecedent a possible noun which satisfies the semantic constraint in the case frame of *kuichigau* (differ) in Table 2 or is semantically similar to examples of components in the case frame. In the *ga*-case (subject), of three candidates, "electronic detectors," "physicists," and "large amounts of data," only "physicists" satisfies the semantic constraint, human, in the case frame of the verb *kaiseki-suru* in Table 2. So "physicists" is selected as the desired antecedent of the *ga*-case. In the *wo*-case (object), two phrases, "electronic detectors" and "large amounts of data" satisfy the semantic constraints, abstract and product. By using the examples "value" and "data" in the case frame, the phrase "large amounts of data," which is semantically similar to "data" in the examples of the case frame, is selected as the desired antecedent of the *wo*-case.

We think that errors made by the substitution of a verb case frame for a noun case frame are rare, but many errors occur when we substitute "X *no* Y" for a noun case frame. This is because "X *no* Y" (Y of X) has many semantic relationships, in particular a feature relationship (ex. "a man of ability"), which cannot be an indirect anaphoric relationship. To reduce the errors, we use the following procedures.

1. We do not use an example of the form "Noun X *no* Noun Y" (Y of X), when noun X is an adjective noun [ex. HONTOU (reality)], a numeral, or a temporal noun. For example, we do not use *hontou* (reality) *no* (of) *hannin* (criminal) (a real criminal).
2. We do not use an example of the form "Noun X *no* Noun Y" (Y of X), when noun Y is a noun that cannot be an anaphor of an indirect anaphora. For example, we do not use "Noun X *no tsuru* (crane)," or "Noun X *no ningen* (human being)."

We cannot completely avoid errors by introducing the above procedure, but we can reduce them to a certain extent.

Nouns such as *ichibu* (part), *tonari* (neighbor) and *betsu* (other) need further consideration. When such a noun is a case component of a verb, we use information on the semantic constraints of the verb. We use a verb case frame dictionary as shown in Table 5.

takusan-no	*kuruma-ga*	*kouen-ni*	*tomatte-ita.*
(many)	(car)	(in the park)	(there were)

(There were many cars in the park.)

ichibu-wa	*kita-ni*	*mukatta*
[A part (of them)]	(to the north)	(went)

(A part of them went to the north.)

(3)

In this example, since *ichibu* (part) is a *ga*-case (subject) of a verb *mukau* (go), we consult the *ga*-case (subject) of the case frame of *mukau* (go). Some noun phrases which can also be used in the case component are written in the *ga*-case (subject) of the case frame. In this case, *kare* (he) and *hune* (ship) are written as examples of things which can be used in the case component. This indicates that the antecedent is semantically similar to *kare* (he) and *hune* (ship). Since *takusan no kuruma* (many cars) is semantically similar to *hune* (ship) in the meaning of vehicles, it is judged to be the proper antecedent.

When such a noun as *tonari* (neighbor or next) modifies a noun X as *tonari no* X, we consider the antecedent to be a noun which is similar to noun X

Table 5: Case frame of verb "*mukau*" (go to)

Surface case	Semantic constraint	Examples
ga-case (subject)	concrete	*kare* (he), *hune* (ship)
ni-case (object)	place	*kouen* (park), *minato* (port)

in meaning.

ojiisan-wa ooyorokobi-wo-shite ie-ni kaerimashita.
(the old man) (in great joy) (house) (returned)
[The old man returned home (house) in great joy,]

okotta koto-wo hitobito-ni hanashimashita
(happened to him) (all things) (everybody) (told) (4)
(and told everybody all that had happened to him.)

tonari-no ie-ni ojiisan-ga mouhitori sunde-orimashita.
(next) (house) (old man) (another) (live)
(There lived in the next house another old man.)

For example, when *tonari* (neighbor or next) modifies *ie* (house), we judge that the antecedent of *tonari* (neighbor or next) is *ie* (house) in the first sentence.

3 Anaphora Resolution System

3.1 Procedure

Before starting the anaphora resolution process, the syntactic structure analyzer transforms sentences into dependency structures (Kurohashi and Nagao, 1994). Antecedents are determined by heuristic rules for each noun from left to right in the sentences. Using these rules, our system gives possible antecedents points, and it determines that the possible antecedent having the maximum total score is the desired antecedent. This is because a several types of information are combined in anaphora resolution. An increase in the points of a possible antecedent corresponds to an increase of the plausibility of the possible antecedent.

The heuristic rules are given in the following form:

Condition ⇒ { *Proposal, Proposal, ...* }
Proposal := (*Possible-Antecedent, Point*)

Surface expressions, semantic constraints, referential properties, for example, are written as conditions in the *Condition* part. A possible antecedent is written in the *Possible-Antecedent* part. *Point* refers to the plausibility of the possible antecedent.

To implement the method mentioned in Section 2, we use the weights W of topics and foci, the distance D, the definiteness P, and the semantic similarity S (in R4 of Section 3.2) to determine points. The weights W of topics and foci are given in Table 3 and Table 4 respectively in Section 2, and represent the preferability of the desired antecedent. In this work, a topic is defined as a theme which is described, and a focus is defined as a word which is stressed by the speaker (or the writer). But we cannot detect topics and foci correctly. Therefore we approximated them as shown in Table 3 and Table 4. The distance D is the number of the topics (foci) between the anaphor and a possible antecedent which is a topic (focus). The value P is given by the score of the definiteness in referential property analysis (Murata and Nagao, 1993). This is because it is easier for a definite noun phrase to have an antecedent than for an indefinite noun phrase to have one. The value S is the semantic similarity between a possible antecedent and Noun X of "Noun X *no* Noun Y." Semantic similarity is shown by level in *Bunrui Goi Hyou* (NLRI, 1964).

3.2 Heuristics for determining antecedents

We wrote 15 heuristic rules for noun phrase anaphora resolution. Some of the rules are given below:

R1 When the referential property of a noun phrase (an anaphor) is definite, and the same noun phrase A has already appeared, ⇒
{ (the noun phrase A, 30)}
A referential property is estimated by this method (Murata and Nagao, 1993). This is a rule for direct anaphora.

R2 When the referential property of a noun phrase is generic, ⇒
{ (generic, 10)}

R3 When the referential property of a noun phrase is indefinite, ⇒
{ (indefinite, 10)}

R4 When a noun phrase Y is not a verbal noun, ⇒
{ (A topic which has the weight W and the distance D, $W - D + P + S$),
(A focus which has the weight W and the distance D, $W - D + P + S$),
(A subject in a subordinate clause or a main clause of the clause, $23 + P + S$)
where the values W, D, P, and S are as they were defined in Section 3.1.

R5 When a noun phrase is a verbal noun, ⇒
{ (A topic which satisfies the semantic constraint in a verb case frame and has the weight W and the distance D, $W - D + P + S$),
(A focus which satisfies the semantic constraint and has the weight W and the distance D, $W - D + P + S$),

kono dorudaka-wa *kyoutyou-wo* *gikushaku saseteiru.*
(The dollar's surge) (cooperation) (is straining)
(The dollar's surge is straining the cooperation.)
jikokutuuka-wo *mamorouto* *nisidoku-ga* *kouteibuai-wo* *hikiageta.*
(own currency) (to protect) (West Germany) (official rate) (raised)
(West Germany raised (its) official rate to protect the mark.)

	Indefinite	*nisidoku*	*jikokutuuka*	*kyoutyou*	*dorudaka*
		West Germany	own currency	cooperation	dollar's surge
R3	10				
R4		25	−23	−24	−17
Subject		23			
Topic Focus (W)			14	14	20
Distance (D)			−2	−3	−2
Definiteness (P)		−5	−5	−5	−5
Similarity (S)		7	−30	−30	−30
Total Score	10	25	−23	−24	−17

Examples of "noun X *no kouteibuai* (official rate)"

"*nihon* (Japan) *no kouteibuai* (official rate)",
"*beikoku* (USA) *no kouteibuai* (official rate)"

Figure 1: Example of indirect anaphora resolution

(A subject in a subordinate clause or a main clause of the clause, $23 + P + S$)

R6 When a noun phrase is a noun such as *ichibu*, *tonari*, and it modifies a noun X, $\Rightarrow$
{ (the same noun as the noun X, 30)}

3.3 Example of analysis

An example of the resolution of an indirect anaphora is shown in Figure 1. Figure 1 shows that the noun *koutei buai* (official rate) is analyzed well. This is explained as follows:

The system estimated the referential property of *koutei buai* (official rate) to be indefinite in the method (Murata and Nagao, 1993). Following rule R3 (ection 3.2) the system took a candidate "Indefinite," which means that the candidate is an indefinite noun phrase that does not have an indirect anaphoric referent. Following R4 (Section 3.2) the system took four possible antecedents, *nisidoku* (West Germany), *jikokutuuka (own currency)*, *kyoutyou* (cooperation), *dorudaka* (dollar's surge). The possible antecedents were given points based on the weight of topics and foci, the distance from the anaphor, and so on. The system properly judged that *nisidoku* (West Germany), which had the best score, was the desired antecedent.

4 Experiment and Discussion

Before the antecedents in indirect anaphora were determined, sentences were transformed into a case structure by the case analyzer (Kurohashi and Nagao, 1994). The errors made by the analyzer were corrected by hand. We used the IPAL dictionary (IPAL, 1987) as a verb case frame dictionary. We used the Japanese Co-occurrence Dictionary (EDR, 1995) as a source of examples for "X *no* Y."

We show the result of anaphora resolution using both "X *no* Y" and a verb case frame dictionary in Table 6. We obtained a recall rate of 63% and a precision rate of 68% when we estimated indirect anaphora in test sentences. This indicates that the information of "X *no* Y" is useful to a certain extent even though we cannot make use of a noun frame dictionary. We also tested the system when it did not have any semantic information. The precision and the recall were lower. This indicates that semantic information is necessary. The experiment was performed by fixing all the semantic similarity values S to 0.

We also estimated the results for the hypothetical use of a noun case frame dictionary. We estimated these results in the following manner: We looked over the errors that had occured when we used "X *no* Y" and a verb case frame dictionary. We regarded errors made for one of the following three reasons as right answers:

Table 6: Results

Non-verbal noun		Verbal noun		Total	
Recall	Precision	Recall	Precision	Recall	Precision
Experiment made when the system does not use any semantic information					
85%(56/66)	67%(56/83)	40%(14/35)	44%(14/32)	69%(70/101)	61%(70/115)
53%(20/38)	50%(20/40)	47%(15/32)	42%(15/36)	50% (35/70)	46% (35/76)
Experiment using "X *no* Y" and verb case frame					
91%(60/66)	86%(60/70)	66%(23/35)	79%(23/29)	82%(83/101)	84% (83/99)
63%(24/38)	83%(24/29)	63%(20/32)	56%(20/36)	63% (44/70)	68% (44/65)
Estimation for the hypothetical use of a noun case frame dictionary					
91%(60/66)	88%(60/68)	69%(24/35)	89%(24/27)	83%(84/101)	88% (84/95)
79%(30/38)	86%(30/35)	63%(20/32)	77%(20/26)	71% (50/70)	82% (50/61)

The upper row and the lower row of this table show rates on training sentences and test sentences respectively.
The training sentences are used to set the values given in the rules (Section 3.2) by hand.
Training sentences {example sentences (Walker et al., 1994) (43 sentences), a folk tale *Kobutori jiisan* (Nakao, 1985) (93 sentences), an essay in *Tenseijingo* (26 sentences), an editorial (26 sentences)}
Test sentences {a folk tale *Tsuru no ongaeshi* (Nakao, 1985) (91 sentences), two essays in *Tenseijingo* (50 sentences), an editorial (30 sentences)}
Precision is the fraction of the noun phrases which were judged to have the indirect anaphora as antecedents. *Recall* is the fraction of the noun phrases which have the antecedents of indirect anaphora. We use precision and recall to evaluate because the system judges that a noun which is not an antecedent of indirect anaphora is an antecedent of indirect anaphora, and we check these errors thoroughly.

1. Proper examples do not exist in examples of "X *no* Y" or in the verb case frame dictionary.
2. Wrong examples exist in examples of "X *no* Y" or in the verb case frame dictionary.
3. A noun case frame is different from a verb case frame.

If we were to make a noun case frame dictionary, it would have some errors, and the success ratio would be lower than the ratio shown in Table 6.

Discussion of Errors

Even if we had a noun case frame dictionary, there are certain pairs of nouns in indirect anaphoric relationship that could not be resolved using our framework.

kon'na hidoi hubuki-no naka-wo ittai dare-ga kita-no-ka-to ibukarinagara, obaasan-wa iimashita.
(Wondering who could have come in such a heavy snowstorm, the old woman said:)
"donata-jana"
("Who is it?")
to-wo aketemiruto, soko-niwa zenshin yuki-de masshironi natta musume-ga tatte orimashita.
(She opened the door, and there stood before her a girl all covered with snow.)

(5)

The underlined *musume* has two main meanings: a daughter or a girl. In the above example, *musume* means "girl" and has no indirect anaphora relation but the system incorrectly judged that it is the daughter of *obaasan* (the old woman). This is a problem of noun role ambiguity and is very difficult to solve.

The following example also presents a difficult problem:

shushou-wa	*teikou-no*	*tsuyoi*
(prime minister)	(resistance)	(very hard)
senkyoku-no	*kaishou-wo*	*miokutta.*
(electoral district)	(modification)	(give up)

(The prime minister gave up the modification of some electoral districts where the resistance was very hard.)

(6)

On the surface, the underlined *teikou* (resistance) appears to refer indirectly to *senkyoku* (electoral district). But actually *teikou* (resistance) refers to the candidates of *senkyoku* (electoral district) not to *senkyoku* (electoral district) itself. To arrive at this conclusion, in other words, to connect *senkyoku* (electoral district) and *teikou* (resistance), it is necessary to use a two-step relation, "an electoral district ⇒ candidates," "candidates ⇒ resist" in sequence. It is not easy, however, to change our system so it can deal with two-step relationships. If we apply the use of two-step relationships to nouns, many nouns which are not in an indirect anaphoric rela-

Table 7: Examples of arranged "X *no* Y"

Noun Y	Arranged noun X
kokumin (nation)	<Human> *aite* (partner) <Organization> *kuni* (country), *senshinkoku* (an advanced country), *ryoukoku* (the two countries), *naichi* (inland), *zenkoku* (the whole country), *nihon* (Japan), *soren* (the Soviet Union), *eikoku* (England), *amerika* (America), *suisu* (Switzerland), *denmaaku* (Denmark), *sekai* (the world)
genshu (the head of state)	<Human> *raihin* (visitor) <Organization> *gaikoku* (a foreign country), *kakkoku* (each country), *poorando* (Poland)
yane (roof)	<Organization> *hokkaido* (Hokkaido), *sekai* (the world), *gakkou* (school), *koujou* (factory), *gasorinsutando* (gas station), *suupaa* (supermarket), *jitaku* (one's home), *honbu* (the head office) <Product> *kuruma* (car), *juutaku* (housing), *ie* (house), *shinden* (temple), *genkan* (entrance), *shinsha* (new car) <Phenomenon> *midori* (green) <Action> *kawarabuki* (tile-roofed) <Mental> *houshiki* (method) <Character> *keishiki* (form)
mokei (model)	<Animal> *zou* (elephant) <Nature> *fujisan* (Mt. Fuji) <Product> *imono* (an article of cast metal), *manshon* (an apartment house), *kapuseru* (capsule), *densha* (train), *hune* (ship), *gunkan* (warship), *hikouki* (airplane), *jettoki* (jet plane) <Action> *zousen* (shipbuilding) <Mental> *puran* (plan) <Character> *unkou* (movement)
gyouji (event)	<Human> *koushitsu* (the Imperial Household), *oushitsu* (a Royal family), *iemoto* (the head of a school) <Organization> *nouson* (an agricultural village), *ken* (prefecture), *nihon* (Japan), *soren* (the Soviet Union), *tera* (temple), *gakkou* (school) <Action> *shuunin* (take up one's post), *matsuri* (festival), *iwai* (celebration), *junrei* (pilgrimage) <Mental> *kourei* (an established custom), *koushiki* (formal)
jinkaku (personality)	<Human> *watashi* (myself), *ningen* (human), *seishounen* (young people), *seijika* (statesman)

tion will be incorrectly judged as indirect anaphora. A new method is required in order to infer two relationships in sequence.

5 Consideration of Construction of Noun Case Frame Dictionary

We used "X *no* Y" (Y of X) to resolve indirect anaphora. But we would achieve get a higher accuracy rate if we could utilize a good noun case frame dictionary. Therefore we have to consider how to construct a noun case frame dictionary. A key is to get the detailed meaning of "*no* (of)" in "X *no* Y." If it is automatically obtainable, a noun case frame dictionary could be constructed automatically. Even if the semantic analysis of "X *no* Y" is not done well, we think that it is still possible to construct the dictionary using "X *no* Y." For example, we arrange "noun X *no* noun Y" by the meaning of "noun Y," arrange them by the meaning of "noun X", delete those where "noun X" is an adjective noun, and obtain the results shown in Table 7. In this case, we use the thesaurus dictionary "*Bunrui Goi Hyou*" (NLRI, 1964) to learn the meanings of nouns. It should not be difficult to construct a noun case frame dictionary by hand using Table 7. We will make a noun case frame dictionary by removing *aite* (partner) in the line of *kokumin* (nation), *raihin* (visitor) in the line of *genshu* (the head of state), and noun phrases which mean characters and features. When we look over the noun phrases for *kokumin* (nation), we notice that almost all of them refer to countries. So we will also make the semantic constraint (or the semantic preference) that countries can be connected to *kokumin* (nation). When we make a noun case frame dictionary, we must remember that examples of "X *no* Y" are insufficient and we must add examples. For example, in the line of *genshu* (the head of state) there are few nouns that mean countries. In this case, it is good to add examples by from the arranged nouns for *kokumin* (nation), which is similar to *genshu* (the head of state). Since in this method examples are arranged by meaning in this method, it will not be very difficult to add examples.

6 Conclusion

We presented how to resolve indirect anaphora in Japanese nouns. We need a noun case frame dictionary containing information about noun relations to analyze indirect anaphora, but no such dictionary exists at present. Therefore, we used examples of "X *no* Y" (Y of X) and a verb case frame dictionary. We estimated indirect anaphora by using this information, and obtained a recall rate of 63% and a precision rate of 68% on test sentences. This indicates that information about "X *no* Y" is useful when we cannot make use of a noun case frame dictionary. We estimated the results that would be given by a noun case frame dictionary, and obtained recall and precision rates of 71% and 82% respectively. Finally, we proposed a way to construct a noun case frame dictionary by using examples of "X *no* Y."

References

Carter, D. 1987. Interpreting Anaphors in Natural Language Texts. Halsted Press.

Electronic Dictionary Research Institute, LTD. 1995. EDR Electronic Dictionary, Japanese Co-occurrence Dictionary, Version 1.5, (in Japanese).

Information-technology Promotion Agency, Japan. 1987. IPA Lexicon of the Japanese Language for Computers IPAL (Basic Verbs), (in Japanese).

Kurohashi, S., Nagao, M. 1994. A Method of Case Structure Analysis for Japanese Sentences based on Examples in Case Frame Dictionary. *The Institute of Electronics, Information and Communication Engineers Transactions on Information and Systems E77–D (2)*, pages 227–239.

Murata, M., Nagao, M. 1993. Determination of referential property and number of nouns in Japanese sentences for machine translation into English. In *Proceedings of the 5th TMI*, pages 218–225, Kyoto, Japan, July.

Murata, M., Nagao, M. 1995. An Estimate of Referents of Pronouns in Japanese Sentences using Examples and Surface Expressions, (in Japanese) IPSJ 95-NL-108, Hukuoka, Japan, July.

Nakao, K. 1985. The Old Man with a Wen. Eiyaku Nihon Mukashibanashi Series, Vol. 7, Nihon Eigo Kyouiku Kyoukai (in Japanese).

National Language Research Institute (1964) *Word List by Semantic Principles.* Syuei Syuppan (in Japanese).

Walker, M., Iida, M., Cote, S. 1994. Japanese Discourse and the Process of Centering Computational Linguistics, Vol. 20, No. 2, pages 193–232.

Watanabe, Y., Kurohashi, S., Nagao, M. 1992. Construction of semantic dictionary by IPAL dictionary and a thesaurus (in Japanese). In *Proceedings of the 45th Convention of IPSJ*, pages 213–214, Tokushima, Japan, July.