Automatic Construction of Nominal Case Frames and its Application to Indirect Anaphora Resolution

Ryohei Sasano, Daisuke Kawahara and **Sadao Kurohashi**
Graduate School of Information Science and Technology, University of Tokyo
{ryohei,kawahara,kuro}@kc.t.u-tokyo.ac.jp

Abstract

This paper proposes a method to automatically construct Japanese nominal case frames. The point of our method is the integrated use of a dictionary and example phrases from large corpora. To examine the practical usefulness of the constructed nominal case frames, we also built a system of indirect anaphora resolution based on the case frames. The constructed case frames were evaluated by hand, and were confirmed to be good quality. Experimental results of indirect anaphora resolution also indicated the effectiveness of our approach.

1 Introduction

What is represented in a text has originally a network structure, in which several concepts have tight relations with each other. However, because of the linear constraint of texts, most of them disappear in the normal form of texts. Automatic reproduction of such relations can be regarded as the first step of "text understanding", and surely benefits NLP applications such as machine translation, automatic abstraction, and question answering.

One of such latent relationship is indirect anaphora, functional anaphora, or bridging reference, such as the following examples.

(1) I bought a ticket. The price was 20 dollars.
(2) There was a house. The roof was white.

Here, "the price" means "the price of a ticket" and "the roof" means "the roof of a house."

Most nouns have their indispensable or requisite entities: "price" is a price of some goods or service, "roof" is a roof of some building, "coach" is a coach of some sport, and "virus" is a virus causing some disease. The relation between a noun and its indispensable entity is parallel to that between a verb and its arguments or obligatory cases. In this paper, we call indispensable entities of nouns *obligatory cases*. Indirect anaphora resolution needs a comprehensive information or dictionary of obligatory cases of nouns.

In case of verbs, syntactic structures such as subject/object/PP in English or case markers such as *ga*, *wo*, *ni* in Japanese can be utilized as a strong clue to distinguish several obligatory cases and adjuncts (and adverbs), which makes it feasible to construct case frames from large corpora automatically (Briscoe and Carroll, 1997; Kawahara and Kurohashi, 2002). (Kawahara and Kurohashi, 2004) then utilized the automatically constructed case frames to Japanese zero pronoun resolution.

On the other hand, in case of nouns, obligatory cases of noun N_h appear, in most cases, in the single form of noun phrase "N_h of N_m" in English, or "N_m *no* N_h" in Japanese. This single form can express several obligatory cases, and furthermore optional cases, for example, "*rugby no coach*" (obligatory case concerning what sport), "*club no coach*" (obligatory case concerning which institution), and "*kyonen* 'last year' *no coach*" (optional case). Therefore, the key issue to construct nominal case frames is to analyze "N_h of N_m" or "N_m *no* N_h" phrases to distinguish obligatory case examples and others.

Work which addressed indirect anaphora in English texts so far restricts relationships to a small, relatively well-defined set, mainly part-of relation like the above example (2), and utilized hand-crafted heuristic rules or hand-crafted lexical knowledge such as WordNet (Hahn et al., 1996; Vieira and Poesio, 2000; Strube and Hahn, 1999). (Poesio et al., 2002) proposed a method of acquiring lexical knowledge from "N_h of N_m" phrases, but again concentrated on part-of relation.

In case of Japanese text analysis, (Murata et al., 1999) proposed a method of utilizing "N_m *no* N_h" phrases for indirect anaphora resolution of diverse relationships. However, they basically used all "N_m *no* N_h" phrases from corpora, just excluding some pre-fixed stop words. They confessed that an accurate analysis of "N_m *no* N_h" phrases is necessary for the further improvement of indirect anaphora resolution.

As a response to these problems and following the work in (Kurohashi and Sakai, 1999), we propose a method to construct Japanese nominal case frames from large corpora, based on an accurate analysis of "N_m *no* N_h" phrases using an ordinary dictionary and a thesaurus. To examine the practical usefulness of the constructed nominal case frames, we also built a system of indirect anaphora resolution based on the case frames.

2 Semantic Feature Dictionary

First of all, we briefly introduce NTT Semantic Feature Dictionary employed in this paper. NTT Semantic Feature Dictionary consists of a semantic feature tree, whose 3,000 nodes are semantic features, and a nominal dictionary containing about 300,000 nouns, each of which is given one or more appropriate semantic features.

The main purpose of using this dictionary is to calculate the similarity between two words. Suppose the word x and y have a semantic feature s_x and s_y, respectively, their depth is d_x and d_y in the semantic tree, and the depth of their lowest (most specific) common node is d_c, the similarity between x and y, $sim(x, y)$, is calculated as follows:

$$sim(x, y) = (d_c \times 2)/(d_x + d_y).$$

If s_x and s_y are the same, the similarity is 1.0, the maximum score based on this criteria.

We also use this dictionary to specify semantic category of words, such as HUMAN, TIME and PLACE.

3 Semantic Analysis of Japanese Noun Phrases N_m *no* N_h

In many cases, obligatory cases of nouns are described in an ordinary dictionary for human being. For example, a Japanese dictionary for children, *Reikai Shougaku Kokugojiten*, or RSK (Tajika, 1997), gives the definitions of the word *coach* and *virus* as follows[1]:

coach a person who teaches technique in some sport

virus a living thing even smaller than bacteria which causes infectious disease like influenza

[1]Although our method handles Japanese noun phrases by using Japanese definition sentences, in this paper we use their English translations for the explanation. In some sense, the essential point of our method is language-independent.

Based on such an observation, (Kurohashi and Sakai, 1999) proposed a semantic analysis method of "N_m *no* N_h", consisting of the two modules: dictionary-based analysis (abbreviated to DBA hereafter) and semantic feature-based analysis (abbreviated to SBA hereafter). This section briefly introduces their method.

3.1 Dictionary-based analysis

Obligatory case information of nouns in an ordinary dictionary can be utilized to solve the difficult problem in the semantic analysis of "N_m *no* N_h" phrases. In other words, we can say the problem disappears.

For example, "*rugby no coach*" can be interpreted by the definition of *coach* as follows: the dictionary describes that the noun *coach* has an obligatory case *sport*, and the phrase "*rugby no coach*" specifies that the *sport* is *rugby*. That is, the interpretation of the phrase can be regarded as matching *rugby* in the phrase to *some sport* in the *coach* definition. "*Kaze* 'cold' *no virus*" is also easily interpreted based on the definition of *virus*, linking *kaze* 'cold' to *infectious disease*.

Dictionary-based analysis (DBA) tries to find a correspondence between N_m and an obligatory case of N_h by utilizing RSK and NTT Semantic Feature Dictionary, by the following process:

1. Look up N_h in RSK and obtain the definition sentences of N_h.
2. For each word w in the definition sentences other than the genus words, do the following steps:
   - 2.1. When w is a noun which shows an obligatory case explicitly, like *kotogara* 'thing', *monogoto* 'matter', *nanika* 'something', and N_m does not have a semantic feature of HUMAN or TIME, give 0.8 to their correspondence[2].
   - 2.2. When w is other noun, calculate the similarity between N_m and w by using NTT Semantic Feature Dictionary, and give the similarity score to their correspondence.
3. Finally, if the best correspondence score is 0.75 or more, DBA outputs the best correspondence, which can be an obligatory case of the input; if not, DBA outputs nothing.

[2]For the present, parameters in the algorithm were given empirically, not optimized by a learning method.

Table 1: Examples of rules for semantic feature-based analysis.

1. N_m:HUMAN, N_h:RELATIVE → <obligatory case(relative)>	e.g. *kare* 'he' *no oba* 'aunt'
2. N_m:HUMAN, N_h:HUMAN → <modification(apposition)>	e.g. *gakusei* 'student' *no kare* 'he'
3. N_m:ORGANIZATION, N_h:HUMAN → <belonging>	e.g. *gakkou* 'school' *no seito* 'student'
4. N_m:AGENT, N_h:EVENT → <agent>	e.g. *watashi* 'I' *no chousa* 'study'
5. N_m:MATERIAL, N_h:CONCRETE → <modification(material)>	e.g. *ki* 'wood' *no hako* 'box'
6. N_m:TIME, N_h:∗ → <time>	e.g. *aki* 'autumn' *no hatake* 'field'
7. N_m:COLOR, QUANTITY, or FIGURE, N_h:∗ → <modification>	e.g. *gray no seihuku* 'uniform'
8. N_m:∗, N_h:QUANTITY → <obligatory case(attribute)>	e.g. *hei* 'wall' *no takasa* 'height'
9. N_m:∗, N_h:POSITION → <obligatory case(position)>	e.g. *tsukue* 'desk' *no migi* 'right'
10. N_m:AGENT, N_h:∗ → <possession>	e.g. *watashi* 'I' *no kuruma* 'car'
11. N_m:PLACE or POSITION, N_h:∗ → <place>	e.g. *Kyoto no mise* 'store'

'∗' meets any noun.

In case of the phrase "*rugby no coach*", "technique" and "sport" in the definition sentences are checked: the similarity between "technique" and "rugby" is calculated to be 0.21, and the similarity between "sport" and "rugby" is calculated to be 1.0. Therefore, DBA outputs "sport".

3.2 Semantic feature-based analysis

Since diverse relations in "N_m *no* N_h" are handled by DBA, the remaining relations can be detected by simple rules checking the semantic features of N_m and/or N_h.

Table 1 shows examples of the rules. For example, the rule 1 means that if N_m has a semantic feature HUMAN and N_h RELATIVE, <obligatory case> relation is assigned to the phrase. The rules 1, 2, 8 and 9 are for certain obligatory cases. We use these rules because these relations can be analyzed more accurately by using explicit semantic features, rather than based on a dictionary.

3.3 Integration of two analyses

Usually, either DBA or SBA outputs some relation. When both DBA and SBA output some relations, the results are integrated (basically, if DBA correspondence score is higher than 0.8, DBA result is selected; if not, SBA result is selected). In rare cases, neither analysis outputs any relations, which means analysis failure.

4 Automatic Construction of Nominal Case Frames

4.1 Collection and analysis of N_m *no* N_h

Syntactically unambiguous noun phrases "N_m *no* N_h" are collected from the automatic parse results of large corpora, and they are analyzed using the method described in the previous section.

Table 2: Preliminary case frames for *hisashi* 'eaves/visor'.

DBA result	
1. a roof that stick out above the window of a house.	
[house]	hall:2, balcony:1, building:1, ···
[window]	window:2, ceiling:1, counter:1, ···
2. the fore piece of a cap.	
[cap]	cap:8, helmet:1, ···
SBA result	
<place>	parking:3, store:3, shop:2, ···
<mod.>	concrete:1, metal:1, silver:1, ···
No semantic analysis result	
<other>	part:1, light:1, phone:1, ···

By just collecting the analysis results of each head word N_h, we can obtain its preliminary case frames. Table 2 shows preliminary case frames for *hisashi* 'eaves/visor'. The upper part of the table shows the results by DBA. The line starting with "[house]" denotes a group of analysis results corresponding to the word "house" in the first definition sentence. For example, "hall *no* hisashi" occurs twice in the corpora, and they were analyzed by DBA to correspond to "house."

The middle part of the table shows the results by SBA. Noun phrases that have no semantic analysis result (analysis failure) are bundled and named <other>, as shown in the last part of the table.

A case frame should be constructed for each meaning (definition) of N_h, and groups starting with "[...]" or "<...>" in Table 2 are possible case slots. The problem is how to arrange the analysis results of DBA and SBA and how to distinguish obligatory cases and others. The following sections explain how to handle these problems.

Table 3: Threshold to select obligatory slots.

type of case slots	threshold of probability
analyzed by DBA	0.5% (1/200)
<obligatory case>	2.5% (1/40)
<belonging>	2.5% (1/40)
<possessive>	5% (1/20)
<agent>	5% (1/20)
<place>	5% (1/20)
<other>	10% (1/10)
<modification>	not used
<time>	not used

Probability = (# of N_m *no* N_h) / (# of N_h)

4.2 Case slot clustering

One obligatory case might be separated in preliminary case frames, since the definition sentence is sometimes too specific or too detailed. For example, in the case of *hisashi* 'eaves/visor' in Table 2, [house], [window], and <place> have very similar examples that mean building or part of building. Therefore, case slots are merged if similarity of two case slots is more than 0.5 (case slots in different definition sentences are not merged in any case). Similarity of two case slots is the average of top 25% similarities of all possible pairs of examples.

In the case of Table 2, the similarity between [house] and [window] is 0.80, and that between [house] and <place> is 0.67, so that these three case slots are merged into one case slot.

4.3 Obligatory case selection

Preliminary case frames contain both obligatory cases and optional cases for the head word. Since we can expect that an obligatory case co-occurs with the head word in the form of noun phrase frequently, we can take frequent case slots as obligatory case of the head word.

However, we have to be careful to set up the frequency thresholds, because case slots detected by DBA or <obligatory case> by SBA are more likely to be obligatory; on the other hand case slots of <modification> or <time> should be always optional. Considering these tendencies, we set thresholds for obligatory cases as shown in Table 3.

In the case of *hisashi* 'eaves/visor' in Table 2, [house-window]-<place> slot and [cap] slot are chosen as the obligatory cases.

4.4 Case frame construction for each meaning

Case slots that are derived from each definition sentence constitute a case frame.

If a case slot of <obligatory case> by SBA or <other> is not merged into case slots in definition sentences, it can be considered that it indicates a meaning of N_h which is not covered in the dictionary. Therefore, such a case slot constitutes an independent case frame.

On the other hand, when other case slots by SBA such as <belonging> and <possessive> are remaining, we have to treat them differently. The reason why they are remaining is that they are not always described in the definition sentences, but their frequent occurrences indicate they are obligatory cases. Therefore, we add these case slots to the case frames derived from definition sentences.

Table 4 shows several examples of resultant case frames. *Hyoujou* 'expression' has a case frame containing two case slots. *Hisashi* 'eaves/visor' has two case frames according to the two definition sentences. In case of *hikidashi* 'drawer', the first case frame corresponds to the definition given in the dictionary, and the second case frame was constructed from the <other> case slot, which is actually another sense of *hikidashi*, missed in the dictionary. In case of *coach*, <possessive> is added to the case frame which was made from the definition, producing a reasonable case frame for the word.

4.5 Point of nominal case frame construction

The point of our method is the integrated use of a dictionary and example phrases from large corpora. Although dictionary definition sentences are informative resource to indicate obligatory cases of nouns, it is difficult to do indirect anaphora resolution 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. Furthermore, sometimes a definition is too specific or detailed, and the example phrases can adjust it properly, as in the example of *hisashi* in Table 2.

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) can not construct comprehensive nominal case frames, because of polysemy and multiple obligatory cases. We can see that dictionary definition can guide the clustering properly even for such difficult cases.

Table 4: Examples of nominal case frames.

	case slot	examples
hisashi:1 'eaves/visor'	(the edges of a roof that stick out above the window of a house etc.)	
	[house, window]	parking, store, hall, $\cdots$
hisashi:2 'eaves/visor'	(the fore piece of a cap.)	
	[cap]	cap, helmet, $\cdots$
hyoujou 'expression'	(to express one's feelings on the face or by gestures.)	
	[one]	people, person, citizen, $\cdots$
	[feelings]	relief, margin, $\cdots$
hikidashi:1 'drawer'	(a boxlike container in a desk or a chest.)	
	[desk, chest]	desk, chest, dresser, $\cdots$
hikidashi:2 'drawer'	<other>	credit, fund, saving, $\cdots$
coach	(a person who teaches technique in some sport.)	
	[sport]	baseball, swimming, $\cdots$
	<belonging>	team, club, $\cdots$
kabushiki 'stock'	(the total value of a company's shares.)	
	[company]	company, corporation, $\cdots$

5 Indirect Anaphora Resolution

To examine the practical usefulness of the constructed nominal case frames, we built a preliminary system of indirect anaphora resolution based on the case frames.

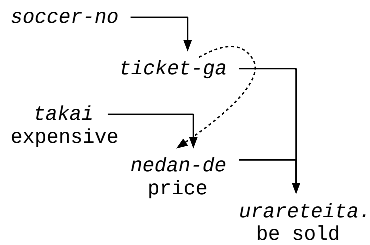
An input sentence is parsed using the Japanese parser, KNP (Kurohashi and Nagao, 1994). Then, from the beginning of the sentence, each noun x is analyzed. When x has more than one case frame, the process of antecedent estimation (stated in the next paragraph) is performed for each case frame, and the case frame with the highest similarity score (described below) and assignments of antecedents to the case frame are selected as a final result.

For each case slot of the target case frame of x, its antecedent is estimated. A possible antecedent y in the target sentence and the previous two sentences is checked. This is done one by one, from the syntactically closer y. If the similarity of y to the case slot is equal to or greater than a threshold α (currently 0.95), it is assigned to the case slot.

The similarity between y and a case slot is defined as the highest similarity between y and an example in the case slot.

For instance, let us consider the sentence shown in Figure 1. *soccer*, at the beginning of the sentence, has no case frame, and is considered to have no obligatory case.

For the second noun *ticket*, *soccer*, which is a nominal modifier of *ticket*, is examined first. The similarity between *soccer* and the examples of the case slot [theater, transport] exceeds the threshold α, and *soccer* is assigned to [theater, transport].



	case slot	examples	result
ticket	[theater, transport]	stage, game,$\cdots$	*soccer*
nedan	[things]	thing, ticket,$\cdots$	*ticket*

ticket a printed piece of paper which shows that you have paid to enter a theater or use a transport
nedan the amount of money for which things are sold or bought

Figure 1: Indirect anaphora resolution example.

Lastly, for *nedan* 'price', its possible antecedents are *ticket* and *soccer*. *ticket*, which is the closest from *nedan*, is checked first. The similarity between *ticket* and the examples of the case slot [things] exceeds the threshold α, and *ticket* is judged as the antecedent of *nedan*.

6 Experiments

We evaluated the automatically constructed nominal case frames, and conducted an experiment of indirect anaphora resolution.

6.1 Evaluation of case frames

We constructed nominal case frames from newspaper articles in 25 years (12 years of *Mainichi* newspaper and 13 years of *Nihonkeizai* newspaper). These newspaper corpora consist of about

Table 5: Evaluation result of case frames.

precision	recall	F
58/70 (0.829)	58/68 (0.853)	0.841

25,000,000 sentences, and 10,000,000 "N_m *no* N_h" noun phrases were extracted from them. The result consists of 17,000 nouns, the average number of case frames for a noun is 1.06, and the average number of case slots for a case frame is 1.09.

We randomly selected 100 nouns that occur more than 10,000 times in the corpora, and created gold standard case frames by hand. For each test noun, possible case frames were considered, and for each case frame, obligatory case slots were given manually. As a result, 68 case frames for 65 test nouns were created, and 35 test nouns have no case frames.

We evaluated automatically constructed case frames for these test nouns against the gold standard case frames. A case frame which has the same case slots with the gold standard is judged as correct. The evaluation result is shown in Table 5: the system output 70 case frames, and out of them, 58 case frames were judged as correct.

The recall was deteriorated by the highly restricted conditions in the example collection. For instance, *maker* does not have obligatory case slot for its products. This is because *maker* is usually used in the form of compound noun phrase, "PRODUCTS *maker*", and there are few occurrences of "PRODUCTS *no maker*". To address this problem, not only "N_m *no* N_h" but also "N_m N_h" (compound noun phrase) and "N_m *ni-kansuru* 'in terms of' N_h" should be collected.

6.2 Experimental results of indirect anaphora resolution

We conducted a preliminary experiment of our indirect anaphora resolution system using "Relevance-tagged corpus" (Kawahara et al., 2002). This corpus consists of Japanese newspaper articles, and has relevance tags, including antecedents of indirect anaphors.

We prepared a small test corpus that consists of randomly selected 10 articles. The test corpus contains 217 nouns. Out of them, 106 nouns are indirect anaphors, and have 108 antecedents, which is because two nouns have double antecedents. 49 antecedents directly depend on their anaphors, and 59 do not. For 91 antecedents out of 108, a case frame of its anaphor includes the antecedent itself or its similar word (the similarity exceeds the threshold, 0.95). Accordingly, the upper bound of the recall of our case-frame-based anaphora resolution is 84.3% (91/108).

Table 6: Experimental results of indirect anaphora resolution.

	precision	recall	F
w dep.	40/46 (0.870)	40/59 (0.678)	0.762
w/o dep.	31/61 (0.508)	31/49 (0.633)	0.564
total	71/107 (0.664)	71/108 (0.657)	0.660

We ran the system on the test corpus, and compared the system output and the corpus annotation. Table 6 shows the experimental results. In this table, "w dep." (with dependency) is the evaluation of the antecedents that directly depend on their anaphors. "w/o dep." (without dependency) is the case of the antecedents that do not directly depend on their anaphors.

Although the analysis of "w dep." is intrinsically easier than that of "w/o dep.", the recall of "w dep." was not much higher than that of "w/o dep.". The low recall score of "w dep." was caused by nonexistence of case frames which include the antecedent itself or its similar word. The antecedents that directly depend on their anaphors were often a part of compound noun phrases, such as "PRODUCTS *maker*", which are not covered by our examples collection.

Major errors in the analyses of the antecedents that do not directly depend on their anaphors were caused by the following reasons.

Specific/generic usages of nouns
Some erroneous system outputs were caused by nouns that have both specific and generic usages.

(3) *kogaisya-no* *kabushiki-wo* *baikyaku-shita.*
subsidiary stock sell

(ϕ sold the stock of the subsidiary.)

In this case, *kogaisya* 'subsidiary' is an obligatory information for *kabushiki* 'stock', which is specifically used. *kogaisya* matches the [*kaisya* 'company'] case slot in Table 4.

However, *kabushiki* 'stock' in the following example is used generically, and does not need specific company information.

(4) *kabushiki* *souba-no* *oshiage* *youin-to* *naru.*
stock price rise factor become

(ϕ become the rise factor of the stock prices.)

Since the current system cannot judge generic or specific nouns, an antecedent which corresponds to [*kaisha* 'company'] is incorrectly estimated.

Beyond selectional restriction of case frames
Selectional restriction based on the case frames usually worked well, but did not work to distinguish candidates both of which belong to HUMAN or ORGANIZATION.

(5) *Bush bei*   *seiken-wa*   *Russia-tono*
American administration
... *Bush* *daitouryou-ga* *shutyou-shita.*
president   claim

(Bush American administration ... with Russia ... President Bush claimed ...)

In this example, *daitouryou* 'president' requires an obligatory case *kuni* 'nation'. The system estimates its antecedent as *Russia*, though the correct answer is *bei* 'America'. This is because *Russia* is closer than *beikoku*. This problem is somehow related to world knowledge, but if the system can carefully exploit the context, it might be able to find the correct answer from "*Bush bei seiken*" 'Bush American administration'.

7 Conclusion

This paper has first proposed an automatic construction method of Japanese nominal case frames. This method is based on semantic analysis of noun phrases "N_m *no* N_h" 'N_h of N_m'. To examine the practical usefulness of the constructed nominal case frames, we built a preliminary system of indirect anaphora resolution based on the case frames. The evaluation indicated the good quality of the constructed case frames. On the other hand, the accuracy of our indirect anaphora resolution system is not satisfactory. In the future, we are planning to make the case frames more wide-coverage, and improve the indirect anaphora resolution by considering larger context and more various factors.

References

Ted Briscoe and John Carroll. 1997. Automatic extraction of subcategorization from corpora. In *Proceedings of the 5th Conference on Applied Natural Language Processing*, pages 356–363.

Udo Hahn, Michael Strube, and Katja Markert. 1996. Bridging textual ellipses. In *Proceedings of the 16th International Conference on Computational Linguistics*, pages 496–501.

Daisuke Kawahara 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.

Daisuke Kawahara 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.*

Daisuke Kawahara, Sadao Kurohashi, and Kôiti Hasida. 2002. Construction of a Japanese relevance-tagged corpus. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, pages 2008–2013.

Sadao Kurohashi and Makoto Nagao. 1994. A syntactic analysis method of long Japanese sentences based on the detection of conjunctive structures. *Computational Linguistics*, 20(4):507–534.

Sadao Kurohashi and Yasuyuki Sakai. 1999. Semantic analysis of Japanese noun phrases: A new approach to dictionary-based understanding. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 481–488.

Masaki Murata, Hitoshi Isahara, and Makoto Nagao. 1999. Pronoun resolution in Japanese sentences using surface expressions and examples. In *Proceedings of the ACL'99 Workshop on Coreference and Its Applications*, pages 39–46.

Massimo Poesio, Tomonori Ishikawa, Sabine Schulte im Walde, and Renata Vieira. 2002. Acquiring lexical knowledge for anaphora resolution. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, pages 1220–1224.

Michael Strube and Udo Hahn. 1999. Functional centering – grounding referential coherence in information structure. *Computational Linguistics*, 25(3):309–344.

Jun-ichi Tajika, editor. 1997. *Reikai Syogaku Kokugojiten.* Sanseido.

Renata Vieira and Massimo Poesio. 2000. An empirically based system for processing definite descriptions. *Computational Linguistics*, 26(4):539–592.