Approaches to Zero Adnominal Recognition

Mitsuko Yamura-Takei
Graduate School of Information Sciences
Hiroshima City University
Hiroshima, JAPAN
`yamuram@nlp.its.hiroshima-cu.ac.jp`

Abstract

This paper describes our preliminary attempt to automatically recognize zero adnominals, a subgroup of zero pronouns, in Japanese discourse. Based on the corpus study, we define and classify what we call "argument-taking nouns (ATNs)," i.e., nouns that can appear with zero adnominals. We propose an ATN recognition algorithm that consists of lexicon-based heuristics, drawn from the observations of our analysis. We finally present the result of the algorithm evaluation and discuss future directions.

1 Introduction

(1) Zebras always need to watch out for lions. Therefore, even while eating grass, so that able to see behind, eyes are placed at face-side.

This is a surface-level English translation of a naturally occurring "unambiguous" Japanese discourse. By "unambiguous," we mean that Japanese speakers find no difficulty in interpreting this discourse segment, including whose eyes are being talked about. Moreover, Japanese speakers find this segment quite "coherent," even though there seems to be no surface level indication of who is eating or seeing, or whose eyes are being mentioned in this four-clause discourse segment.[1] However, this is not always the case with Japanese as a Second Language (JSL) learners.[2]

What constitutes "coherence" has been studied by many researchers. Reference is one of the linguistic devices that create textual unity, i.e., cohe-

sion (Halliday and Hasan, 1976). Reference also contributes to the semantic continuity and content connectivity of a discourse, i.e., coherence. Coherence represents the natural and reasonable connections between utterances that make for easy understanding, and thus lower inferential load for hearers.

The Japanese language uses ellipsis as its major type of referential expression. Certain elements are elided when they are recoverable from a given context or from relevant knowledge. These ellipses may include verbals and nominals; the missing nominals have been termed "zero pronouns," "zero pronominals," "zero arguments," or simply "zeros" by researchers.

How many zeros are contained in (1), for example, largely depends on how zeros are defined. In the literature, zeros are usually defined as elements recoverable from the valency requirements of the predicate with which they occur. However, does this cover all the zeros in Japanese? Does this explain all the content connectivity created by nominal ellipsis in Japanese?

In this paper, we introduce a subgroup of zeros, what we call "zero adnominals," in contrast to other well-recognized "zero arguments" and investigate possible approaches to recognizing these newly-defined zeros, in an attempt to incorporate them in an automatic zero detecting tool for JSL teachers that aims to promote effective instruction of zeros. In section 2, we provide the definition of zero adnominals, and present the results of their manual identification in the corpus. Section 3 describes the theoretical and pedagogical motivations for this study. Section 4 illustrates the syntactic/semantic classification of the zero adnominal examples found in the corpus. Based on the classification results, we propose lexical information-based heuristics, and present a preliminary evaluation. In the final two sections, we present related work, and discuss possible future directions.

---

[1] This was verified by an informal poll conducted on 15 native speakers of Japanese.
[2] Personal communication with a JSL teacher.

2 Zero Adnominals

2.1 Definition

Recall the discourse segment in (1). Its original Japanese is analyzed in (2).

(2) a. *simauma-wa raion ni itumo*
   zebra-TOP lion-DAT always
   ki-o-tuke-nakereba-narimasen.
   watch-out-for-need-to
   "Zebras always need to watch out for lions."

   b. *desukara, Ø kusa-o tabete-ite-mo,*
   so   **Ø-NOM** grass-ACC eating-even-while
   "So even while (they) are eating grass,"

   c. *Ø Ø usiro-no-ho-made   mieru-yo-ni*
   Ø-NOM Ø-ADN-behind-even see-can-for
   "so that (they) can see even what is behind (them),"

   d. *Ø me-ga   Ø kao-no-yoko-ni*
   Ø-ADN-eye-NOM Ø-ADN-face-side LOC
   tuite-imasu.
   placed-be
   "(their)eyes are on the sides of (their) faces."

Zero arguments are unexpressed elements that are predictable from the valency requirements of their heads, i.e., a given predicate of the clause. Zero nominatives in (2b) and (2c) are of this type. Zero adnominals, analogously, are missing elements that can be inferred from some features specified by their head nouns. A noun for body-part, *me* 'eyes' in (2d) usually calls hearers' attention to "of-whom" information and hearers recover that information in the flow of discourse. That missing information can be supplied by a noun phrase (NP) followed by an adnominal particle *no*, i.e., *si-mauma-no* 'zebras'(= their)' in the case of (2d) above. Hence, as a first approximation, we define a zero adnominal as an unexpressed "NP *no*" in the NP *no* NP (a.k.a., A *no* B) construction.

2.2 The Corpus

Before we proceed, we will briefly describe the corpus that we investigated. The corpus consists of a collection of 83 written narrative texts taken from seven different JSL textbooks with levels ranging from beginning to intermediate. Thus, it is a representative sample of naturally-occurring, but maximally canonical, free-from-deviation, and coherent narrative discourse.

2.3 Identification

Our primary goal is to identify relevant information for recognizing zero adnominals. Since such information is unavailable in the surface text, the identification of missing adnominal elements and their referents in the corpus was based on the native speaker intuitions and the linguistic expertise of the author, who used the definition in 2.1, with occasional consultation with a JSL teaching expert/linguist. As a result, we located a total of 320 zero adnominals. These adnominals serve as the zero adnominal samples on which our later analysis is based.

3 Theoretical/Pedagogical Motivations

3.1 Centering Analysis

One discourse account that models the perceived degree of coherence of a given discourse in relation to local focus of attention and the choice of referring expressions is centering (e.g., Grosz, Joshi and Weinstein, 1995).

The investigation of zeros behavior in our corpus, within the centering framework, shows that zero adnominals make a considerable contribution to center continuity in discourse by realizing the central entity in an utterance (called Cb) just as well-acknowledged zero arguments do.

Recall example (2). Its center data structure is given in (3). The Cf (forward-looking center) list is a set of discourse entities that appear in each utterance (U_i). The Cb (backward-looking center) is a special member of the Cf list, and is meant to represent the entity that the utterance is most centrally about; it is the most highly ranked element of the Cf (U_{i-1}) that is realized in U_i.

(3) a. Cb: none   [Cf: zebra, lion]
   b. Cb: zebra   [Cf: zebra, grass]
   c. Cb: zebra   [Cf: zebra, what is behind]
   d. Cb: zebra   [Cf: zebra, eye, face-side]

In (3b) and (3c), the Cb is realized as a zero nominative, and in (3d), it is realized by the same entity (zebra) as a zero adnominal, maintaining the

CONTINUE transition that by definition is maximally coherent. This matches the intuitively perceived degree of coherence in the utterance. Our corpus contains a total of 138 zero adnominals that refer to previously mentioned entities (15.56% of all the zero Cbs), and realize the Cb of the utterance in which they occur, as in (3d=2d).

Our corpus study shows that discourse coherence can be more accurately characterized, in the centering account, by recognizing the role of zero adnominals as a valid realization of Cbs (see Yamura-Takei *et al*., ms. for detailed discussion). This is our first motivation towards zero adnominal recognition.

3.2 Zero Detector

Yamura-Takei *et al*. (2002) developed an automatic zero identifying tool. This program, Zero Detector (henceforth, ZD) takes Japanese written narrative texts as input and provides the zero-specified texts and their underlying structures as output. This aims to draw learners' and teachers' attention to zeros, on the basis of a hypothesis about ideal conditions for second language acquisition, by making invisible zeros visible. ZD regards teachers as its primary users, and helps them predict the difficulties with zeros that students might encounter, by analyzing text in advance. Such difficulties often involve failure to recognize discourse coherence created by invisible referential devices, i.e., the center continuity maintained by the use of various types of zeros.

As our centering analysis above indicates, inclusion of zero adnominals into ZD's detecting capability enables a more comprehensive coverage of the zeros that contributes to discourse coherence. This is our project goal.

4 Towards Zero Adnominal Recognition

4.1 Semantic Classification

Unexpressed elements need to be predicted from other expressed elements. Thus, we need to characterize B nouns (which are overt) in the (A *no*) B construction, assuming that zero adnominals (A) are triggered by their head nouns (B) and that certain types of NPs tend to take implicit (A) arguments. Our first approach is to use an existing A *no* B classification scheme. We adopted, from among many A *no* B works, a classification modeled on Shimazu, Naito and Nomura (1985, 1986, and 1987) because it offers the most comprehensive classification (Fais and Yamura-Takei, ms). Table 1 below describes the five main groups that we used to categorize (A *no*) B phrases.

4.2 Results

We classified our 320 "(A *no*) B" examples into the five groups described in the previous section. Group V comprised the vast majority, while approximately the same percentage of examples was included in Groups I, II and III. There were no Group IV examples. The number and percentage of examples of each group are presented in Table 2.

Group	# of examples
I	33 (10.31%)
II	23 (7.19%)
III	35 (10.94%)
IV	0 (0.00%)
V	229 (71.56%)
Total	320 (100%)

Table 2: Distribution of semantic types

Group #	Definition	Example from Shimazu *et al*. (1986)
I	A: argument B: nominalized verbal element	*kotoba no rikai* 'word-*no*-understanding'
II	A: noun denoting an entity B: abstract relational noun	*biru no mae* 'building-*no*-front'
III	A: noun denoting an entity B: abstract attribute noun	*hasi no nagasa* 'bridge-*no*-length'
IV	A: nominalized verbal element B: argument	*kenka no hutari* 'argument-*no*-two people'
V	A: noun expressing attribute B: noun denoting an entity	*ningen no atama* 'human-*no*-head'

Table 1: (A no) B classification scheme

We conjecture that certain nouns are more likely to take zero adnominals than others, and that the head nouns which take zero adnominals, extracted from our corpus, are representative samples of this particular group of nouns. We call them "argument-taking nouns (ATNs)." ATNs syntactically require arguments and are semantically dependent on their arguments. We use the term ATN only to refer to a particular group of nouns that can take implicit arguments (i.e., zero adnominals).

We closely examined the 127 different ATN tokens among the 320 cases of zero adnominals and classified them into the four types that correspond to Groups I, II, III and V in Table 1. We then listed their syntactic/semantic properties based on the syntactic/semantic properties presented in the *Goi-Taikei Japanese Lexicon* (hereafter GT, Ikehara, Miyazaki, Shirai, Yokoo, Nakaiwa, Ogura, Oyama, and Hayashi, 1997). GT is a semantic feature dictionary that defines 300,000 nouns based on an ontological hierarchy of approximately 2,800 semantic attributes. It also uses nine part-of-speech codes for nouns. Table 3 lists the syntactic/semantic characterizations of the nouns in each type and the number of examples in the corpus. What bold means in the table will be explained later in section 4.3.

Type	Syntactic properties	Semantic properties	#	Examples
I	Nominalized verbal, derived (from verb) noun, common noun	Human activity	21	*zikosyokai* 'self-introduction'
		phenomenon	3	*entyo* 'extension'
II	formal noun, common noun	**Location**	**13**	*mae* 'front'
		Time	1	*yokuzitu* 'next day'
III	Derived (from verb/adjective) noun, suffix noun, common noun	**Amount**	**9**	*sintyo* 'height'
		Value	2	*nedan* 'price'
		Emotion	1	*kimoti* 'feeling'
		Material phenomenon	1	*nioi* 'smell'
		Name	1	*namae* 'name'
		Order	1	*ichiban* 'first'
V	Common noun	**Human (kinship)**	**14**	*haha* 'mother'
		Animate (body-part)	**14**	*atama* 'head'
		Organization	**7**	*kaisya* 'company'
		Housing (part)	**7**	*doa* 'door'
		Human (profession)	**4**	*sensei* 'teacher'
		Human (role)	**4**	*dokusya* 'reader'
		Human (relationship)	**3**	*dooryoo* 'colleague'
		Clothing	**3**	*kutu* 'shoes'
		Tool	2	*saihu* 'purse'
		Human (biological feature)	2	*zyosei* 'woman'
		Man-made	2	*kuruma* 'car'
		Facility	1	*byoin* 'hospital'
		Building	1	*niwa* 'garden'
		Housing (body)	1	*gareeji* 'garage'
		Housing (attachment)	1	*doa* 'door'
		Creative work	1	*sakuhin* 'work'
		Substance	1	*kuuki* 'air'
		Language	1	*nihongo* 'Japanese'
		Document	1	*pasupooto* 'passport'
		Chart	1	*chizu* 'map'
		Animal	1	*petto* 'pet'
		? (unregistered)	2	*hoomusutei* 'homestay'
	Total		127	

Table 3: Subtypes of ATNs

When we examine these four types, we see that they partially overlap with some particular types of nouns studied theoretically in the literature. Tera-mura (1991) subcategorizes locative relational nouns like *mae* 'front', *naka* 'inside', and *migi* 'right' as "incomplete nouns" that require elements to complete their meanings; these are a subset of Type II. Iori (1997) argues that certain nouns are categorized as "one-place nouns," in which he seems to include Type I and some of Type V nouns. Kojima (1992) examines so-called "low-independence nouns" and categorizes them into three types, according to their syntactic behaviors in Japanese copula expressions. These cover sub-sets of our Type I, II, III and V. In computational work, Bond, Ogura, and Ikehara (1995) extracted 205 "trigger nouns" from a corpus aligned with English. These nouns trigger the use of possessive pronouns when they are machine-translated into English. They seem to correspond mostly to our Type V nouns. Our result offers a comprehensive coverage which subsumes all of the types of nouns discussed in these accounts.

Next, let us more closely look at the properties expressed by our samples. The most prevalent ATNs (21 in number) are nominalized verbals in the semantic category of human activity. The next most common are kinship nouns (14 in number) and body-part nouns (14), both in the common noun category; location nouns (13), either in the common noun or formal noun category; and nouns that express amount (9) whose syntactic category is either common or de-adjectival. The others in-clude some "human" subcategories, etc.

The part-of-speech subcategory, "nominalized verbal" (*sahen-meishi*) is a reasonably accurate indicator of Type 1 nouns. So is "formal noun" (*keishiki-meishi*) for Type II, although this does not offer a full coverage of this type. Numeral noun and counter suffix noun compounds also represent a major subset of Type III.

Semantic properties, on the other hand, seem helpful to extract certain groups such as location (Type II), amount (Type III), kinship, body-part, organization, and some human subcategories (Type V). But other low-frequency ATN samples are problematic for determining an appropriate level of categorization in GT's semantic hierarchy tree.

4.3 Algorithm

Our goal is to build a system that can identify the presence of zero adnominals. In this section, we propose an ATN (hence zero adnominal) recogni-tion algorithm. The algorithm consists of a set of lexicon-based heuristics, drawn from the observa-tions in section 4.2.

The algorithm takes morphologically-analyzed text as input and provides ATN candidates as out-put. The process consists of the following three phases: (i) bare noun extraction, (ii) syntactic cate-gory (part-of-speech) checking, and (iii) semantic category checking.

Zero adnominals usually co-occur with "bare nouns." Bare nouns, in our definition, are nouns without any pre-nominal modifiers, including de-monstratives, explicit adnominal phrases, relative clauses, and adjectives.[3] Bare nouns are often sim-plex as in (4a), and sometimes are compound (e.g., numeral noun + counter suffix noun) as in (4b). These are immediately followed by case-marking, topic/focus-marking or other particles (e.g., *ga, o, ni, wa, mo*).

(4) a. ***atama**-ga*   head-NOM
     b. *70-**paasento**-o* 70-percent-ACC

The extracted nouns under this definition are initial candidates for ATNs.

Once bare nouns are identified, they are checked against our syntactic-property- (i.e., part-of-speech, POS) based-, followed by semantic-attribute (SEM) based-heuristics. For semantic filtering, we decided to use the noun groups of high frequency (more than two tokens categorized in the same group; indicated in bold in Table 3 above) to minimize a risk of over-generalization.

The algorithm checks the following two condi-tions, for each bare noun, in this order:

[1] If POS = [nominalized verval, derived noun, formal noun, numeral + counter suffix com-pound], label it as ATN.

[2] If SEM = [2610: location, 2585: amount, 362: organization, 552: animate (part), 111: hu-man (relation), 224: human (profession), 72:

human (kinship), 866: housing (part), 813: clothing], label it as ATN. [4]

Therefore, nouns that pass condition [1] are labeled as ATNs, without checking their semantic properties. A noun that fails to pass condition [1] and passes condition [2] is labeled as ATN. A noun that fails to match both [1] and [2] is labeled as non-ATN. Consider the noun *sintyo* 'height' for example. Its POS code in GT is common noun, so it fails condition [1] and goes to [2]. This noun is categorized in the "2591: measures" group which is under the "2585: amount" node in the hierarchy tree, so it is labeled as ATN. In this way, the algorithm labels each bare noun as either ATN or non-ATN.

4.4 Evaluation

To assess the performance of our algorithm, we ran it by hand on a sample text.[5] The test corpus contains a total of 136 bare nouns. We then matched the result against our manually-extracted ATNs (34 in number). The result is shown in Table 4 below, with recall and precision metrics. As a baseline measurement, we give the accuracy for classifying every bare noun as ATN. For comparison, we also provide the results when only either POS-based or semantic-based heuristics are applied.

	Recall	Precision
Baseline	34/34 (100%)	34/136 (25.00%)
POS only	2/34 (5.88%)	2/6 (33.33%)
Semantic only	30/34 (88.23%)	30/35 (85.71%)
POS/Semantic	32/34 (94.11%)	32/41 (78.04%)

Table 4: Algorithm evaluation

Semantic categories make a greater contribution to identifying ATNs than POS. However, the POS/Semantic algorithm achieved a higher recall but a lower precision than the semantic-only algorithm did. This is mainly because the former produced more over-detected errors. Closer examination of those errors indicates that most of them (8 out of 9 cases) involve verbal idiomatic expressions that contain ATN candidate nouns, as example (5) shows.

(5) *me-o-samasu* eye-ACC-wake 'wake up'

Although *me* 'eye' is a strong ATN candidate, as in example (2) above, case (5) should be treated as part of an idiomatic expression rather than as a zero adnominal expression.[6] Thus, we decided to add another condition, [0] below, before we apply the POS/SEM checks. The revised algorithm is as follows:

[0] If part of idiom in [idiom list],[7] label it as non-ATN.

[1] If POS = [nominalized verval, derived noun, formal noun, numeral + counter suffix compound], label it as ATN.

[2] If SEM = [2610: location, 2585: amount, 362: organization, 552: animate (part), 111: human (relation), 224: human (profession), 72: human (kinship), 866: housing (part), 813: clothing], label it as ATN.

When a noun matches condition [0], it will not be checked against [1] and [2]. When this applies, the evaluation result is now as shown below.

	Recall	Precision
POS only	2/34 (5.88%)	2/4 (50.00%)
Semantic only	30/34 (88.23%)	31/35 (88.57%)
POS/Semantic	32/34 (94.11%)	32/33 (96.96%)

Table 5: Revised-algorithm evaluation

The revised algorithm, with both syntactic/semantic heuristics and the additional idiom-filtering rule, achieved a precision of 96.96%. The result still includes some over/under-detecting errors, which will require future attention.

5 Related Work

Associative anaphora (e.g., Poesio and Vieira, 1998) and indirect anaphora (e.g., Murata and Nagao, 2000) are virtually the same phenomena that this paper is concerned with, as illustrated in (6).

---

[4] These numbers indicate the numbers assigned to each semantic category in *Goi-Taikei Japanese Lexicon* (GT).
[5] This is taken from the same genre as our corpus for the initial analysis, i.e., another JSL textbook.

[6] Vieira and Poesio (2000) also list "idiom" as one use of definite descriptions (English equivalent to Japanese bare nouns), along with same head/associative anaphora, etc.
[7] The list currently includes eight idiomatic samples from the test data, but it should of course be expanded in the future.

(6) a. a house – the roof
   b. *ie* 'house' – *yane* 'roof'
   c. *ie* 'house' – (*Ø-no*) *yane* '(Ø's) roof'

We take a zero adnominal approach, as in (6c), because we assume, for our pedagogical purpose discussed in section 3.2, that zero adnominals, by making them visible, more effectively prompt people to notice referential links than lexical relations, such as meronymy in (6a) and (6b).

However, insights from other approaches are worth attention. There is a strong resemblance between bare nouns (that zero adnominals co-occur with) in Japanese and definite descriptions in English in their behaviors, especially in their referential properties (Sakahara, 2000). The task of classifying several different uses of definite descriptions (Vieira and Poesio, 2000; Bean and Riloff, 1999) is somewhat analogous to that for bare nouns. Determining definiteness of Japanese noun phrases (Heine, 1998; Bond *et al*., 1995; Murata and Nagao, 1993)[8] is also relevant to ATN (which is definite in nature) recognition.

6   Future Directions

We have proposed an ATN (hence zero adnominal) recognition algorithm, with lexicon-based heuristics that were inferred from our corpus investigation. The evaluation result shows that the syntactic/semantic feature-based generalization (using GT) is capable of identifying potential ATNs. The evaluation on a larger corpus, of course, is essential to verify this claim. Implementation of the algorithm is also in our future agenda.

This approach has its limitations, too, as is pointed out by Kurohashi *et al*. (1999). One limitation is illustrated by a pair of Japanese nouns, *sakusya* 'author' and *sakka* 'writer,' which fall under the same GT semantic property group (at the deepest level).[9] These nouns have an intuitively different status for their valency requirements; the former requires "of-what work" information, while the latter does not.[10] We risk over- or undergeneration when we designate certain semantic properties, no matter how fine-grained they might

be. We proposed the idiom-filtering rule to solve one case of over-detection. A larger-scale evaluation of the algorithm and its error analysis might lead to additional rules that refine extracted ATN candidates. Insights from the works presented in the previous section could also be incorporated.

Determining an appropriate level of generalization is a significant factor for this type of approach, and this was done, in this study, according to our introspective judgments. More systematic methods should be explored.

A related issue is the notoriously hard-to-define argument-adjunct distinction for nouns, which is closely related to the distinction between ATNs and non-ATNs. We experimentally tested seven native-Japanese-speaking subjects in distinguishing these two. We presented 26 nouns in the same GT semantic category (at the deepest level): "persons who write." There were six nouns which all the subjects agreed on categorizing as ATNs, including *sakusha* 'author.' Five nouns, including *sakka* 'writer,' on the other hand, were judged as non-ATNs by all the subjects. For the remaining 15 nouns, however, their judgments varied widely. As Somers (1984) suggests for verbs, binary distinction does not work well for nouns, either. This distinction might largely depend on the context in some cases. This is also something we will need to address.

In this study, we focused on "implicit argument-taking nouns." There may be a line (although it may be very thin) between nouns which take explicit arguments and those which take implicit arguments. This distinction also needs further investigation in the corpus.

---

[8] Their interests are in machine-translation of Japanese into languages that require determiners for their nouns.

[9] This example pair is taken from Iori (1997).

[10] This intuition was verified by an informal poll conducted on seven native speakers of Japanese.

References

Bean, David L. and Ellen Riloff. 1999. Corpus-based identification of non-anaphoric noun phrases. In *Proceedings of the 37th Annual Meeting of the ACL*, 373-380.

Bond, Francis, Kentaro Ogura, and Satoru Ikehara. 1995. Possessive pronouns as determiners in Japanese-to-English machine translation. In *Proceedings of the 2nd Pacific Association for Computational Linguistics conference.*

Bond, Francis, Kentaro Ogura, and Tsukasa Kawaoka. 1995. Noun phrase reference in Japanese-to-English machine translation. In *Proceedings of the 6th International Conference on Theoretical and Methodological Issues in Machine Translation*, 1-14.

Fais, Laurel and Mitsuko Yamura-Takei (under review). Salience ranking in centering: The case of a Japanese complex nominal. ms.

Grosz, Barbara J., Aravind Joshi, and Scott Weinstein. 1995. Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2), 203-225.

Halliday, M.A.K. and Ruqaiya Hasan. 1976. *Cohesion in English*. Longman, New York.

Heine, Julia E. 1998. Definiteness prediction for Japanese noun phrases. In *Proceedings of the COLING/ACL'98,* Quebec, 519-525.

Ikehara, Satoru, Masahiro Miyazaki, Satoshi Shirai, Akio Yokoo, Hiromi Nakaiwa, Kentarou Ogura, and Yoshifumi Oyama, editors. 1997. *Goi-Taikei – Japanese Lexicon.* Iwanami Publishing, Tokyo.

Iori, Isao. 1997. *Aspects of Cohesion in Japanese Texts*. Unpublished PhD dissertation, Osaka University (in Japanese).

Kojima, Sachiko. 1992. Low-independence nouns and copula expressions. In *IPA Technical Report* No. 3-125, 175-198 (in Japanese).

Kurohashi and Sakai. 1999. Semantic analysis of Japanese noun phrases: A new approach to dictionary-based understanding. In *Proceedings of the 37th Annual Meeting of the ACL*, 481-488.

Murata, Masaki and Makoto Nagao. 1993. Determination of referential property and number of nouns in Japanese sentences for machine translation into English. In *Proceedings of the 5th International Conference on Theoretical and Methodological Issues in Machine Translation*, 218-225.

Murata, Masaki and Makoto Nagao. 2000. Indirect reference in Japanese sentences. In Botley, S. and McEnerry, A. (eds.) *Corpus-based and Computational Approaches to Discourse Anaphora*, 189-212. John Benjamins, Amsterdam/Philadelphia.

Poesio, Massimo and Renata Vieira. 1998. A corpus-based investigation of definite description use. *Computational Linguistics*, 24(2): 183-216.

Sakahara, Shigeru. 2000. *Advances in Cognitive Linguistics*. Hituzi Syobo Publishing, Tokyo, Japan (in Japanese).

Shimazu, Akira, Shozo Naito, and Hirosato Nomura. 1985. Classification of semantic structures in Japanese sentences with special reference to the noun phrase (in Japanese). In *Information Processing Society of Japan, Natural Language Special Interest Group Technical Report*, No. 47-4.

Shimazu, Akira, Shozo Naito, and Hirosato Nomura. 1986. Analysis of semantic relations between nouns connected by a Japanese particle "no." *Mathematical Linguistics*, 15(7), 247-266 (in Japanese).

Shimazu, Akira, Shozo Naito, and Hirosato Nomura. 1987. Semantic structure analysis of Japanese noun phrases with adnominal particles. In *Proceedings of the 25th Annual Meeting of the ACL*, Stanford, 123-130.

Somers, Harold L. 1984. On the validity of the complement-adjunct distinction in valency grammar. *Linguistics* 22, 507-53.

Teramura, Hideo. 1991. *Japanese Syntax and Meaning II*. Kurosio Publishers, Tokyo (in Japanese).

Vieira, Renata and Massimo Poesio. 2000. An empirically based system for processing definite descriptions. *Computational Linguistics*, 26(4): 525-579.

Yamura-Takei, Mitsuko, Laurel Fais, Miho Fujiwara and Teruaki Aizawa. 2003. Forgotten referential links in Japanese discourse and centering. ms.

Yamura-Takei, Mitsuko, Miho Fujiwara, Makoto Yoshie, and Teruaki Aizawa. 2002. Automatic linguistic analysis for language teachers: The case of zeros. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, Taipei, 1114-1120.