# JAPANESE LANGUAGE SEMANTIC ANALYZER BASED ON AN EXTENDED CASE FRAME MODEL

Akira Shimazu, Syozo Naito, and Hirosato Nomura
Musashino Electrical Communication Laboratory, N.T.T.
3-9-11, Midoricho, Musashino, Tokyo, 180, Japan

## ABSTRACT

This paper describes a Japanese language semantic analyzer based on an extended case frame model, which consists of a relatively large collection of case relations, modalities and conjunctive relations. The analyzer performs four stage analysis using a frame type knowledge base. It also utilizes plausibility scores for dealing with ambiguities and local scene frames for the prediction of omitted case elements.

## 1 INTRODUCTION

Many natural language understanding systems employ the concept of Case Frames [Bruce, 1975], which is suitable for Japanese sentence analysis. However, most systems deal with rather simple sentences, in which the analysis is a relatively easy task. However, in the case of more complex everyday sentences, a case analysis is not a simple matching of case frame slots and phrases (candidates of case elements). Moreover, these sentences result in a variety of problems such as an increase in the possible number of semantic structures in a sentence (ambiguities), the presence of various modality expressions, and frequent occurrences of ellipses.

As part of a research program on language understanding, the authors are observing actual sentences in scientific or engineering articles, and are developing a language understander. The observations and the language understander are based on an extended case frame model. The characteristics of the model are 1) its relatively large collection of case relations for distinguishing between meanings, and 2) emphasis on modalities and conjunctive relations (relations between sub-sentences in a sentence).

The analyzer, designed in accordance with the model; 1) performs four stage analysis using a frame type knowledge base, 2) constructs a frame type semantic structure (FTSS), 3) uses plausibility scores to determine a better structure within the possible number of semantic structures in a sentence, and 4) predicts omitted case elements by using local scene frames.

## 2 Extended Case Frame Model

An extended case frame model used as an understanding model consists of predicates, semantic case relations (roles), modalities, and conjunctive relations (Table 1, Appendix 1).

### 2.1 CASE RELATIONS

One characteristic of this model is the relatively large collection of case relations. A small set of case relations is not sufficient for representing the differences in various expressions, which is a significant feature of natural languages. Therefore, from the viewpoint of language understanding and analysis, a large collection of relations is better for capturing the richness of the language phenomena as has been pointed out by Minsky [Minsky, 1975].

The case collection is based on the following criteria; 1) adoption of surface cases, 2) differentiation between the meanings of various linguistic expressions, 3) adoption of a semantic interpretation model rather than primitive decomposition [Bobrow and Winograd, 1979], [Lehnert, 1979]. The interpretation is applied to a) different case views of the same phrase (a6 in "Tom gave Mary a present", "Tom" is an agent of "gave" and source of "a present" movement) [Charniak, 1981], b) a hierarchy between analogous cases (Appendix 1), and c) different expressions of the same meaning (as in "add six to five." and "add six and five.").

Our collection includes; 1) basic cases like Fillmore's case system, 2) case relations such as extent, manner, and degree, which appear as adverbial phrases, and 3) additional cases represented by the inflection of a certain verb such as "tsuite" (concerning).

Table 1 Extended Case Frame Model

```
sentence = simple_sentence |
     sentence + conjunctive_relation + sentence
simple_sentence =
     predicate + case_relations + modalities
ca8e_relations = see Appendix 1
modalities - tense | aspect | manner | intention
     | guess | attitude | negation | ability |
     necessity | ...
conjunctive_relation ■ time | cause | reason |
     result | contrast | goal | assumption |
     circumstance | ...
```

### 2.2 MODALITIES

In addition to these propositional matters, semantic structures must include information on modality such as tense, aspect, intention, manner, attitude and assumption. Real (not toy) sentences convey much of this type of information. Unfortunately, however, most understanding systems and linguistic theories have not dealt sufficiently with modalities.

### 2.3 CONJUNCTIVE RELATIONS

A sentence generally consists of several subsentences, each of which represents a unit event. In the model, event relations are expressed by conjunctive relations such as time, cause, reason, result, goal, assumption, contrast and circumstance. These are essential for understanding a sentence, and must be organized into semantic structures.

## 3 Analysis Framework based on the Model

The design policies of the semantic analyzer are; (a) to reflect the case frame model, and (b) to integrate semantic and syntactic analyses. In accordance with these policies, the semantic analyzer performs four stage analysis using a knowledge base. The knowledge base is made of frame systems, and consists of lexical and syntactic structure information, case frames, concept relations, event conjunctive relations, common sense knowledge, and expertise.

```
CASE90: [a CASE with (*CASE-NAME UNFILLED)
         (♦CASE-INSTANCE CASE-INSTANCE89)
         (♦CASE-MARKER KARA-K (from))
         (SCORE 0) ... ]

CASE-INSTANCE89:
        [a CASE-INSTANCE with (*TYPE COMPLEX)
         (*CATEGORY OBJECT) (*INSTANCE COMPLEX88]

COMPLEX88: [a COMPLEX with (*MODIFIERS MODIFIER87)
            (*MODIFICANT CASE-INSTANCE88) ... ]

MODIFIER87:
        [a MODIFIER with
         (♦RELATION-CATEGORY WHOLE-PART)
         (*MODIFIER-INSTANCE CASE-INSTANCE55)
         (♦RELATION-INSTANCE NO-K (of) ... )

CASE-INSTANCE55 ...
        "KYUUKOU DENSHA' (express train),
         kind of DENSHA (train).

CASE-INSTANCE88 ... "MADO' (window),
         part of "DENSHA' (train).
```

Figure 1 Semantic Structure (abbreviated) to a Case Element

### 3.1 Four Stage Analysis Process

This analysis consists of morphology, case element, simple sentence, and whole sentence stage. In each stage, the analyzer uses information obtainable in its own stage. In the morphological stage, the analyzer relates a string of characters to words, which will be related to concepts in the latter stages. Case element candidates analyzed in the second stage are simple noun phrases (not including embedded sentences) or adverb phrases. Analyses of case relations, modalities, and event conjunctive relations are done in the third and the fourth stages. The analyses of embedded sentences are also accomplished in the fourth stage.

### 3.2 Growth of a Semantic Structure

As the analysis proceeds, more information is used and a FTSS of the sentence grows. For example, as for the underlined part of the sentence,

"KYUUKOU DENSYA NO MADO KARA TSUUKA EKI NO EKIMEI HYOUZIBAN WO MIRU TOKI. KANTAN NA EKIMEI NARABA HITOME DE YOMITORU KOTOGADEKIRU GA, ZISUU GA 001 TO NAKANAKA MUZUKASHII."
(When we see the signboard of a passing station from an express train's window, we can read it if the name is simple. However, it is difficult in the case of a lot of characters.),

the second stage analysis constructs FTSSs for case element candidates, which include compound nouns. FTSSs represent concepts and their relations in addition to surface morphological expressions. The frame structures represent KYUUKOU DENSYA (an express train) and MADO (a window) as prototypes (not a constant), where MADO is represented as a part of KYUUKOU DENSYA using the knowledge base (Figure 1). In the third and fourth stages, the semantic structure grows to

```
SSENT96: [a SIMPLE_SENTENCE with
          (♦PREDICATE PREDICATE97)
          (♦CASE CASE90 CASE98)
          (♦UNFILLED-SLOT (AGENT ... ))
          (SCORE 1]
         ;assumption of SSENT121.
         ;happens at the same time.

CASE90: [a CASE with (♦CASE-NAME SOURCE)
         (♦CASE-INSTANCE CASE-INSTANCE89) ... ]

PREDICATE97 ...
        "MIRU' (see) with (MODALITIES ...)

CASE-INSTANCE89 ... "MADO' (window),
        part of 'KYUUKOU DENSHA' (express train)

CASE98 ... "EKIMEI HYOUJI BAN' (signboard),
        at "TSUUKA EKI' (passing station)
```

Figure 2 Semantic Structure (abbreviated) to a Simple Sentence

include case relations, modalities, event relations, and time relations. The frame structures represent KYUUKOU DENSYA NO MADO KARA (from a window of an express train) as a source relation to MIRU (see), and the underlined part is represented as a reference time for the rest of the sentence (Figure 2).

### 3.3 Dealing with Ambiguities

The analyzer generally produces multiple possible semantic structures at each stage. To control this multiplicity, the analyzer attaches a plausibility score to each structure. The score represents the matching appropriateness between a sentence and the knowledge base as will be described in the case analysis in section 4.1. As a result of the analysis, a better FTSS is applied to the succeeding process such as a structural transfer and a query command generation. If necessary, the next better FTSS is applied. FTSSs with the same score are dealt with as ambiguities.

## 4 Analysis Characteristics

### 4.1 Matching Case Frame Slots with Phrases

Filling case frame slots is not a simple task. This is because; (a) meanings can be variously expressed, (b) obligatory cases are not necessarily present, especially in Japanese, (c) there may be sub-sentences between a case and a predicate for case ellipses, (d) some cases must be present to determine the meaning of a predicate, and (e) there may be optional cases.

Considering such phenomena, a sophisticated pattern matcher performs case analyses using case frames and an agenda control [Charniak, 1980], and attaches plausibility scores to each possible FTSS. The matcher predicts case slots (relations) from a particle as a human does, and compares a case slot with a case element by checking semantic categories including inheritance, particles, obligatoricity, and specific information such as the fact that a certain word requires another word to have a certain meaning. A case slot is taken from a (obligatory) case frame given to each verb (or adjective), a default case frame (used when there are no case frames), or an optional case frame. The matcher assigns an approapriate score to a FTSS by subtracting points for each unmatched checking element in the comparison. When the match is not complete and there are other possible case slots, the matcher tries another matching.

### 4.2 Filling Elliptical Cases

Cases and events in a sentence are considered to constitute a local scene frame (LSF). It is recognized that an omitted case element is generally present as the same or similar case in a preceding simple sentence. Based on such a phenomena, the analyzer constructs a LSF as the analysis proceeds, and fills the unfilled slots from the LSF.

### 4.3 Analysis of Modality

Modalities, which characterize natural languages, are expressed by various words such as auxiliary verbs, particles, and adverbs, whereas in the extended case frame model, adverbs are regarded as special case relations. Each modality is analyzed by the semantic categories (features) or the attached procedure (in the case of ambiguities) of each word. The procedure uses a verb's semantic categories, the information in the case slots, and contextual information.

### 4.4 Analysis of Conjunctive Relation

A series of simple sentences (events and states) expressed in a sentence constitutes a discourse, as does a series of sentences in a story. To grasp such a discourse, semantic conjunctive relations are analyzed by syntactic relations between simple sentences, the semantic characteristics of predicates, and event/state relations. As a part of this analysis, time relations between events/states are analyzed according to the types of relations between simple sentences, case elements involving time, and tense marker (auxiliary verb "ta").

## 5 Conclusion

This paper describes ongoing research on a Japanese language semantic analyzer based on an extended case frame model. The first version is being used in a Japanese multiwindow editing system JMACS developed by Maclisp on DEC 20. It has been applied to an intelligent machine translation system LUTE (Appendix 2), and a question answering system. A more advanced version is under development.

## REFERENCES

[I] Bobrow, D. G. and T. Winograd: KRL: Another perspective, Cognitive Science, 3, pp. 29-42, 1979.
[2] Bruce, B: Case Systems for Natural Language, Artificial Intelligence, vol.6, no.4, pp. 327-360, 1975.
[3] Charniak, E, et al.: Artificial Intelligence Programming, LEA, 1980
[4] Charniak, E.: The Case-Slot Identity Theory, Cognitive Science 5, pp. 285-292, 1981.
[5] Lehnert, W.: A Critical Perspective on KRL, Cognitive Science, 3, pp. 1-28, 1979.
[6] Minsky, M. A.: Framework for Representing Knowledge, in P. H. Winston (ed.), the Psychology of Computer Vision, 1975, McGraw-Hill.

Appendix 1 Case Relations Examples

OBJECT-TYPE:
"Tarou ha (agent) Jirou to (co-object) benkyosuru."
(Tarou studies with Jirou.)
"sore o (object) VIS to (secondary-object) yobu."
(We call it VIS.)
"Tarou ha (object-modified) Jirou yori
(object-compared) se ga takai."
(Tarou is taller than Jirou.)
"Tarou ha (object-predicated) gakusei
(predicative-object) da (be)."
(Tarou is a student.)
"18do to (additional-obj.) jouon ni naru."
(It got normal temparature, 18 degrees.)

METHOD-TYPE:
"jitensya de (method) tsuugakusuru."
( ... go to school by bycicle.)
"tamago de (material) tsukuru."
( ... make it with eggs.)
"koukyou de (cause/reason) moukeru."
( ... earn because of prosperity.)

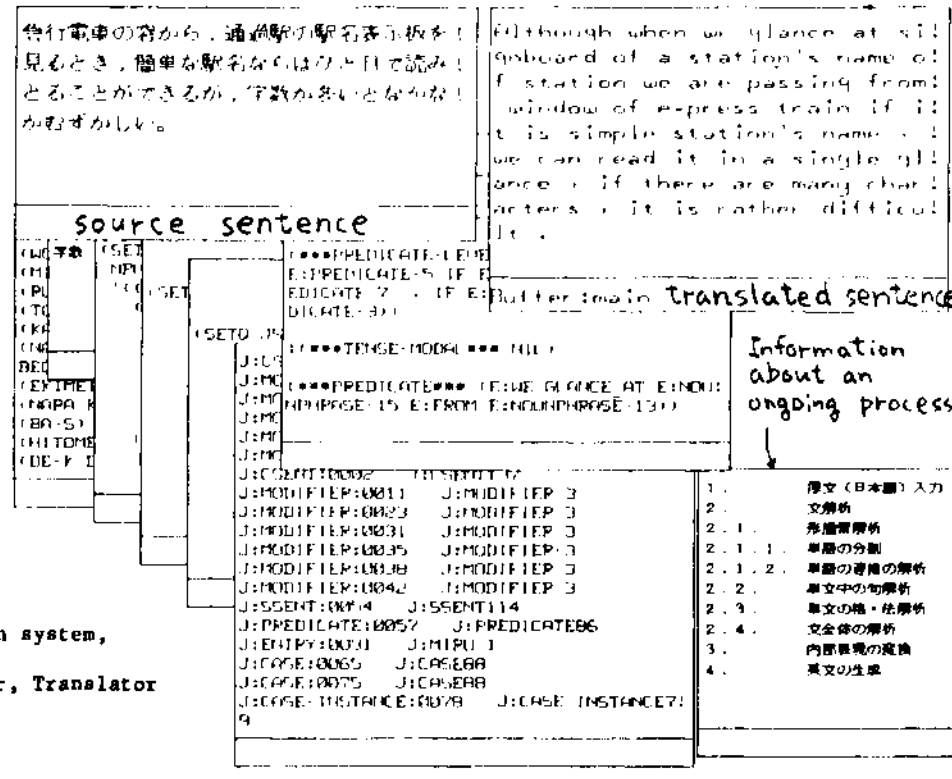
DIRECTION-TYPE:
"mado kara (source) yama o miru."
( ... see a mountain from a window.)
"karera niyoruto (news-origin) ... "
(From what they said, ... )
"minami ni (destination) susumu."
( ... proceed south.)
" ... kanozyo ni ai ni (aim) kita."
( ... came here to see her.)
" ... shouri to (result) naru."
( ... result in a victory.)

TIME&SPACE-TYPE:
"shako ni (location) kuruma o oku."
( ... put a car in the garage.)
" ... asu (time) iku."
( ... go tommorow.)
" ... 30 pun (duration) oyogu."
( ... swim for 30 minutes.)
"yoi tenki no baai (circumstance) ... "
(in the case of good weather ... )

SUPPLEMENT-TYPE:
"kyonen nitaishi (distinction) kotoshi wa ... "
( ... this year against last year.)
"kanojo nitsuite (object-described) hanasu."
( ... talk about her.)
"eigo nokawarini (substitution) doitsugo o narau
( ... learn German instead of English.)
"yuujin toshite (object-as) ... iu."
( ... say ... as a friend.)
"waga kuni nitotte (object-for) rieki da."
( ... is advantageuos for our country.)

MODIFICATION-TYPE:
"shibashiba (rate) okoru."
( ... happen frequently.)
"tadashiku (extent) kotaeru."
( ... answer correctly.)
"osoraku (guess) ame ga furu daro."
(Probably it will rain.)
"meiryou ni (manner) arawareru."
( ... appear clearly.)

---




Appendix 2

Display of
a Machine Translation system,
LUTE
(Language Understander, Translator
& Editor)