Resolving 'incognito' ellipsis: treatment for constructions that disguise ellipsis

Shigeko Nariyama

University of Melbourne / Nara Institute of Science and Technology
Melbourne Australia / Nara Japan
shigeko@unimelb.edu.au

Abstract

This paper raises a neglected issue in the study of ellipsis resolution. The existence of ellipsis under certain constructions is often disguised due to the structure that assigns the nominative marking to what is typically the object. This kind of ellipsis deserves attention in view of the fact that its referent is the agent of the sentence and that these constructions are observed in diverse languages. A problem is posed by virtue of the fact that English is not one of those languages, and it overtly expresses the referent of ellipsis that is implicit in those languages that use those constructions. Hence, the recognition and resolution of such ellipses is of importance particularly in machine translation systems that translate sentences with "incognito ellipsis" from those languages into English. After presenting the types of constructions, the paper explicates the mechanisms that govern the constructions in Japanese, and proposes a method to resolve such incognito ellipses along with common ellipses in a unified manner.

1 Introduction

Ellipsis resolution, anaphora resolution more generally, has been one of the most pressing issues and challenges in the design of machine translation systems. This paper raises an issue in ellipsis resolution that is observed in certain constructions of sentence that contain a different type of nominal ellipsis, which is referred to in this paper as "incognito" ellipsis. This issue has been neglected in the analysis of ellipsis resolution, presumably because the existence of this type of ellipsis is inconspicuous. It occurs in constructions which appear syntactically complete but are in actuality missing an agent.

The following examples from Japanese newspaper articles demonstrate the point.[1] Both (1) and (2) appears syntactically complete, containing all subcategorised arguments, including, and most importantly, the subject of the sentence. The earlier studies of ellipsis resolution have focused on resolving subject ellipsis, since the subject is by far the most frequently ellipted argument (Nariyama, 2002). Hence, ellipsis contained in this type of sentences can easily be overlooked from the study of ellipsis resolution. However, the literal translation of these sentences produces peculiar English sentences, and the natural English translation (shown by the arrows �) requires the specification of an argument that is not expressed in the Japanese sentences.

(1) *Bosnia–o hoomonsuru ishi-ga aru.*
Bosnia-OB visiting intention-SB exist/be

(lit.) 'An intention to visit Bosnia exists. / There is an intention to visit Bosnia.'

� 'ø has the intention to visit Bosnia.'

(2) *Aite-no shin'i-ga wakara-nai.*
partner-of true intention-SB understand-not

(lit.) 'The partner's real intention is not understandable/comprehensible.'

� 'ø don't understand the person's real intention.'

---

[1] The example sentences in this paper only show part of the sentences that succinctly demonstrates the point of argument, instead of quoting the whole long sentences.

Some abbrivations used in this paper are: IO= indirect object, NP=noun phrase, OB= object, SB=subject, Top=topic.

Although this type of sentence appears complete syntactically without missing arguments, semantically essential information that is needed in English, i.e. the agent of a sentence, is not expressed, namely, 'who has the intention' in (1) and 'who does not understand' in (2). In other words, despite its nominative marking, the overt NP is essentially the object of the sentence.

In addition, this problem may arise in a variety of contexts, because these constructions are observed in diverse languages, such as Spanish and Russian as well as Japanese (Shibatani, 1982:106); for example, Spanish expresses (1) as (3):

(3) *Hay intenci'on de visitar Bosnia.*
     Exist intention of visit   Bosnia
(lit.) 'Intention of visiting Bosnia exists'.

The verb *hay* has no agreement with the subject, *intenci'on*. Analogous to Japanese, it functions more like the object. Indeed, *intenci'on* takes an accusative pronoun when pronominalised.

The problem is posed by virtue of the fact that English is not one of those languages, and it overtly expresses the referent of ellipsis that is incognito in those languages which use the 'incognito ellipsis' constructions. Furthermore, this type of sentence is prevalent in Japanese (12% of clauses are of this type; see Table 1 in Section 2.2). Hence, the recognition and resolution of such ellipses is of importance particularly in machine translation systems that translate sentences with incognito ellipsis from those languages into English.

This paper focuses on the phenomenon seen in Japanese. There have been numerous papers on resolving Japanese ellipsis (Nakaiwa et al. 1995; Walker et al. 1994; Kameyama, 1985; inter alia). The book by Mazuka and Nagai (eds.) (1995) is a collection of 14 papers addressing different aspects of Japanese sentences from the point of view of sentence processing. However, the phenomenon of 'incognito ellipsis' raised in this paper is not dealt with in those papers.

The rest of this paper is organized as follows. Section 2 presents types of constructions that disguise the existence of ellipsis, such as (1) and (2), and the implications of incognito ellipsis. Section 3 explains the linguistic mechanisms that govern those constructions. Finally, Section 4 proposes a method for resolving incognito ellipsis along with common ellipsis in a unified manner.

2   Incognito ellipsis

2.1   Types of constructions

The constructions that disguise an agentive argument can be summarised as the following:

I.	Intransitive constructions	(e.g. (4b))
II.	Passive constructions	(e.g. (5b))
III.	Constructions with a pleonastic pronoun	
		(e.g. (6))
IV.	Existential constructions	(e.g. (1,7))
V.	Nominative object constructions	(e.g. (2))

Apart from a prototypical intransitive sentence, such as 'I walk', we can see in the following pairs of sentences that intransitive and passive sentences can be used to reduce the valency of the sentence, i.e. in (4b) and (5b) the agent is not expressed:

(4a)   I opened the door.
(4b)   The door opened.

(5a)   I held a meeting yesterday.
(5b)   A meeting was held yesterday (by me).

These types of constructions and the relationship with the corresponding transitive sentences are also observed in English; hence do not raise a problem that this paper is addressing. Similarly, in the third constructions such as (6), the agent is non-specific, generic or obfuscated and the implicit agent of 're-solve' is not treated as elliptical.

(6)   It is difficult to resolve pronominal anaphors.

The problem is posed by the next two types of constructions. The fourth type of construction is the existential constructions, "There is/are … ." in English. In Japanese and many other languages, this type of construction is also used to express one's possession, for example:

(7) *Ø Imooto-ga iru.*
     sister-SB   exist/be
(lit.) 'A/my sister exists. / There is a sister.'
➔   'ø have a sister.'

(7) appears complete syntactically, as it has the subject in the intransitive sentence. However, analogous to the example (1), the literal English translation of the Japanese sentence, 'A/my sister exists' or 'There is a sister', may be grammatical but strange or carry another meaning; hence the

possessor argument must be specified. The semantic content is commonly expressed in English by the 'have' construction, i.e. 'I have a sister.'

Note that in Japanese the possessor is not usually expressed and definiteness (the *a/the* distinction) is not marked, as they are not grammatically required (Bond, 2001). Note also that (7) can express the physical presence of 'sister', particularly if it is accompanied by locational or temporal adverbs, as in 'My sister is over there now,' in which case it is the prototypical existential sentence and there is no ellipsis, conceding that the exact distinction between the two meanings must be investigated further.

The fifth type of construction is the nominative object constructions (known as the 'dative subject' constructions in linguistics), of the kind that are not observed in modern English. In these constructions, the prototypical subject is marked by the dative (indirect object, *–ni* (or *–de*) in Japanese) or the topic marker (*-wa*), while the prototypical object is marked by the nominative (*-ga*), which predominantly marks the subject; that is, having a construction of [X-*ni/de/wa* Y-*ga* (SB?) Verb], where X is prototypically the subject of a sentence and semantically the agent, while Y is prototypically the object of the sentence and non-agent, but nonetheless marked by the nominative marker, hence the name 'nominative object'. This surface structure is identical for the existential constructions, since *–ni* and *–de* also denote location in Japanese. For this constructional similarity, these two constructions are treated together as a low transitive clause (LT) in the algorithm presented in Section 4.

The problem arises from the fact that X is often ellipted and only [Y-*ga* (SB?) Verb] is overtly expressed, as in (1) and (2). Hence, syntactically it seems difficult to detect the existence of the ellipted agent. However, a close examination of the constructions reveals the linguistic cues that can distinguish these types of constructions. This is presented in Section 3.

2.2   Implications of incognito ellipsis

As a reference point for getting some idea of the frequency of incognito ellipsis, a small examination was conducted using 5 newspaper articles and 5 magazine articles (PHP). The results are shown in Table 1.

Table 1: Proportion of incognito ellipsis and related figures

	news-paper	PHP	total
# of clauses/sentences	110/36	410/172	520/208
# of ellipsis	66	267	333
% of ellipsis/clause	**60.0%**	**65.1%**	**64.0%** ①
# of incognito ø / incognito agent	9 / 11	46 / 52	55/63
% of incognito ø	**81.8%**	**88.5%**	**87.3%** ③
% of incognito ø/ all ø	**13.6%**	**17.2%**	**16.5%** ②

The results do not seem to show significant differences between the newspaper articles and magazine articles with regard to the frequency of incognito ellipsis compared with other ellipses, so the following comments are made on the total figures. 333 ellipses are found in 520 clauses; ① on average every clause contains .64 ellipsis; ② on average 16.5% of those ellipses are incognito ellipses found under the existential and nominative object constructions; and ③ the agents are ellipted on average 87.3% of the time. In other words, 12.1% (63/520) of clauses form the constructions that create incognito ellipsis; the agent in these constructions are ellipted 87.3% of the time; and 16.5% of all ellipses derive from those constructions.

As a further reference, the sentence (2), renumbered as (8a), was translated into English using two free machine translation services on the internet, in order to ascertain how sentences under the constructions are currently translated.

(8a) Ø *Aite-no shin'i-ga   wakara-nai.*
    partner-of intention-SB understand-not
 (lit.) 'The true intention of the partner is
     not understandable/comprehensible.'

➡ 'ø don't understand the person's true intention.'

MT1: ? 'Real intention of the partner is
     not recognized.'
MT2: ? 'A partner's real intention is
     not understood.'

The translated sentences reveal that 'real intention' is treated as the subject of the sentence, not recognising the structure of the sentence as the nominative object and the existence of agent of (8a). So the referent and context are added to (8a) as (8b) to see if the sentence can be better translated.

(8b) *Watashi-wa shachoo-no hanashi-o kiita ga,*
ø shin'i-ga wakara-nai.

➡ 'I listened to the president's talk, but (I) don't understand (his) true intention.'

MT1: ? 'I heard the story of the president, but real intention is not understood.'
MT2: ? 'Real intention is not understood although I heard the president's talk.'

The translation remains unchanged, still showing that the system is not recognising the constructions that contain an incognito agent. So this time the agent is input to (8a) without ellipsis. The results with regard to the construction were identical for MT1, but improved for MT2.

(8c) *Watashi-wa shachoo-no shin'i-ga wakara-nai.*
➡ 'I don't understand the president's true intention.'

MT1: ? 'As for me, the real intention of the president is not recognized.'
MT2: 'I do not understand the president's real intention.'

It is highly probable that MT2 treats sentences, such as (8c), as an idiom when the sentence does not contain ellipsis, but does not recognise the structure of the sentence itself. It may be that these particular systems happen not to be of high standard, but it is equally true that currently the constructions are not widely recognised as a structure of importance.

3 Mechanisms behind the constructions

This section presents the linguistic cues that can distinguish the types of constructions that contain incognito ellipsis. The existential constructions can express possessions where the nominative object can be 'possessed' by human; usually when there is a discourse topic and no locational or temporal adverb, as in (1) and (7). The nominative object constructions tend to express emotions as in (2), ability as in (9), states as in (10), and sometimes involve adjectives instead of verbs as in (11). Hence, the semantics of the verb and adjective can signal the constructions that involve incognito ellipsis to a great extent. In all of these examples the agent of the sentence can be ellipted and be grammatical in Japanese, with the knowledge that the ellipsis is coreferential with the topic of the discourse.

(9) *(Nihon-wa) anzenhoshoo mende kokuren-ni*
Japan-Top security area the U.N.-to
Nanraka-no kooken-ga dekiru.
Something-of contributions-SB can

(lit.) '(As for Japan,) some contribution to the U.N. in the area of security is possible.'
➡ '(Japan) is able to contribute in some way to the U.N. in the area of security.'

(10) *(Bsha-ni) insider torihiki-no utagai-ga deteiru.*
Bcompany-in insider trading-of doubt-SB out

(lit.) '(In Company B,) suspicions of insider trading have appeared.'
➡ '(Company B) has been suspected for insider trading'

(11) *(John-wa) seikaku-ga warui.*
(John-Top) personality-SB bad

(lit.) '(As for John, his) personality is bad.'
➡ '(John) has a bad personality.'

Things are further complicated, in that the existential constructions can involve two ellipses, for example:

(12) *Denwa-ga atta.*
telephone-SB was
'There was a call (from $\emptyset_i$ to $\emptyset_j$).'
➡ '$\emptyset_i$ called $\emptyset_j$.'

Since both ellipses typically refer to humans (or organizations representing people in them), ellipsis resolution is unwieldy under the systems relying solely on selectional restrictions and may be unreliable with stochastic models alone. Hence, more detailed linguistic mechanisms behind the constructions that bear incognito ellipsis are needed for ellipsis resolution.

Nariyama (in press) proposes a set of principles applicable to Japanese that governs the pattern of ellipsis (the principle of ellipsis), which in turn is governed by the structure of sentences (the principle of direct alignment). Japanese sentences are structured in such a way as to express an argument high in person/animacy [1>2>3> animate>inanimate] and discourse salience (topicality/prior mention) as the subject. Namely, a subject must be higher than non-subject arguments (SB>nonSB) in terms of person/animacy and discourse salience. For example, a sentence, such as 'I surprised my husband', which has a first person subject acting on a third person object, i.e. [1→3]

forming a direct alignment, is acceptable. On the other hand, the reverse, such as 'The news surprised my husband', (i.e. [inanimate→3]) violates the principle of direct alignment, and indeed the sentence sounds unusual in Japanese, although this sentence is acceptable in English. Instead, the semantic content is commonly expressed using the intransitive construction, 'My husband got surprised by the news', which then satisfies the SB>nonSB requirement.

The same is true of a set of examples involving (12). (12a) with [1→3] is acceptable, while (12b) with [3→1] is infelicitous. Instead the semantic content is expressed using the existential construction, as in (12c).

(12a) [1→ 3]
Watashi-ga otto-ni   denwa-o   shita.
I-SB   my husband-to telephone-OB did
'I telephoned my husband.'

(12b) [3 → 1]
⁽*⁾*Otto-ga   watashi-ni denwa-o   shita.*
my husband-SB I-to   telephone-OB did
'My husband telephoned me.'

(12c) [use of the exsitential construction]
Otto-kara   watashi-ni denwa-ga   atta.
my husband-from I-to   telephone-SB was
'There was a call for me from my husband.

The type of restructuring from (12b) to (12c) can be formulated as follows:

$$[A_{SB} \ B_{IO} \ C_{OB} \ Verb_{transitive}]$$
$$\rightarrow \quad [B_{to} \ A_{from} \ C_{SB} \ Verb_{existential}]$$
where A<B in terms of animacy and discourse salience

Figure 1: Restructuring formula

This restructuring typically occurs when a sentence involves a 'Sino-Japanese' verb (Chinese origin verb; nominal verb). The roots of Sino-Japanese verbs are in reality simply nouns which have been converted into verbs by virtue of the inflection-bearing capability of –*suru* 'do' (-*shita* in past tense) (Jacobsen, 1992:205). Indeed, the literal translation of (12a) is 'I made a phone call to my husband.' Since those Sino-Japanese verbs that involves two human arguments are prevalent, but not limitless and predictable from their semantics (e.g. *shootai-suru* 'give an invitation' must involve two people: one who invites and another who is

invited), this information can be incorporated into a morphological parser for finer subcategorisation.

Thus, the principle of direct alignment reflected in Figure 1 resolves ellipsis from the type of sentence structure. Namely, a transitive sentence signals that, in the case of (12), the caller is higher than the receiver of the call in terms of animacy and discourse salience, while the existential construction is used when the caller is lower than the receiver. This is built into the makeup of the algorithm presented in Section 4.1.

Moreover, the principle of ellipsis provides useful information for resolving ellipsis. It specifies that the higher an argument in terms of animacy and discourse salience, the more prone to ellipsis. Indeed, when the higher argument is ellipted, as in 'There was a call (for me) from my husband', the sentence sounds natural. On the other hand, when a lower argument is ellipted while a higher argument is retained, as in 'There was a call for me (from my husband)', the sentence is unnatural in Japanese. Thus, the overt argument can be used as a reference point to resolve ellipsis.

The claim that sentences that violate the principle of direct alignment are restructured into the existential constructions or nominative object constructions (i.e. intransitive sentences) is verified in the small corpus analysis. The types of sentence structure in the PHP magazine articles are analysed and compared with their English translation. The results shown in Table 2 suggest that what would be transitive sentences (or passive) in English are often expressed using intransitive in Japanese.

Table 2: Proportion of sentence structures in Japanese and English

(n = 383 sentences for Japanese texts)

	Japanese	English
intransitive	**42.8%**	**23.4%**
transitive	**29.5%**	**36.1%**
passive	5.5%	14.4%
copula	22.2%	26.1%
	100%	100%

4 Resolution of ellipsis

Anaphora resolution often adopts elaborate and complicate rules and stochastic approaches. This paper, as a preliminary investigation, adopts one

simple method and algorithm demonstrated in Nariyama (2002), after which can be combined with a stochastic model such as the tournament model by Iida et al. (2003). For simplicity given the scope of this paper, this paper describes only the part of the algorithm pertinent to the topic addressed in this paper.

4.1 Algorithm

The core of the algorithm is captured in the use of "salient referent list". This is like a memory bank that pools old referents from the previous sentences, reflecting how humans may store referential information, and hence it builds context and inference. It is this input information that provides cues to resolve various types of ellipsis, including subject ellipsis, non-subject ellipsis and multiple ellipses (more than one ellipsis per clause). This paper extends its application to resolving incognito ellipsis.

The salient referent list details all overt arguments in the sentence, merged with the arguments that have appeared up until 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. Note that NomOB (nominative object) is added in the list, in order to account for the existential constructions and the nominative object constructions.

> Topicalised SB (Global > Local > Quotation)
> > SB > IO > NomOB > OB > other

Figure 2:   Salient referent order list

4.2   Resolving ellipsis

This subsection explains the way in which salient referent lists (SRL) are created and used to resolve ellipses using fragments of newspaper articles. Each sentence is numbered, noted as [sX]. Each subordinate clause is indicated by square brackets

[2] The salient referent list and its order were eclectically adapted from variety of sources, including the Japanese version of Expected Center Order in Centering Theory (Kameyama, 1985), Givon's (1979) Topicality hierarchy, Kuno's (1987) Thematic hierarchy. They also share basic ideas with the Lappin and Leass algorithm (1994).

In Japanese, the topicalised subject is morphologically differentiated from the non-topicalised subject by the use of different postpositional markers: *wa* and *ga* respectively.

[ ] with the clause number on the right side. The matrix clause is numbered but not bracketed.

Text 1 [s1]
Baxter[a] -wa   [" … ."][1] *to chooshuu*[b]*-ni katatta.*[2]
         -TopSB   that audience-to talked

'Mr Baxter[a], the coach, said to the audience that . .'

[s1] has two overtly expressed arguments – the topicalised subject (TopSB), Baxter, and the indirect object *chooshuu* 'audience'. By following the salient referent order list, TopSB is listed higher than IO. 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 referent for the ellipsis; T2 is the next highest, and so forth. These overt arguments are listed in the SRL accordingly, provided with detail-ed lexical information, including the grammatical relation, topicality and animacy. Accordingly, the SRL for [s1] is formulated as follows:

SRL: [s1] {T1[a]: *Baxter* (TopSB; third person) >
T2[b]: *chooshuu* (IO; mass human)}

Ellipsis is resolved based on the information contained in the SRL for the sentence where the ellipsis appears. [s1] contains no ellipsis, so we process the next sentence [s2].

Text 1 [s2]
[ϕ[a] [*Taidan-go,   shidoo*[c]*-o   tsuzuke tai*][1]
SB leaving team-after coaching-OB continue want
kimochi[d]*-ga   aru*][2/LT] *to* ϕ[a] *itta*[3]
feeling-NomOB exist   that SB said

'(He[a]) said that after leaving the team, (he[a]) has the desire to continue coaching.'

Clause 2 uses the existential verb *aru*. This structure together with the verbal semantics of the nominative subject (*kimochi* 'feelings' that must be possessed by someone) reveals that the clause has the missing agent. So it is detected as the construction containing an incognito agent, noted by LT (low-transitive clause, encompassing the existential and nominative object constructions) next to the clause number. Note that for simplicity, the subject of coaching in Clause 1 is precluded, as it is not required in the English translation.

The SRL is updated with each new sentence. In [s2], there are two overt arguments: *shidoo* (the object) and *kimochi* (NomOB). These are merged with the previous SRL [s1]. By following the salient referent order list, the SRL for [s2] is formulated as follows.

SRL: [s2] {T1$_a$: *Baxter* (TopSB; third person) >

T2$_b$: *chooshuu* (IO; mass human) >
T3$_d$: *kimochi* (NomOB: inaminate) >
T4$_c$: *shidoo* (OB; inanimate)}

The verbal semantics and the structure of the clauses (i.e. subcategorisation) signal that [s2] has two ellipses: the subject in Clause 2 and the subject in Clause 3. Ellipsis is resolved per clause. So the T1 argument is applied as referent to each clause, i.e. both ellipses in this case. This interpretation, following the method, correctly selects the referents for ellipses including incognito ellipsis.

This operation reflects the mechanisms presented in Section 3 that the argument high in animacy and discourse salience is expressed as the subject (which is listed high in the SRL list) and most prone to ellipsis (i.e. taking the highest argument in the SRL as referent).

Observe another example, a part of which is (2), this time involving the nominative subject construction. For simplicity, the SRL for the previous sentence is provided as follows:

SRL: [s1] {T1$_a$: *shachoo* (TopSB; third person) >
T2$_b$: *handan* (OB; inanimate)}

Text 2 [s2]
[φ$_a$ Aite-no   shin'i$_c$-ga   wakara-nai]$_{1/LT}$
SB partner-of true intention-NomOB understand-not

to   φ$_a$ shinchoona shisei$_d$-o   miseta.$_2$
that SB cautious   attitude-OB showed

➡ '(He$_a$) appeared cautious by saying that (he$_a$) didn't understand the person's real intention.'

This sentence has two overt arguments: *shin'i* (NomOB, which is detected from the type of verb *wakaru* 'understand', whose semantics require an agent of 'understand') and *shisei* (OB). These are incorporated into the SRL [s1] to make SRL [s2]. The handling of the possessive *aite* is left for further study. T2 *handan* in SRL [s1] is replaced by the new object *shisei* in [s2] for the reasons of re-

cency (as the current system lists only one argument under any one grammatical relation), and the SRL for [s2] is formulated as follows:

SRL: [s2] {T1$_a$: *shachoo* (TopSB; third person) >
T2$_c$: *shin'i* (NomOB; inanimate) >
T3$_d$: *shisei* (OB; inanimate)}

From subcategorisation, [s2] has two missing arguments. So the T1 argument is applied as referent to ellipsis in each clause, to both ellipses in this case. This provides the correct interpretation for the ellipses.

[s3] is the problematic sentence, similar to (12).

Text 2 [s3]
Futsuka go, sono aite$_e$ kara φ$_a$ renraku$_f$-ga   atta.$_1$
2 days later, that person from   contact-NomSB existed
(lit.) 'Two days later, there was a contact (to φ$_a$)
    from that person$_d$.'
➡ 'Two days later, the person$_d$ contacted (him$_a$).'

Analogous to creating SRLs for the previous sentences, SRL [s3] is created as follows:

SRL: [s3] {T1$_a$: *shachoo* (TopSB; third person) >
T2$_f$: *renraku* (NomOB(SJV+2); inanimate) >
T3$_d$: *shisei* (OB; inanimate)
T4$_e$: *aite* (other: SJV'from'; third person)}

The sentence structure and the semantics of the Sino-Japanese verb 'contact' singal that [s3] takes two human arguments (from ø and to ø), noted as 'NomOB(SJV+2)' in the SRL. So [s3] is missing an argument: 'who is contacted.' The T1 argument is chosen as the referent, which provides the correct interpretation.

4.3   Test results and evaluation

The salient referent list was hand-tested on the same texts used for Table 1. The results are shown in Table 3, which duplicates some of the information from Table 1 for convenience.

There are 333 ellipses. ④ in Table 3 shows that the use of SRL resolves ellipses, including incognito ellipses, with an accuracy of 85.6%. 48 incorrect selections were made by SRL, which were due to the following reasons. The most frequent mis-

takes were caused by the method not distinguishing generic referents from particular referents.

Table 3: Effectiveness of Salient Reference List

	news-paper	PHP	total
# of clauses/sentences	110/36	410/172	520/208
# of ellipsis	66	267	333
% of ellipsis/clause	60.0%	65.1%	64.0%
# of incognito ø / incognito agent	9 / 11	46 / 52	55/63
% of incognito ø	81.8%	88.5%	87.3%
% of incognito ø / all ø	13.6%	17.2%	16.5%
④SRL: % of ✓ for all ø	**87.9%**	**82.0%**	**85.6%**
SRL: % of ✓ for incognito ø	100%	71.7%	77.2%

1. 22/48 (6.6% of all ellipses): ellipsis referring to generic (non-specific) referents that do not appear in the context. All of these occurred in PHP and none in the newspaper articles.
2. 11/48 (3.3%): two topics simultaneously dominating the story. They are distinguished by commonsense knowledge, for example, the knowledge about a mother and her child.
3. 13/48 (3.9%): number problem. The salient referent lists select "I", but from the context "we" is more appropriate.
4. 2/48 (.6%): referring to the whole sentence, instead of a particular referent.

5   Conclusions

This paper has provided a resolution method for 'incognito ellipsis', whose constructions are apparently syntactically saturated but which nevertheless require resolution for full interpretation. The proposed method using 'salient referent list' seems promising, as it can resolve various types of ellipsis with a high accuracy. However, this is a preliminary report based on hand-simulated analysis using a small number of texts. Empirical examinations of large corpus will be more assuring, after which the proposed method can be combined with a stochastic model in order to account for vagarious aspects of language.

In addition, the explicit specifications of at least two issues must be provided for structurally identical sentences: the differentiation of sentences that involve incognito ellipsis (such as (12)) from the ones that do not (such as proto-typical intransitive/existential sentences), and the handling of the determiner on the nominative object, as in 'A (of B)' and 'a/the A' (Bond 2001).

References

Francis Bond. 2001. Determiners and number in English contrasted with Japanese, as exemplified in machine translation, Ph.D. thesis, University of Queensland.

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

Ryu Iida, et al. 2003. Incorporating contextual cues in trainable models for coreference resolution, EACL Workshop on Computational treatment of anaphora, Budapest, pp.23-30.

Wesley Jacobsen. 1992. *The transitive structure of events in Japanese*, Kuroshio, Tokyo.

Megumi Kameyama. 1985. Zero anaphora: the case of Japanese, Ph.D. thesis, Stanford University.

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

Shalom Lappin and Herber J. Leass. 1994. An algorithm for pronominal anaphora resolution, *Computational Linguistics* 20.4: pp.535-561.

Mazuka R. and Nagai N. eds., 1995. *Japanese sentence processing*, Lawrence Erlbaum Associates, NJ.

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

Shigeko Nariyama. 2002. Grammar for ellipsis resolution in Japanese, the 9th TMI, Kyoto, pp.135-145.

Shigeko Nariyama. in press. *Ellipsis and reference tracking in Japanese,* SLCS 66, John Benjamins.

Masayoshi Shibatani. 1982. Japanese grammar and universal grammar. *Lingua* 57: pp.103-123.

Marilyn Walker, Iida M. and Cote S. 1994. Japanese discourse and the process of centering. *Computational Linguistics* 20: pp.193-232.