Grammar for ellipsis resolution in Japanese

Shigeko Nariyama
The University of Melbourne, Australia
shigeko@unimelb.edu.au

Abstract

This paper elucidates the linguistic mechanisms for resolving ellipsis (zero anaphor). The mechanisms consist of three tiers of linguistic system. [1] Japanese sentences are structured in such a way to anchor the topic, which is predominantly the subject (by Sentence devices), [2] with argument inferring cues on the verbal predicate (by Predicate devices), and [3] are cohesively sequenced with the topic as a pivot (by Discourse devices). This topicalised subject is most prone to ellipsis. This agrees with the fact that over 90% of ellipsis is of topicalised subjects. An interplay of these mechanisms constitute the keys to ellipsis resolution. I have developed an algorithm capturing the mechanisms. The initial hand-simulated test based on a set of narrative texts has shown that the algorithm achieves a high level of accuracy, over 85%.

1. Introduction

In Japanese-to-English machine translation, it has been widely recognised as a challenge that Japanese frequently unexpresses nominal arguments (ellipses), such as the subject and the object, which must be identified and made explicit in order to be translated into grammatical English. Consider the following Japanese sentence (taken from PHP magazine 2.1999. p.29).

(1) *Tuma ni hanasu to, douisite kureta.*
wife-to-talk-when, agree-gave
'When (I) talked to (my) wife, (she) agreed with (me)'.

Literally this sentence is 'wife-to-talk-when-agree-gave'. English has to reconstruct 'I', 'my', 'she' and 'me' which are not expressed explicitly in the Japanese sentence. This problem of recovering unexpressed arguments is common to the translation of many Asian languages to English.
In order to see the current treatment of the problem, observe the following translations for (1) which came out from some of the existing machine translation systems on the market.

[Human translation] 'When (I) talked to (my) wife, (she) agreed with (me)'.
[MT system 1] ALT J/E 'When (it) was spoken, (it) had the kindness to agree with (one's) wife.'
[MT system 2] Lycos '(It) will be agreed, if talked to the wife.'
[MT system 3] LogoVista '(It) agreed when (it) talked to the wife.'
[MT system 4] Inforseek '(It) will have agreed, if talked to the wife.'

The main reason why these unacceptable sentences are generated is their inability to recover unexpressed arguments and the current systems lack implementations of linguistic rules governing ellipsis resolution. English sentences cannot be adequately generated without knowing the referent, particularly the subject, of the sentence, and it is this subject that is often absent in Japanese.

In this paper, I explicate the linguistic mechanisms with which to identify the referent of ellipsis from a linguistic view point. I have developed an algorithm, based on studies of these linguistic mechanisms. The initial hand-simulated test based on a small set of narrative texts has shown that the algorithm achieves an unprecedented level of high accuracy, over 85%. It looks not only at sentence-level grammar but also at inter-sentential information – contextual information gained from previous sentences.

2. Earlier treatments

Notable studies on ellipsis resolution in Japanese are Kameyama (1985) and a series of work led by Nakaiwa (1995, 1998, inter alia). While they have made great contributions to ellipsis resolution, there still remain shortcomings.

The strength of Kameyama's account is that it integrates two theories to deal with complex phenomena of ellipsis resolution: Lexical Functional Grammar (Bresnan 1982) to account for grammatical aspects and Centering Theory (Grosz, Joshi & Weistein 1983) for discourse aspects with the notion of topic to better deal with Japanese ellipsis (Kameyama 1985). However, the process utilising the two theories was not exemplified by her. The main weakness of her account is that it is not equipped to adequately account for: [1] complex sentences, which comprise 87.5% of sentences in written narratives (Nariyama 2000), [2] cataphors, due to the linear nature attributed to Centering Theory, and [3] multiple argument ellipsis (more than one ellipsis per clause). In more recent work, Kameyama (1998) proposed an account for intrasentential Centering by breaking a complex sentence into a hierarchy of center-updating units, that is, more or less clauses in more general terms. My assumption on her account is to utilise conjunctive particles, by which a hierarchy of center-updating units for each complex sentence is determined. However, this account requires additional convoluted hierarchies and yet its results were shown to be still inadequate by Strube (1998). Even with the potential increase in the accuracy, this method still retains an unsolved problem – resolving non-subject ellipsis. Moreover, it is designed for English complex sentences, and implications for Japanese are not addressed in her work.

In terms of work by Nakaiwa, it requires more grammatical input. It utilises only three grammatical constraints: verbal semantics, conjunctions and modal expressions. They provide some keys to ellipsis resolution. However, as will be explained in the next section, they can be overridden under certain grammatical environments. Furthermore, the three constraints are all placed on the verbal predicates. When the cue to ellipsis resolution is not on the predicate, but on the subject, as demonstrated by the following minimal pair of sentences,[1] it is not unequivocal how Nakaiwa's account can differentiate the meaning of sentences.

(2a) *Taro-**wa** nihon-ni kaette kara, ø hataraita.*
-TopSB Japan-to return after (SB) worked
'After Taro returned to Japan, (he) worked.'

(2b) *Taro-**ga** nihon-ni kaette kara, ø hataraita.*
-SB Japan-to return after (SB) worked
'After Taro returned to Japan, (someone else) worked.'

[1] The following abbreviations are used in the examples: ø=ellipsis, IO=indirect object, NP=noun phrase, OB=object, Obl=oblique (arguments other than the subject and the object), SB=subject, TopSB =topicalised subject.

Furthermore, although intersentential ellipsis resolution is purported by Nakaiwa, precise grammar for it has not been addressed. I will demonstrate in the next section that the three constraints used in Nakaiwa's account constitute only one tier of grammar, and that there are two other tiers to ellipsis resolution.

3. Linguistic mechanisms

The mechanisms for ellipsis resolution consist of three tiers of linguistic devices, and are presented in the following subsections:

[1] ***Predicate devices*** (word level):
morphological signal about the type of subject on the predicates
[2] ***Sentence devices*** (sentence level): rules as to how a sentence should be structured
[3] ***Discourse devices*** (inter-sentence level): rules as to how a sentence should be structured following another

3.1. Predicate devices

Predicate devices work like a subject-verb agreement signalling the referential identity of subjects on the predicate (verbs, adjectives and nominal predicates), not in terms of person, number and gender commonly found in Romance languages, but in terms of four elements: 1) Verbal semantics, 2) Switch-reference, 3) Epistemic morphemes, 4) Honorifics. These are explained below.

3.1.1. Verbal semantics

Verbal semantics provide information about valency (how many arguments/NPs a verb takes and what grammatical functions each argument has) and selectional restrictions (what semantic attributes each argument has). In addition, Japanese has rich morphology to distinguish transitive verbs [Vt] from intransitive verbs [Vi] (see Jacobsen 1992), which provides extra information. For example, the verbal semantics of *atumeru* [Vt] 'to gather' requires three arguments: the subject (human/organization), the object (human/organization, concrete object, evidence), and the oblique (place/location), while its intransitive counterpart *atumaru* [Vi] selects two arguments: the subject (human/organization, concrete object, evidence) and the oblique (place/location). 'Goitaikei' - A Japanese valency dictionary (Ikehara et al. 1997) lists such information in detail. It can distinguish 2700 types of semantic attributes for common noun and 130 types of proper noun on the basis of the verbal semantics.

3.1.2. Switch-reference

Switch-reference is a syntactic mechanism used to indicate whether the subject of a dependent clause is the same as [SS] or different from [DS] the subject of the matrix clause in complex sentences (Stirling 1993). The following Japanese conjunctive particles are reported as switch-reference markers (Iwasaki 1993, Nariyama in press, inter alios):

SS markers: *-nagara, -te, -si, -tutu, -ø, tameni*
DS markers: *-to, -tara, -ga, -node, yooni*

The following set of examples show that the SS conjunctive particle *nagara* in (3) signals that the subjects of the two clauses are the same, while the DS conjunctive particle *to* in (4) signals that the subjects of the two clauses are different (coreference is noted by subscripts):

SS (3) **ø** *Terebi-o mi nagara, haha-ga naitei-ta.*
(SB) TV-OB watch while[SS], my mother-SB crying-Past
'My mother$_i$ was crying, while (she$_i$) watched TV.'

DS (4) **ø** *Miru to, haha-ga naitei-ta.*
(SB) look when[DS], my mother-SB crying-Past
'My mother$_i$ was crying, when (someone$_j$) looked at (her$_i$).'

Minami (1974) makes an analogous statement without the linguistic term 'Switch-reference'. Based on Minami, the SS/DS distinction derived from the conjunctive particles is utilised as one constraint by Nakaiwa and a number of other researchers in the field. However, it is also noted that the SS/DS distinction is valid approximately 60 to 90% depending on the particle (Iwasaki 1993, Nariyama in press); a similar statement was made by Minami (1974:130).

3.1.3. Epistemic morphemes

Japanese has a rigid grammatical constraint on subjective statements, in that the speaker is only in a position to assert his own feelings and thoughts, but he cannot describe feelings about others in the same way, due to lack of direct/adequate evidence. This constraint, thus, distinguishes first person from non-first person, for example:

(5) ø *Uresii*.
(SB) happy
'(**I/*someone**) am happy.'

(6) ø *Uresi-soo-da*.
(SB) happy-look-be
'(**Someone**) looks happy (to my observation)'

Nakaiwa refers this pheneomenon as modal expressions. Although epistemic morphemes provide some cues to ellipsis resolution, the problems lie in the fact that they do not appear in every sentence, and that they can be overridden in subordinate clauses, past tense, and empathy phenomena (see Nariyama 2001: 357, Nariyama 2000).

3.1.4. Honorifics

Japanese utilises honorifics which distinguish whether the subject is shown respect, as in (7), or non-subject is, as in (8), in terms of the social categories of referents in relation to the speaker (Harada 1976, inter alios), for example:

(7) ø *Mesiagaru*. [subject honorific]
'(Honouree) eats.'

(8) ø *Itadaku.* [non-subject honorific]
'(Honourer) eats (something offered by honouree.)'

In summary, Predicate devices consist of: 1) Verbal semantics (applicable for all sentences), 2) Switch-reference (applicable for some conjunctive particles with 60-90% predictability in complex sentences), 3) Epistemic morphemes (applicable for some sentences with conditions), 4) Honorifics (applicable for some sentences).

3.2. Sentence devices

Nariyama (2000, 2001) demonstrated that Japanese sentences are structured in such a way to express an argument high in person/animacy and discourse salience (topicality) as the subject (the principle of direct alignment); namely, a subject must be higher than non-subject arguments (SB>nonSB) in terms of the two constraints. This subject is most prone to ellipsis (the principle of ellipsis). Thus, the two principles jointly claim that the topicalised subject is most prone to ellipsis. Although this has been well accepted as correct, no literature has previously elucidated the grammatical reasons for the phenomenon.

Due to the limited space in this paper, the below I note the statistics based on a corpus analysis as the substantiation for the principles. They correctly reflect the fact that 93.5% of 216 ellipted arguments are subjects.

Figure 1: Proportion of ellipsis

(Nariyama 2001: 8 written narrative texts from PHP magazines, n=216)

Subject ellipsis [ø nonSB V]	**93.5%**	(c.f. 91.3%, Nakaiwa et al. 1995b, n=3718)
Non-subject ellipsis	6.5%	
(with SB ellipsis [ø ø V]	4.6%)	
(without SB ellipsis [SB ø V]	1.9%)	
	100.0%	

The remaining (6.5%) were non-subject ellipses, most of which are objects. Only 1.9% were the cases where the subject was overt and the non-subject argument was ellipted, i.e. [SB ø V]. [SB ø V] happens when the subject is focused, often marked by *ga*. A focused argument cannot be ellipted, since it conveys important and/or new information (Kuno 1987).

3.3. Discourse devices

Discourse devices specify how a sentence should be structured following another. In other words, the subject oriented sentences governed by Sentence devices are cohesively sequenced with the topic as a pivot. Here, the interplay of *wa* and *ga* in relation to ellipsis holds the major key to ellipsis resolution. *Wa*, being the topic marker, extends its scope over to another clause/sentence until a new topic is introduced, whereas *ga*, being the subject marker, has its scope generally only within the clause. This difference induces the following basic rule for ellipsis resolution (Nariyama 2000) - the referent of ellipsis is the nearest previous *wa*-marked argument, except for complex sentences, namely:

Basic rule: **'X-*wa* ... . $\underline{ø_X}$ ... .'**

If the sentence is a complex sentence (see Nariyama 2000 for sentences with more than two clauses),

1) Apply the *wa*-marked referent within the sentence; i.e. in terms of surface structures (Square brackets denote subordinate clauses. Note that *wa* cannot generally appear in subordinate clauses):

'[$\underline{ø_X}$ ... ,] X-*wa* ... .' (cataphora; 13%)

or 'X-*wa* [$\underline{ø_X}$ ... ,] ... .' (this structure is much more common; 87%)

2) If the matrix *wa*-marked referent is not found within the complex sentence, apply the nearest previous *wa*-marked referent); that is, the *ga*-marked referent has no bearing on ellipsis resolution.

'X-*wa* ... . [Y-*ga* ... ,] $\underline{ø_X}$ ... .'

or 'X-*wa* ... . [$ø_X$ ... ,] Y-*ga* ... .'

or 'X-*wa* ... . [$\underline{ø_X}$ ... ,] $\underline{ø_X}$ ... .'

3) If a *ga*-marked referent is found within the complex sentence and if the two adjacent clauses are treated as a mono-clause, for example, indicated by a SS conjunctive particle (see §3.1.2), apply that *ga*-marked referent:

'**{**[ø$_Y$ ... SS conjuctive,] Y-*ga* ... **}**.'

The mechanisms are exemplified in the following contrived structure of hypothetical sentences:

X-*wa* ... . ø$_{1X}$ ... , **{**[Y-*ga* ... SS,] ø$_{2Y}$ ... .**}** ø$_{3X}$ ... . Z-*ga* ... .
ø$_{4X}$ ... . A-*wa* ..., ø$_{5A}$ ... .
(new topic)

Ellipsis 1 (ø$_1$) is coreferential with the nearest previous *wa*-marked argument X following the basic rule; ø$_2$ with Y derived from the rule 3); ø$_3$ with X derived from the basic rule; ø$_4$ with X from the rule 2), because the previous subjects are marked by *ga*; and ø$_5$ with A, which is a new topic introduced just before ø$_5$ following the rule 1).

The next section proposes an algorithm for ellipsis resolution which incorporates the three tiers of linguistic devices.

4. Algorithm

Due to the limited space in this paper, the detailed procedures and analysis of the proposed algorithm for ellipsis resolution have to be referred to Nariyama (2000), and this paper shows the main part of the algorithm. The proposed algorithm utilises "salient referent list", which is a memory bank that pools old referents from the previous sentences, and hence it builds context and inference. It is designed to reflect how humans store referential information.
Each sentence contains one or more referents. The subsequent sentences may retain one or more of these referents, some or all of which may be expressed by ellipses, and may also introduce one or more new referents. It is plausible to assume that when processing sentences, addressees store new referents by incorporating them into a pool of old referents from the previous sentences which have been stored in their cognition, and repeat this process as they process each new sentence. The salient referent list does just that. It functions like a memory bank in a cognitive sense, listing overt arguments appearing in the sentence by incorporating arguments that have appeared in the previous sentences. It is this input information which provides cues to resolve ellipses, not only for subject ellipsis but also for non-subject ellipses and multiple ellipses (more than one ellipsis per clause).

The salient referent list lists all overt arguments which have appeared up to the sentence in question. These overt arguments are listed in the following hierarchical order, called the "salient referent order list",[2] which accords the topicalised subject the highest saliency. In Japanese, the topicalised subject is morphologically differentiated from the non-topicalised subject by the use of different markers: *wa* and *ga* respectively.

Topicalised SB (Global > Local > Quotation) > SB > IO > OB > Obl

Figure 1: Salient referent order list

[2] The salient referent list order was eclectically adapted from the Japanese version of Expected Center Order in Centering Theory (Kameyama 1985), Keenan and Comrie's (1977) Noun accessibility hierarchy, Givon's (1979) topicality hierarchy, and Kuno's (1987) Thematic hierarchy.

'Topicalised SB (Global > Local)' is to cover the fact that although when a new topic is introduced, normally the new topic replaces the old one, when there is a global topic (usually the writer or the main topic/protagonist of the text), sometime it is still carried over after a long absence of the mention, while the current (local) topic is still in effect.

A salient referent list is created for each new sentence by modifying the one from the preceding sentence. If an argument appears with an identical grammatical relation to another argument already existing in the list, for example, where a subject exists and a new subject appears, the new subject takes its place for reasons of recency, except for topicalised subjects.[3]

5. Creation of salient referent lists and ellipsis resolution

This section explains how salient referent lists are created and used to resolve ellipses using a fragment of a text (*Seikachoo* Newspaper 2.1999). Each sentence is numbered, noted as [s1] being the first sentence in the text. Each subordinate clause is indicated by square brackets [ ] with the clause number on the right side. The matrix clause is numbered but not bracketed.

[s1]
[Watasi$_a$*-wa senshuu no doyoobi hotondo ne nai de]*$_1$
I-TopSB last week of Saturday hardly sleep not and[SS]

ϕ_a *terebi*$_b$*-o mi tuzuketa.*$_2$
SB TV-OB watch continued

"Last Saturday, I_a hardly slept, and instead (I_a) kept on watching TV_b."

[s1] has only one human argument - the topicalised subject *watasi*, and one inanimate object *terebi*. Each listed argument is given a number, for example, 'T1'. The argument under T1 has the highest saliency and is therefore the best candidate as the referent for the ellipsis; T2 is the next highest, and so forth. They are listed in the salient referent list accordingly, provided with the grammatical relation, topicality and person/animacy.[4] Hence, the salient referent list (SRL) for [s1] is formulated as follows:

SRL: [s1] {T1$_a$: *watasi* (TopSB; first person) >
T2$_b$: *terebi* (OB; inanimate)}

Ellipsis is resolved based on the information in the salient referent list for the sentence where the ellipsis appears. [s1] contains one ellipsis, so that T1 argument is applied as the referent, which is indeed the case. This coreference is indexed by subscript after T1 as 'T1$_a$', which is also coindexed in the text for easy recognition.

The next sentence is denoted as [s2].

[3] This method of listing only one argument under any one slot of grammatical relation works satisfactorily in the texts analysed. However, this needs be further investigated in more texts and larger texts.

[4] For simplicity, this paper notes only the grammatical relation, topicality and person/animacy. However, in practice, other information should be also noted; e.g. detailed semantic attributes of arguments except for first and second person, number, and the in-group/out-group distinction.

[s2]

Nazenara, [[watasi$_a$*-wa Goo*$_c$*-no fan de,]*$_1$
Because I-TopSB *Goo*-Gen fan be-and,

[Goo$_c$*-ga marason*$_d$*-ni choosensuru node,]*$_2$
Goo-SB marathon-Obl challenge because[DS]

ϕ_a ϕ_c *ooen sitakatta]*$_3$ *kara da.*$_4$
SB OB cheer wanted because be

"The reason is that I_a am a fan of Goo_c, and because Goo_c was competing in the $marathon_d$, (I_a) wanted to cheer for (him_c)."

The salient referent list needs to be updated with each new sentence, so that each salient referent list also needs to be numbered. In [s2], there are three overt arguments '*watasi*', '*Goo*', and '*marason*'. The referent '*watasi*' appears again with the same function of topicalised subject, so it remains as T1 in the list. Another argument '*Goo*' is a non-topicalised subject, so that it is listed as T2. The other argument '*marason*' is oblique, so that it is listed as T3. However, all arguments from the previous salient referent list must be carried over to the salient referent list for [s2], so that the inanimate object argument '*terebi*' must be incorporated into SRL [s2]. Because the object is listed higher than the oblique according to the salient referent order list,'*terebi*' is listed as T3 and '*marason*' as T4. Hence, the salient referent list for [s2] is formulated as follows:

SRL: [s2] {T1$_a$: *watasi* (TopSB; first person) >
T2$_c$: *Goo* (SB; third person) >
T3$_b$: *terebi* (OB; inanimate)
T4$_d$: *marason* (Obl: inaminate)}

[s2] has multiple ellipses in Clause 3: the subject and the object. Multiple ellipses are also ranked by the same salient referent order list, so that the subject ellipsis is ranked higher than the object ellipsis. The method of multiple argument ellipses resolution works as follows - the T1 argument in the salient referent list is chosen to be the referent for the highest ranked ellipsis in the salient referent order list. Similarly, T2 is selected as the referent for the next highest ellipsis, T3 is for the next highest ellipsis, and so forth. Accordingly, in [s2], the subject ellipsis is ranked higher than the object ellipsis, so that T1 'a' is chosen to be the referent of the subject ellipsis, and T2 'c' as the referent of the object ellipsis. This interpretation, following the proposed method, correctly selects the referents for ellipses including multiple ellipses. Thus, the salient referent list offers the key to resolving ellipses in Japanese.

6. Results and evaluation

The salient referent list was hand-tested on 7 short essays written by non-professional writers, which eliminate any potential bias caused by individual writing styles and topics. One of these essays (Text 1) is taken from *Seikachoo* Newspaper (2.1999) and the rest from PHP magazines (2.1999). The results are shown in Table 1. There are 210 ellipses.

Texts	Tx1	Tx2	Tx3	Tx4	Tx5	Tx6	Tx7	Σ
No. of sentence	24	25	9	18	67	9	19	171
No. of ellipsis	39	33	17	25	53	16	27	210
✗ No of incorrect ø	0	2	1	1	15	5	6	30
% of ✗	0	6.0	5.9	4.0	*28.3*	*31.3*	*22.2*	14.3

Table 1: Effectiveness of Salient Reference List

Table 1 shows that the texts are divided into two sharply contrasted groups in terms of accuracy; the salient reference list is extremely effective for Texts 1~4, but not so for Texts 5~7. There were mainly five factors responsible for the incorrect selections.

The first factor is caused by the lack of precise differentiation of 'global' topic and 'local' topic as to when 'global' topic overrides 'local'. In Texts 5~7, the writers usedellipted 'I' as the global topic seemingly at random points.

The second factor is the anomalous use of *ga* (the non-topicalised subject marker) which had scope over to the next sentence, which is normally the function of *wa*. What happened was that *ga* which also has another function of exhaustive listing (focus) overtook the topic marker *wa*. Namely, *wa* would have been used, if the subject were not focused.

Note that the first and the second factors comprise of 18/30 errors, most of which occurred within the same sentence or in succession, so that the actual occurrence was less than half the times. They are the main triggers for the poor performance for Texts 5~7.

The third occurred when the particle *wa* is used not as the topic marker but as the contrastive marker. The differentiation of the two functions of *wa* is murky and an unresolved issue in linguistics. When *wa* is used as the topic, the ellipsis is coreferential with the *wa*-marked referent. However, when it is used as the contrast, the ellipsis is coreferential with the previous *wa*-marked referent.

The fourth is the problem of mismatch caused by a part-whole relationship. For example, T1 may list 'John's life', but the ellipsis refers to 'John'.

The fifth is the notorious problem of world knowledge, comprising of 5/30 errors. In fact, 5 errors in relation to 210 ellipses, namely 5/210 = 2.4%, is not too significant.

Note that the complete algorithm consists of three components: [1] input information, [2] select referent for ellipsis, and finally [3] double check that selection utilising Predicate devices and Sentence devices. However, due to the limited space, the last component was not considered. The accuracy rate in Table 1 is, therefore, expected to be higher when the third component is incorporated and picks up incorrect selection.

7. Discussions and conclusions

This paper mainly discussed linguistic mechanisms for ellipsis resolution in order to demonstrate the need for more grammatical input into existing systems for improvement. I have shown briefly that three tiers of linguistic devices must be considered to fully capture the mechanisms of ellipsis resolution in Japanese: Predicate devices, Sentence devices, and Discourse devices. In a nutshell, the basic mechanisms of ellipsis are that Japanese sentences

are structured in such a way to anchor the topicalised subject which is selected on the basis of person/animacy and topicality, and it is this topicalised subject that is most prone to ellipsis.

These mechanisms are transformed into an algorithm. It uses 'salient referent list' that stores contextual information which is a must for ellipsis resolution. The proposed algorithm is therefore a promising method which can resolve subject ellipsis as well as non-subject ellipsis and multiple ellipsis. However, this is a preliminary report based on hand-simulated analysis using short narrative texts. The proposed method requires a large corpus analysis and corpora from difference genres (e.g. newspapers, conversation scripts) to be fully evaluated with consideration to those problems described in Section 6. This is the next step in future research.

References

Bresnan, Joan ed.: 1982, *The mental representation of Grammatical relations*, Cambridge Mass: MIT Press

Givón, Talmy: 1979. *On understanding grammar*, New York: Academic Press

Grosz, B. A.K.Joshi, & S.Weistein: 1983, Providing a unified account of definite noun phrases in discourse, In Proceedings of the 21st Annual Meeting of the American Association for Computational Linguistics, Cambridge. MA:ACL. pp. 44-50

Harada, Shigeyuki: 1976, Honorifics. In M. Shibatani ed. *Japanese Generative Grammar*. Syntax and Semantics Series Vol.5. New York: Academic Press. pp. 499-561

Ikehara, Satoshi et al.: 1997, *Goi-taikei* - A Japanese lexicon, Tokyo: Iwanami Shoten

Iwasaki, Shoichi: 1993, *Subjectivity in grammar and discourse: theoretical considerations and a case study of Japanese spoken discourse*, John Benjamins

Jacobsen, Wesley M: 1992, *The transitive structure of events in Japanese*, Tokyo: Kurosio

Kameyama, Megumi: 1985, *Zero anaphora: the case of Japanese*, Stanford Dissertation

Kameyama, Megumi: 1998. Intrasentential Centering: a case study, In M. Walker, K. Joshi and E. Prince (eds.). *Certering theory in discourse*. Oxford: Clarendon Press. pp. 89-112

Keenan, Edward & Bernard Comrie: 1977. Noun phrase accessiblility and universal grammar. *Linguistic Inquiry* 8: 63-99

Kuno, Susumu: 1987, *Functional syntax: anaphora, discourse and empathy*. Chicago: The University of Chicago Press

Minami, Fujio: 1974, *Gendai Nihon-go no kouzou* (Structures of modern Japanese). Tokyo: Taishukan

Nakaiwa, Hiromi: 1998, *Resolving Japanese zero pronouns in machine translation*, ms

Nakaiwa, Hiromi. & Ikehara, Satoshi: 1995. 'Intrasentential resolution of Japanese zero pronouns in a machine translation system using semantic and pragmatic constraints'. In Proc. of TMI 95: pp. 96-105

Nakaiwa, Hiromi et al. 1995, *Extrasentential resolution of Japanese zero pronouns using semantic and pragmatic constraints*, AAAI '95 Spring Symposium. pp. 99-105

Nariyama, Shigeko: 2000, *Referent identification for ellipted arguments in Japanese.* Dissertation: University of Melbourne

Nariyama, Shigeko: 2001, NIHONGO KOUBUN TO KOU SHOURYAKU NO GENSOKU (Principles of Japanese sentence structure and the order of deletion), In Proceedings of the 7th Annual Conference on Natural Language Processing, Tokyo University, vol.7: pp. 355-8

Nariyama, Shigeko: (in press). The WA/GA distinction and switch-reference for ellipted subject identification in Japanese complex sentences. *Studies in language.* John Benjamins

Stirling, Lesley: 1993, *Switch-reference and discourse representation*, Cambridge University Press

Strube, Michael: 1998, Never look back: An alternative to Centering. In proceedings of the 36th Annual Meeting of the *ACL*: 1251-1257

Corpus

PHP magazines. 9-12.1997, Kyoto: PHP Publishing
Seikacho newspaper. 2.1999, Kyoto