Japanese Sentence Processing: Evidence from Topic sentences

Noriko Nagai
Assistant Professor
Duke University
Asian and African Languages and Literature
Duke University
Durham, North Carolina 27706 U.S.A.

1. Introduction

Sentence processing is one of the fastest growing fields, and much is known about the mechanisms of English sentence processing. However, much of Japanese sentence processing mechanisms remain unknown. This paper attempts to explore the mechanisms of Japanese sentence processing, drawing examples from topic sentences in Japanese.

Japanese topic sentences violate Subjacency because topic phrases are not derived via movement but rather base generated and empty categories in topic sentences are <u>pro</u> (Saito 1985). Given this explanation, topic sentences in which topic phrases are coindexed with <u>pros</u> in relative clauses should be all grammatical; they are properly licensed by principles of the grammar within the framework of Government and Binding. However, some topics are acceptable, while some others are not, as illustrated by the following examples:

1.a Hanako$_i$-wa [$_S$[$_{NP}$[$_S$ e_i e_j aishite-ita] hito$_j$]-ga
     TOP   love-PAST   person-NOM
     jisatsu shite-shimat-ta]
     suicide do-PRFT-PAST
     'Speaking of Hanako, the person who (s/he) loved committed suicide.'
  b.*Hanako$_j$-wa [$_S$[$_{NP}$ [$_S$ e_i e_j aishite-ita] hito$_i$]-ga
      TOP   love-PAST   person-NOM
     jisatsu shite-shimat-ta]
     Suicide do-PRFT-PAST
     'Speaking of Hanako, the person who loved him/her committed suicide.'

There remain questions about circumstances under which those topic sentences become unacceptable. To account for such contrast in acceptability of topic

sentences, various attempts were made in the past.
For instance, Kuno (1973) proposes a pragmatic
constraint called "aboutness relation", which must
exist between a topic and the rest of the sentence.
More strict grammatical conditions are also proposed.
Hasegawa (1981) proposes a constraint appealing to
grammatical relations between the topic and the head of
a relative clause, on one hand, and their corresponding
gaps. Ue (1982) claims Fodor's nested dependency
constraint is applicable to Japanese topic sentences.
Yet, none of these studies can successfully explain the
acceptability and unacceptability of all the topic
sentences that I will examine in this paper. The
failure of the past analyses lies in the failure to
consider parsing contexts, such that gaps are not all
detected at once nor gaps are all filled at once in the
parsing.

This paper examines a variety of topic sentences
containing multiple gap dependencies, some of which
have never investigated before, and shows that topic
sentences become unacceptable due to some parsing
difficulty; more specifically, it demonstrates that
topic sentences result in parsing breakdown because
the parser immediately and obligatorily constructs a
syntactic representation of a word string and
deterministically assigns a filler to a potential gap,
which turn out illegitimate at the end of the sentence.
This paper does not only explain how the
acceptability/unacceptability of topic sentences with
multiple gaps are determined but also suggests how the
human parser comprehend topic sentences with multiple
gaps in Japanese, which has not yet been studied in
detail in the past.

Topic sentences containing multiple gaps create
two major problems from a parsing perspective: gap
detection and assignment of a filler to a potential
gap[1]. In what follows, we will show that topic

---

[1]When we investigate sentences from a parsing
perspective, there is one more factor we have to
consider, namely filler identification (cf. Cliftion
and Frazier (1989)). In this paper I will not refer to
this problem at all simply because the topic sentence
does not create any interesting problems in identifying
a filler, The filler in the topic sentence is clearly

sentences become unacceptable when the human parser
fails to detect a gap and/or to assign a filler to an
intended gap, .

2. Gap detection
     Japanese is a head-final language and hence
complements of a head precede the head. When
complements of a verb are phonetically null (empty
categories), the existence of them in a sentence cannot
be detected until a verb is encountered (until which
point, nothing indicates the existence of gaps). Upon
reaching a verb, the parser postulates gaps in
accordance with the subcategorization information of
that verb (see Inoue (1984) and Mazuka (1990) for
similar views). For instance, in sentence 2 when the
parser encounters the first verb, namely narat-ta, it
realizes that two arguments of that verb are
phonetically null and postulates two gaps as
illustrated in 3:

2.   sono-gakusee-wa kyonen narat-ta sensee-ga
     that student-TOP last year learn-PAST teacher-NOM
     naku-nat-ta
     die-PAST
     'Speaking of the student, the teacher who he
     learned from last year died.'
3.   sono-gakusee-wa [$_S$ e e kyonen narat-ta]

     However, gaps are not always identified by the
parser. When gaps are not successfully detected, the
sentence is judged unacceptable. In the following two
subsections, we will examine those unacceptable
sentences and explore mechanisms employed by the parser
to detect gaps.

2.2.1 Failure to detect a gap (I)
     When there is a lexical NP that can be an argument
of a verb, it is impossible to detect a gap for that
position. For instance, in parsing example 4, there is
a choice between taking the lexical NP, namely Taroo-ga
as an argument of the first verb, aishite-i-ta and

---

marked with the particle wa and placed at the sentence
initial position. Hence, I assume that the parser
recognizes it without any special difficulties.

positing a gap for that position, as indicated 5.a and
5.b respectively:

4.   Hanako-wa Taroo-ga aishite-i-ta hito-o
           TOP   NOM love-PAST   person-ACC
      koroshite-shimat-ta
      kill-PRFT-PST

5.a   Hanako$_i$-wa [$_S$ e$_i$ [$_{NP}$[$_S$ Taroo-ga   e$_j$ aishite-i-ta]
      hito$_j$]-o koroshite shimat-ta]
      'Speaking of Hanako, she killed the person Taro
      loved.'

  b   *Hanako$_i$-wa [$_S$ Taroo-ga [[e$_i$ e$_j$ aishite-i-ta]
      hito$_j$]-o koroshite shimat-ta]
      'Speaking of Hanako, Taro killed the person she
      loved.'

In 5.a the lexical NP, <u>Taroo-ga</u> is analyzed as the
subject argument of the verb, <u>aishite-i-ta</u> and a gap is
posited as the object argument of the verb. In 5.b two
gaps are postulated as the arguments of the verb,
<u>aishite-i-ta</u>, and the lexical NP, <u>Taroo-ga</u> is analyzed
as the subject argument of the main verb,
<u>koroshite-shimat-ta</u>.
    Although these two interpretations of example 4
should be equally possible, (a)-interpretation is
acceptable, while (b)-interpretation is not.
    Similarly, the following example can be
interpreted in two ways, yet only one interpretation is
acceptable, as contrasted in acceptable example 7.a and
unacceptable example 7.b:

6.   Hanako-wa Taroo-o aishite-ita hito-ga
           TOP   ACC loved   person-NOM
      koroshite-shimat-ta
      killed

7.a   Hanako$_i$-wa [$_{S[NP}$[$_S$ e$_i$ Taroo-o aishite-ita]
      hito$_j$]-ga e$_i$ koroshite-shimat-ta]
      'Speaking of Hanako, the person who loved Taroo
      killed her.'

  b   *Hanako$_i$-wa [$_S$ Taroo-o$_j$ [$_{S[NP}$[$_S$ e$_i$ e$_k$ aishite-ita]
      hito$_k$-]ga e$_i$ koroshite-shimat-ta]]
      'Speaking of Hanako, the person who she loved
      killed Taroo.'

In 7.a, the lexical NP, namely <u>Taroo-o</u> is interpreted
as the object argument of the verb, <u>aishite-i-ta</u> and a
gap is posited as the subject argument of the verb. In
7.b, two gaps are posited as the arguments of the verb,

<u>aishite-i-ta</u> and the lexical NP, <u>Taroo-o</u> is interpreted as the object argument of the main verb, <u>koroshite-shimat-ta</u>.

The contrast in acceptability between the (a)-interpretation and the (b)-interpretation of each example above clearly demonstrates the intervening lexical NP prevents the parser from detecting a gap[2]. Two pieces of evidence further support this claim.

First, when the intervening lexical NP in each example 5.b and 7.b, namely <u>Taroo-ga</u> and <u>Taroo-o</u>, respectively, is scrambled so that it does not precede the first encountering verb, the example with the intended interpretation results in acceptable, as shown in 8.b and 9.b:

8.a *Hanako$_i$-wa [$_S$ **Taroo-ga** [$_{NP}$[$_S$ $\underline{e}_i$ $\underline{e}_j$ aishite-i-ta] hito$_j$]-o koroshite shimat-ta]^i (=5.b)
'Speaking of Hanako, Taroo killed the person she loved.'

  b Hanako$_i$-wa [$_S$[$_{NP}$[$_S$ $\underline{e}_i$ $\underline{e}_j$ aishite-i-ta] hito]-o$_k$ **Taroo-ga** $\underline{e}_k$ koroshite shimat-ta]$_k$

9.a *Hanako$_i$-wa [$_S$ **Taroo-o**$_k$ [$_{NP}$[$_S$ $\underline{e}_i$ $\underline{e}_j$ aishite-ita] hito$_j$]-ga $\underline{e}_k$ koroshite-shimat-ta]^j (=7.b)
'Speaking of Hanako, the person who she loved killed Taroo.'

  b Hanako$_i$-wa [$_S$[$_{NP}$[$_S$ $\underline{e}_i$ $\underline{e}_j$ aishite-ita] hito$_j$]-ga **Taroo-o** koroshite-shimat-ta]

Unlike 8.a, 8.b is acceptable because the object of the main verb, namely <u>aishite-i-ta</u> <u>hito-o</u> is fronted over the subject of the main verb, <u>Taroo-ga</u>, so that no lexical NPs that can be arguments of the first verb, <u>aishite-i-ta</u> precede the verb. Likewise, 9.b is acceptable because the object of the main verb follows the first verb and no lexical NPs that can be arguments of that verb precede it.

Second, when the Case particle of an intervening lexical NP indicates that it can not be an argument of

---

[2] In parsing English sentences, it is known that the parser sometimes prefers a lexical NP over a gap for some doubtful gaps (the "last-resort model of gap finding (Fodor 1978)). For a detail discussion of this principle, see Fodor.

a verb, gaps for that verb are all successfully
identified. Observe the following examples:

10.a Hanako-wa Taroo-to aishite-i-ta hito-o
              TOP   with love-PAST person-ACC
      koroshite-shimat-ta
      kill-PRFT-PAST
      'Speaking of Hanako, she killed the person who
      she loved with Taroo.'
   b Hanako-wa Taroo-ni aishite-i-ta hito-ga
              TOP   with love-PAST   person-NOM
      atte-shimat-ta
      meet-PRFT-PAST
      'Speaking of Hanako, the person she loved met
      with Taroo.
   c Hanako-wa Taroo-o dosookai-de at-ta
              TOP   ACC school-reunion-at meet-PAST
      hito-ga shitte-i-ta
      person-NOM know-PAST
      'Speaking of Hanako, the person she met at the
      school reunion knew Taroo.'

In each of the above examples, the particle attached to
the intervening lexical NP, namely <u>Taroo</u> indicates that
this NP is not an argument of the first verb. For
instance, in 10.b the first verb, <u>aishite-i-ta</u> takes
only subject and object arguments. The NP immediately
prior to the verb is marked with <u>ni,</u> which indicates
that it is not the subject nor object. Hence, the
parser successfully detects two gaps for the verb and
the sentence results in acceptable, as in 10.b':

10.b' Hanako$_i$-wa [$_S$ Taroo-ni$_k$ [$_S$[$_{NP}$[$_S$ e_i e_j
       aishite-i-ta] hito$_j$]-ga e_k atte-shimat-ta]]

     The above two pieces of evidence confirm that the
parser cannot detect a gap when there is a lexical NP
that can be an argument of a verb.
     The preference for a lexical NP over a gap is not
arbitrary but follows a very basic parsing principle.
In the on-line processing of sentences with filler-gap
dependencies, the processor's task is not limited to
detecting gaps. Obviously, the processor identifies a
gap while constructing a syntactic representation of an
incoming word string. Although exact mechanisms of
Japanese sentence processing remains to be more fully
explored, I assume here that the Japanese parser make
an initial structural representation on-line upon
reaching an incoming constituent. Because Japanese is

a head-final language, complements of a head remain
unattached until the head is reached. I will assume
that when the head is reached, all the potential
complements of a head is immediately and obligatorily
attached to the head. Because of this parsing
principle the parser fails to identify a gap when an NP
that can be an argument of a verb precedes that verb.
For instance, in parsing word string 11.a the second
NP, <u>Taroo-ga</u> is attached as a complement of the verb,
<u>aishite-i-ta</u>, as in 11.b, when that verb is reached,
instead of being held unattached.

11.a Hanako-wa Taroo-ga aishite-i-ta hito-o
         TOP   NOM love-PAST   person-ACC
     koroshite-shimat-ta
     kill-PRFT-PST
   b Hanako-wa [Taroo-ga e̱ aishite-i-ta]

     There are three pieces of evidence that supports
this parsing principle.
     The first evidence is based on a garden path
sentence. The following example exhibits a strong
garden path effect:

12. **Taroo-wa imooto-ni Hanako-o shookai-shi-ta**
         TOP sister-with   ACC introduce-PAST
     hito-ga koi-shite-i-ta
     person-NOM love-PAST
     'Speaking of Taroo, the person who he introduced
     Hanako to loved (his) sister.'

In parsing example 12, the parser initially analyzes
the word string in the bold, <u>Taroo-wa imooto-ni
Hanako-o shookai-shi-ta</u>, as a simple sentence, as in
13.a.

13.a Taroo-wa [e̱ imooto-ni Hanako-o shookai-shi-ta]
   b *Taroo-wa [e̱ imooto-ni Hanako-o shookai-shi-ta]
        hito-ga koi-shite-i-ta

This analysis, however, breaks down when the main verb,
<u>koi-shite-i-ta</u> is reached for the following two
reasons. First, the topic NP, <u>Taroo-wa</u> is stranded
without any corresponding gap. Second, the main verb,
<u>koi-shite-it</u> does not have the dative argument and it
is impossible to comprehend who the person that
introduced Hanako to his sister loves. Hence, the
parser must reanalyze the sentence and posit a dative

argument gap for the verb, shookai-shi-ta, as shown in
14.

14. Taroo-wa [$_S$ imooto-ni [$_{NP}$[$_S$ e e Hanako-o
        TOP sister-to   ACC
    shookai-shi-ta] hito-ga e koi-shite-i-ta
    introduce-PAST person   love-PAST
    'Speaking of Taroo, the person he introduced
    Hanako to loved (his) sister.'

The second evidence is based on a developmental
study conducted by Clancy and her associates (1986).
They report that Japanese children analyze the first
word string, zoo-ga kirin-o taoshi-ta in example 15 as
a simple sentence, which is ultimately proven to be
incorrect.

15. Zoo-ga kirin-o taoshi-ta shika-o nade-ta
    elephant-NOM giraffe-ACC knocked deer-ACC patted
    'The elephant patted the deer that knocked down the
    giraffe.'

The third evidence is from an experimental study
conducted by Mazuka and her associates (1989). They
report the existence of mild garden path sentences in
Japanese like example 16[3]. In 16, the word string
roojin-ga kodomo-o yonda is analyzed as a main
sentence, which ultimately turns out incorrect.

16. Roozin-ga kodomo-o yonda josee-to hanasi-o
    old man-NOM child-ACC called woman-with talk-ACC
    shita
    did
    'The old man talked to the woman who called the
    child.'

The above three pieces of evidence clearly
demonstrate that the Japanese parser makes an initial
syntactic representation of an incoming constituent.
More specifically, the parser attaches all the
potential NP arguments to a head immediately and

---

[3]Mazuka et al note that Japanese garden path
sentences like 15 is not as sever as English garden
path sentences fro unknown reasons.

obligatorily when the head is reached. This in turn proves that the parser fails to detect a gap when there is an NP that can be an argument of a verb.

2.2. Failure to detect a gap (II)

Because in the on-line processing, gaps are identified when a verb is reached, gaps of a higher verb in the syntactic structure cannot be detected until that verb is encountered later on. Observe the following example:

17. Tanaka kyooju$_j$-wa [$_{NP}$[$_S$ e$_i$ [$_{NP}$ [$_S$ e$_j$ e$_k$ kai-ta]
    professor-TOP   write-PAST book-ACC
hon$_k$]-o koosee-shite-i-ta] gakusee$_i$]-ga

rectify-PROG-PAST student-NOM
inakunatte-shimat-ta
disappear-PRFT-PAST
'Speaking of Prof. Tanaka, the student who was rectifying the paper that he wrote cannot be located.'

Notice in 17, the leftmost gap is not an argument of the verb the parser first encounters, but it is an argument of the next highest verb, koosee-shi-ta. Therefore, at the time the first verb, kai-ta is encountered the gap associated with the next highest verb cannot be identified in the left-to-right processing. This delay of the gap detection causes a serious problem in assigning a filler to a gap, which I will discuss in 3.

3. Assignment of a filler to a gap

In the previous section we have seen topic sentences result in unacceptable when an intended gap is not successfully detected. Even though gaps are successfully detected, some topics result in unacceptable because the parser fails to assign a filler to an intended gap. In this section, we will observe such topic sentences.

Some topic sentences become unacceptable when the parser fails to assign a filler to an intended gap. Observe the following oft-cited examples:

18.a Hanako$_i$-wa [$_{NP}$[$_S$ e$_i$ e$_j$ aishite-i-ta] hito$_j$-ga
    TOP   love-PAST   person-NOM
shinde-shimat-ta
die-PRFT-PAST

'Speaking of Hanako, the person Hanako loved
died.'

b *Hanako$_i$-wa [$_{NP}$[$_S$ $\underline{e}_j$ $\underline{e}_i$ aishite-i-ta] hito$_j$-ga
    TOP   love-PAST   person-NOM

shinde-shimat-ta
die-PRFT-PAST
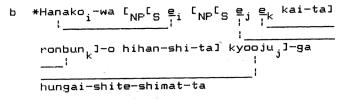'Speaking of Hanako, the person who loved Hanako
died.'

Because of the examples like those cited above, it is
assumed that Fodor's (1978) Nested Constraint is
applicable to Japanese (for instance, in Ue (1982)).
However, this constraint cannot explain
acceptability/unacceptability of the following
examples[4].:

19.a Tanaka kyooju$_j$-wa [$_{NP}$[$_S$ $\underline{e}_i$ [$_{NP}$ [$_S$ $\underline{e}_j$ $\underline{e}_k$ kai-ta]
      professor-TOP   write-PAST

hon$_k$]-o koosee-shite-i-ta] gakusee$_i$]-ga
book-ACC rectify-PROG-PAST student-NOM
inakunatte-shimat-ta
disappear-PRFT-PAST
'Speaking of Prof. Tanaka, the student who was
rectifying the paper that he wrote cannot be
located.'

b *Hanako$_i$-wa [$_{NP}$[$_S$ $\underline{e}_i$ [$_{NP}$[$_S$ $\underline{e}_j$ $\underline{e}_k$ kai-ta]
    TOP   write-PAST

ronbun$_k$]-o hihan-shi-ta] kyooju$_j$]-ga
paper-ACC criticize-PAST
hungai-shite-shimat-ta
professor-NOM resent-PRFT-PAST
'Speaking of Hanako, the professor resented her
criticism of his paper.'

20.a Tanaka kyooju$_j$-wa [$_{NP}$[$_S$ $\underline{e}_i$ [$_{NP}$ [$_S$ $\underline{e}_j$ $\underline{e}_k$ kai-ta]

hon$_k$]-o koosee-shite-i-ta] gakusee$_i$]-ga

inakunatte-shimat-ta

---

[4]Hasegawa (1981) and Kornfilt et al (1981) also
argue against the crossing constraint for reasons which
differ from mine. I will not discuss their arguments
here due to the limited space of this paper.

b *Hanako$_i$-wa [$_{NP}$[$_S$ e$_i$ [$_{NP}$[$_S$ e$_j$ e$_k$ kai-ta]

ronbun$_k$]-o hihan-shi-ta] kyooju$_j$]-ga

hungai-shite-shimat-ta

Notice that example 19.a is acceptable even though it
exhibits crossing lines of filler-gap dependencies, as
shown in 20.a. Hence, the crossing constraint
incorrectly rules out 19.a. On the contrary, 19.b is
unacceptable even though it does not exhibit crossing
lines of filler-gap dependencies, as shown in 20.b.
Hence, the crossing constraint does not rule out
unacceptable example 19.b.
     In order to explain the contrast in acceptability
in the above examples, let us first examine differences
between them in detail. 19.a contrasts sharply with
19.b in that the propositions associated with 19.a are
all locally comprehended in the parse of that sentence,
while the propositions associated with 19.b are not
comprehended locally. There are three propositions
associated with 19.a: i) Tanaka-kyooju-ga hon-o kaita
"Prof. Tanaka wrote a book."; ii) gakusee-ga hon-o
koosee-shi-ta "A student rectified the book."; and iii)
gakusee-ga inakunat-ta "The student is missing." In
the parse of 19.a these propositions are immediately
comprehended without any delay when each verb and each
head of the relative clauses are reached. That is,
because the topic NP is assigned to one of the two gaps
of the first verb and the first head of the relative
clause is assigned to the other gap, the first
proposition is comprehended at this point. The second
proposition is understood when the second verb and the
head of the second relative clause are encountered;
the gap of the second verb is filled by the head of the
second relative clause. Finally, the third proposition
is comprehended when the third verb is encountered.
Each of the three proposition in 19.a, hence is
understood locally. On the contrary, in unacceptable
19.b three propositions are not locally comprehended.
19.b involves the following three propositions: i)
kyooju-ga ronbun-o kai-ta "A professor wrote a paper.";
ii) Hanako-ga ronbun-o hihan-shi-ta "Hanako criticized
the paper." and iii) kyooju-ga hungai-shi-ta "The
professor got angry (about it)." When the first verb
is encountered two gaps associated with that verb are

detected. However, the topic NP is not assigned to one
of these two gaps, but rather is held unassigned until
the second verb is encountered. The first proposition
cannot be comprehended until the head of the second
relative clause is reached because the subject gap of
the first verb is coindexed with that head. Hence,
the first proposition fails to be comprehended locally.
This seems to be the cause of unacceptability of 19.b.

In the parsing literature, it is assumed that the
assignment of a filler to a gap is accomplished
immediately when the gap is identified (Frazier 1985).
Without much discussion, I also assume that in the
parse of Japanese the filler is immediately assigned to
a gap when it is detected. Under this assumption the
parser successfully comprehend each proposition of a
sentence locally if the filler is assigned correctly to
the first detected gap, as we have seen in 19.a. In
19.b, however, the parser initially assigns the topic
NP to the first detected gap and realizes this
assignment is incorrect, resulitng in the parsing
breakdown. Hence, 19.b is unacceptable.

4. Conclusion
This paper discussed a long-noted problem, under
what conditions topicalization from relative clauses
result in unacceptable. I argue that the
topicalization from relative clauses become
unacceptable due to two types of parsing difficulties.
One is that the parser fails to detect an intended gap
due to the preference for a lexical NP over a gap and
the other is that the parser fails to assign the topic
NP to an intended gap due to the immediacy of the
assignment of a filler to a gap.

References
Chomsky, N. (1981) Lectures on Government and Binding.
    Foris: Dordrecht.
             (1986) Knowledge of Language: Its Nature,
    Origin, and Use. New York: Praeger.
Clancy, P.M., H.Lee, and M.Zoh (1986) 'Processing
    Strategies in the Acquisition of Relative Clauses:
    Universal Principles and Language-specific
    Realizations,' Cognition 14, 225-262.
Clifton, C. and L,Frazier (1989) 'Comprehending
    Sentences with Long-distance Dependencies,' In
    G.N. Carlson and M.K. Tanenhause (eds), Linguistic
    Structure in Language Processing, 273-317. Kluwer

Academic Publishers.

Fodor, J.D. (1978) 'Parsing Strategies and Constraints on Transformations,' Linguistic Inquiry 9,427-473.

Hasegawa, N. (1981) A Lexical Interpretive Theory with Emphasis on the Role of Subject. University of Washington dissertation.

Inoue, A. (1984) 'On the Sentence Processing of Japanese,' ms. University of Connecticut.

Kornfilt, K., S. Kuno, and E. Sezer (1980) 'A Note on Crisscrossing Double Dislocation' In S.Kuno (ed), Harvard Studies in Syntax and Semantics, Harvard University, Cambridge, MA.

Mazuka, R. (1990) 'Processing of Empty Categories in Japanese,' ms. Duke University.

Mazuka, R. et al (1989) 'Processing of Japanese Garden Path, Center-embedded, and Multiply-left-embedded Sentences, Annual Bulletin of Research Institute of Logopedics and Phoniatrics 23. 187-212.

Saito, M. (1985) Some Asymmetries in Japanese and their Theoretical Implication. MIT dissertation.

Ue N. (1982) A Crossing Constraint in Japanese Syntax. University of Michigan dissertation.