Analysis of Japanese Sentences by Using Semantic and Contextual Information (II)-Contextual Analysis

Makoto NAGAO*, Jun-ichi TSUJII* and Kazutosi TANAKA*

Abstract

In the previous paper, we explained the semantic descriptions of verbs and nouns. In this paper, we explicate how contextual information is organized in our system and how contextual information and semantic descriptions are combined in order to perform contextual analysis; for example, to fill up omitted words from preceding sentences.

In most of the artificial intelligence approaches to computer understanding of natural language, logical expressions are used to represent context, and rather complex logical operations are required to perform even a simple contextual analysis. On the contrary we memorize context in the form similar to the short-term memory of human beings. According to several heuristics the words in the memory are re-ordered in order to emphasize the theme words which express the key subjects of the sentences.

We describe the detailed construction of this memory. Several examples of complicated sentences parsed by our system are also shown in this paper.

1. Introduction

In order to resolve the syntactic ambiguity of a sentence, it is also necessary to utilize contextual information obtained from preceding sentences. When one knows a certain event has occurred, he can anticipate successive events that will occur and what changes the objects participating in the event will undergo. This kind of expectation plays an important role in understanding sentences. Various kinds of associations cluster conceptually around individual activities. One can perform contextual analysis of language by explicating these associations.

2. Verb Descriptions for Contextual Analysis

We append various kinds of empirical knowledge to the case frames of verbs. The following two items are described for each case frame in the verb dictionary:

(1) CON : this refers to the consequent activities which are likely to follow the activity of the verb, but not necessarily.

This paper first appeared in Japanese in Joho－Shori (Journal of the Information Processing Society of Japan), Vol. 17, No. 1 (1976), pp. 19～28.

* Department of Electrical Engineering, Kyoto University

(2) NTRANS : this refers to the resultant effects on objects in view of how the objects are influenced by the activity. In our system the influence on the objects is described by the following three expressions:

(a) (ADD case a-set-of-(A V)-pairs)

(b) (DELETE case a-set-of-attributes)

(c) (CREATE lexical-name-of-an-object a-set-of-(A V)-pairs)

(a) means that the object in the case indicated by the second element comes to have a set of properties indicated by the third element. (b) is for the deletion of a set of properties from the object. (c) shows that some objects will be created by the activity. Figure 2.1 shows the description of the verb TOKASU (melt, dissolve).

```
(TOKASU     (   CF
 melt
 dissolve
            (((  ACT    NINGEN)(OBJ    KOTAI)(INST))
                 (human beings)    (solid)

                (NTRANS    (ADD  OBJ (STATE  EKITAI))))
                                          (liquid)

            (((  ACT    NINGEN)(OHJ    KOTAI)(IN  EKITAI))
                 (human beings)    (solid)    (liquid)

                (NTRANS  (CREAT    YOUEKI
                                  (solution)

                          (  SOLVER     (  IN))

                          (  SOLVENT    (  OBJ)))))    ))
```

Figure 2.1 Example of an NTRANS expression

3. Noun Stack and Hypothetical Noun Stack

The analysis of a sentence is primarily grounded in the semantic description --case frame -- of a main verb. Contextual analysis is mainly grounded in accumulated information about nouns. The objects or concepts that are the themes of the sentences and what has been predicated of them can usually be characterized in terms of the nouns appearing in the sentences;and these offer important clues for contextual analysis.

We assign a different LISP atom (produced by the LISP function 'gensym') to each noun which appears. Information about each is entered on the respective property list. We can retrieve all the information about an object from this property list, such as what predication was made on the object, in what event the object participated, and so on. We call these LISP atoms assigned to individual objects Noun Atoms and stack them on two special stacks, NS (Noun Stack) and HNS (Hypothetical Noun Stack). These two stacks are used as an index attached to the semantic network established by preceding sentences. Through these two stacks, we can retrieve necessary information about context.

When we start to analyze a sentence, we stack a list of noun atoms which are assigned to the nouns in the sentence. After the analysis of the sentence, these noun

atoms are reordered according to their degrees of importance. To decide how important a word is, we use the following heuristics.

(1) In Japanese, a theme word is often omitted or expressed by a pronoun in succeeding sentences after it appears once.

(2) A theme word may appear as a "subject" in the surface structure.

(3) The importance of a head noun in a noun phrase is greater than that of other nouns.

By this ordering algorithm, we can retrieve focused words from context more easily than unfocused ones.

We show an example which cannot be properly analyzed without HNS.

```
SHOKUEN  5gr          -O     MISU   100cc  -NI   TOKASU
salt     five grams (OBJ)  water  100cc  (IN)  dissolve.

KONO   SUITYOUEKI  -WA - - - - - - - - -
the    solution    (SUBJ)

(Someone) dissolves 5 grams of salt in 100cc of water.
The solution - - - - - - -
```

In this example, though the demonstrative KONO (the, this) is used, the object referred to does not appear explicitly in the preceding sentence. The object referred to is produced as the result of the event which is expressed by the preceding sentence. As mentioned before, this newly created object is expected by using the NTRANS description in the verb dictionary. We stack this kind of object in HNS instead of NS.

4. Estimation of the Omitted Words and Anaphoric Expressions

In order to supply omitted words drawing from preceding or succeeding sentences, we must be able to:

1. recognize that word is omitted and
2. search for an appropriate word to fill the gap.

Our contention is that an individual syntactic unit such as a noun phrase or a simple sentence conveys a definite idea; a noun phrase may designate a certain definite object, a concept or whatever, and a simple sentence may describe a definite event. In order that a simple sentence describe a definite event, each intrinsic case element of the case frame must be specified by particular objects. We can detect an omitted word by searching for unspecified case elements in a case frame. Moreover, we can guess from the case frame what kind of nouns should be supplied to fill any gaps.

The omitted words in a noun phrase are properly supplied in similar fashion. Most nouns have definite meaning by themselves and are regarded as entity nouns. However,

some kinds of nouns have relational meaning. That is to say, they have slots in their meanings to be filled in by other words in order that they express definite ideas. Sometimes a relational noun is used alone in a noun phrase. In this case, the relational noun must be semantically connected with other words which are omitted in the present noun phrase.

As the final step in the analysis of a noun phrase, we check whether there remain relational nouns which have no definite meaning. If found, we search through NS for words which are suitable to fill in the slots of the nouns. The searching process is the same as for omitted words in simple sentences.

5. Detailed Description of the Trapping List (TL)

Most anaphoric expressions and omitted words are well analyzed by searching through the preceding sentences. However, we need sometimes to refer to succeeding sentences in order to analyze a sentence properly. The sentence shown in Figure 5.1 is an example.

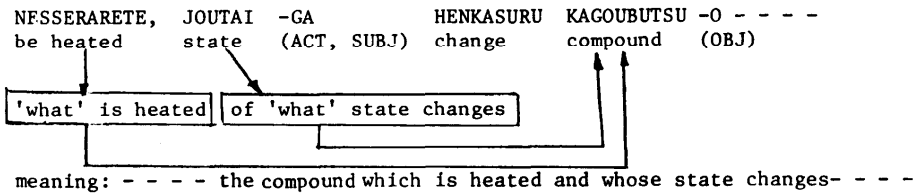


Figure 5.1 Example where omitted words appear in succeeding sentences

To solve this problem, we set up a trapping list TL. Then we cannot find an appropriate word in the preceding sentences for an omitted word or anaphoric expression, we put a new trapping element in TL. When a noun phrase in a succeeding sentence is analyzed, we pick up nouns from the noun phrase one-by-one and check whether the present noun can resolve a pending problem in TL by evaluating the function in the trapping element.

By using the idea of TL, we can separate various checking mechanisms from the main program. They can be invoked automatically when a noun appears in a sentence.

6. Conclusion

We can summarize our interpretive procedure as follows:

(I) Through the use of a grammatical case we describe patterns of activity in the verb dictionary. The descriptions also contain information as to how activities are connected with each other and how activities change objects.

(II) The meaning descriptions of nouns are based upon the upper and lower concept relationships and attribute value pairs. Some kinds of nouns are regarded as having relational meanings.

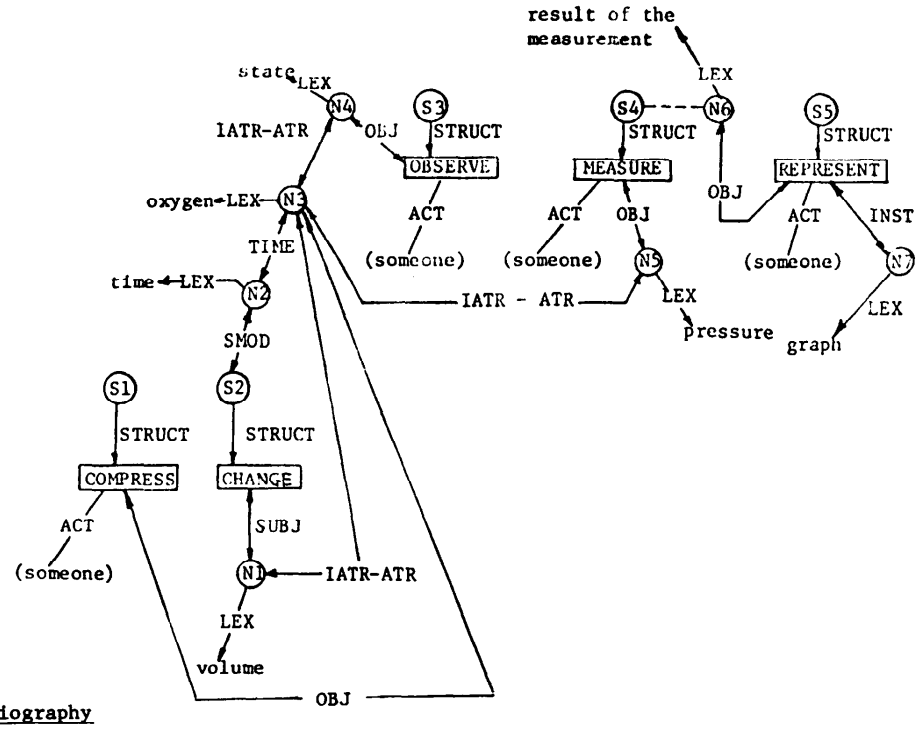
(III) We do not use logical expressions to represent context. Contextual information is represented in the form of semantic network. This, in combination with the semantic descriptions of words, has enabled us to perform efficient analysis dependent on contextual information

The following is an example sentence analyzed by the parser.

Input sentence:

ASSHUKUSARETE,	TAISEKI	-GA		HENKASURU	TOKI	-NO	SANSO	-NO
be compressed	volume	(ACT SUBJ)		change	when		oxygen	
JOUTAI	-O	KANSATSUSHI,	SONO	ATSURYOKU	-O	SOKUTEISHI,	SORE	-O
state	(OBJ)	observe	the	pressure	(OBJ)	measure	it	(OBJ)
GURAFU	-NI		SURU.					
graph	(IOBJ, RESULT, etc.)		represent					

meaning: (Someone) observes the state of oxygen which is compressed and whose volume changes. (Someone) measures the pressure and represents it as a graph.



Bibliography

Nagao, M. and Tsujii, J. : 1976, "PLATON - A New Programming Language for Natural Language Analysis", AJCL

Nagao, M. and Tsujii, J. : 1976, "Analysis of Japanese Sentences by Using Semantic and Contextual Information", AJCL