

GENERATION OF JAPANESE SENTENCES FROM CONCEPTUAL REPRESENTATION

Shun Ishizaki

Electrotechnical Laboratory
1-1-4 Umezono Sakura-mura Niihari-gun
Ibaraki, Japan 305

ABSTRACT

This paper describes an attempt to generate Japanese sentences from conceptual representation. This generator infers the temporal order of events included in the conceptual representation using causal chains and MOPs. Appropriate conjunctives between the events and case markers for subjects are used based on the representation. This generator was first built as part of the Machine Translation Project in the Computer Science Department of Yale University in 1982. It has been subsequently improved at ETL in Japan. About 15 stories are parsed into conceptual representations from Spanish newspaper stories (Lytinen and Schank, 1982) and then Japanese sentences are generated.

I Introduction

This is an attempt to generate Japanese sentences from conceptual representation obtained from Spanish newspaper stories [1]. This generator was first built as part of the Machine Translation Project in the Computer Science Department of Yale University in the summer of 1982. It has been subsequently improved at the Electrotechnical Laboratory in Japan.

Conceptual representation is based on the Conceptual Dependency theory [2] and also on Memory Organization Packets (MOPs) [3]. The general idea behind MOPs is to store knowledge which is common to many different situations in only one processing structure, and then to make this processing structure available for use in a variety of appropriate situations. MOPs are useful for a Japanese generator because their contents are independent from the language, although CD theory and MOPs are constructed by using English. Such independency may be verified by applying these theories to Japanese sentences, structurally different from Romance languages, such as English, Spanish and French.

This generator has several features. First, it infers the temporal order of events included in the conceptual representation using causal chains and MOPs. Such inference is necessary to generate Japanese sentences because most clauses and sentences are ordered temporally in Japanese. Second, appropriate conjunctives are inserted between the events using their causal relations. Third, subjects are omitted in relevant situations and two case markers for subjects are used in the generator to make the sentences suitable for Japanese. Other features are described in the following sections.

Input stories are extracted from Spanish newspapers. All the stories concern terrorism and each consists of a few sentences. The parser analyzes them to obtain their conceptual representations by using appropriate MOPs and causal relations among events in the stories. About 15 stories are parsed into conceptual representations and then Japanese sentences are generated.

II JAPANESE LANGUAGE STRUCTURE

The word order in Japanese is somewhat different from that in English. The main differences are (1) the verb is the last element in a clause, (2) negation is expressed by endings attached to the verb (3) adjectives and adverbs always precede their modified nouns, verbs and other words.

The verb determines the structure of the sentence in Japanese as well as in English, although the verb is placed at the end of the sentence. The role of the verb is one of the reasons why CD theory is also effective for Japanese.

The most important difference between English and Japanese is in the order of clauses, phrases, and sentences. In English, clause modifiers usually follow the modified word. But modifiers are never placed after the modified words in Japanese. Subordinate clauses are never placed after their main clauses. All the clauses are temporally ordered in Japanese. The only exception is a goal clause. There are, of course, exceptions in literary expressions but this generator does not treat them.

III CONCEPTUAL REPRESENTATION

The Japanese generator starts from the conceptual representation which the parser extracts from Spanish newspaper stories. For example, Fig.1 shows a conceptual representation for Story 1. Fig.2 shows Japanese sentences generated from the representation and their correspondence with English words.

Each conceptual representation is expressed by a tree structure. Nodes of the tree are either nouns or events (verbs). Each node has its own concept. For example, event node EXE0 in Fig.1 has EXECUTE as its concept. Its actor is HUM11, its object is HUM9, and the place where the event occurred is LOCO. IR-FROM is a causal relationship which means that UN11 caused the event of EXE0.

```

CONCEPT M-MOCK-TRIAL
ACTOR HUM11=
    CONCEPT TERRORIST
    ORG OBJ3= CONCEPT TERRORIST-ORG
    MEMBERS HUM11
    GENDER MALE
    TYPE GUERRILLA
    WEARING OBJ0= CONCEPT CLOTHING
    TYPE SUIT
    COLOR OLIVE-COLORED
OBJECT HUM6= CONCEPT PERSON
    NUMBER AT-LEAST 60
    RESIDENCE COUNTRY
SCENE2 ACC1= CONCEPT ACCUSE
    OBJECT HUM6
    BAD-ACT UNDI
    ACTOR HUM11
    IR-FROM UNDI
SCENE4 EXE0=
    CONCEPT EXECUTE
    ACTOR HUM11
    PLACE LOC0= CONCEPT CITY
    NAME SAN PEDRO PERULAPAN
    OBJECT HUM6
    IR-FROM UNDI
SCENE1 UNDI= CONCEPT UNDESIRABLE-ASSISTANCE
    JUDGER HUM11
    OBJECT OBJ5= CONCEPT GOVERNMENT
    ACTOR HUM6
SCENE3 TRY0= CONCEPT TRY
    ACTOR HUM11
    OBJECT HUM6

```

Fig.1 Conceptual Representation for Story 1.

```

sukunakutomo 60 nin no noumintachi ga seifu ni
AT-LEAST 60 PEASANTS GOVERNMENT
kyoryokushita node, terososhiki ni zokusu
COLLABORATED BECAUSE TERRORIST-ORG
oribuiro no fuku o kita geriratachi wa sono
OLIVE-COLORED SUIT WEARING GUERRILLAS THE
noumintachi o kokuhatsushite, saiban-ni-kaketa.
PEASANTS ACCUSED TRIED
sorekara, sono geriratachi wa san pedomo perurapan
THEN THE GUERRILLAS SAN PEDRO PERULAPAN
toiu machi de sono noumintachi o shokeishita.
CITY THE PEASANTS EXECUTED

```

Fig.2 Generated Japanese Sentences and Correspondence with English Words.

The generator has a Japanese dictionary. It has Japanese entries for verbs, nouns and others which correspond to the concepts and the proper nouns such as shown in Fig.1.

IV GENERATION OF JAPANESE SENTENCES

One of the main functions of this generator is the inference of temporal order of events using MOPs and causal chains. The temporal order is important because most clauses and sentences are ordered temporally in Japanese. The parser gives the conceptual representation and its main node to the generator. The generator first extracts all the

events connected to the main node.

Then the events are sorted temporally and appropriate conjunctives are inserted between them. For example, in Story 1,

(UND1 node, ACC1, TRY0, sorekara EXE0) is the whole structure to be generated for the Story 1, where all the events are ordered temporally and two conjunctives are inserted there. The generator starts to translate the left side events and then processes the next to the right until the list becomes NIL.

A. Temporal Ordering of Events Using MOPs

1. Ordering With Scenes in MOPs

Every event is related to another event by a causal chain or a scene relation in MOPs. The causal chain, indicated by such relations as LEAD-TO, REASON, GOAL, PRECONDITIONS etc. specifies its temporal order between the two events. The scene relation shown in Fig.1 gives a temporal order with other scenes in the MOP; SCENE1, SCENE2, etc. are in a temporal order. For example, in the MOP of M-POLICE-CAPTURE, SCENE1 is a description of a crime and the scene causes SCENE2, POLICE-SEARCH. As a result of the scene, AUTHORITY will arrest BAD-GUY in SCENE3. These events are ordered temporally using these scenes if ACTORS and OBJECTS satisfy the conditions at each scene.

2. Ordering With Causal Relations And MOPs

The parser extracts causal relationships among events in stories when some of the events can not be expressed with scenes in MOPs.

However there are some cases where the temporal orders can not be determined by the causal chains given by the parser. For example, in Story3, the main node contains three events:

```

KIL2 : A POLICEMAN KILLED A CRIMINAL.
TRA0 : A POLICEMAN TRANSPORTED A CRIMINAL.
ESC1 : A CRIMINAL TRIED TO ESCAPE.

```

The input story has a structure such as, KIL2 during TRA0. KIL2 when ESC1.

Then the parser gives the following temporal orders in the conceptual representation,

```
TRA0 → KIL2 and ESC1 → KIL2.
```

The temporal relation between TRA0 and ESC1 is not clear here. The generator has to infer the relation using MOPs and the conceptual representation to generate Japanese sentences.

Connectivity of locations is one of the clues of inference when either of the event has a concept of PTRANS(position transition). If the location where ESC1 occurred is confirmed to be the destination or on the way of TRA0, then the generator can infer that ESC1 occurred during or after TRA0 occurred. But there is no such information in the conceptual representation of Story 3.

Next, the generator looks at CONTROL and DISABLE-CONTROL relation. The concept of TRA0, TRANSPORT-CRIMINAL, has the template as follows,

```

TEMPLATE ACTOR AUTHORITY
          OBJECT BAD-GUY
          PRECONDITIONS (CONTROL ACTOR OBJECT).

```

This template means that an authority keeps a bad-

guy under control during the transportation. On the other hand, ESCAPE is the ESC1's concept and it has an antecedent, DISABLE-CONT. Therefore, ESCAPE satisfies the properties of DISABLE-CONT. After checking the correspondence between ACTOR of TRA0 and OBJECT of ESC1 and also OBJECT of TRA0 and ACTOR of ESC1, the generator infers that the temporal order of the events should be TRA0 → ESC1. The whole structure of Story 3 thus becomes

(TRA0 toki, ESC1 node, KIL2).

B. EVENT GENERATION

An event is generated by the following functions,

(GEN-SUBJECT), (GEN-VERB-MODIFIERS), (GEN-VERB).
(GEN-SUBJECT) picks up an ACTOR slot from the event. The function chooses an OBJECT slot and the event is expressed in a passive form when the ACTOR slot is not filled.

A case marker "wa" usually follows a subject. Another case marker "ga" is also used often for a subject. The difference between them is so subtle and highly semantic that it is not easy to treat. Roughly speaking, "ga" is used in subordinate clauses which express time, reason, goal, etc. And "wa" is usually used in main clauses. This approximation is adopted in this generator.

(GEN-VERB-MODIFIERS) generates OBJECT and other modifiers for the verb. It has standard pairs of properties and case markers listed below,

((TIME ni) (PLACE de) (FROM kara) (TO e)
(INST de) (OBJECT o)).

These modifiers for verbs are generated with this order. A case marker must be specified at the verb's entry in the Japanese dictionary when the property has another type of case markers.

(GEN-VERB) generates a verb which is appropriately inflected according to the tense and the mode of the event. The tense inflection of the verbs are specified according to their endings.

C. NOUN PHRASE GENERATION

A noun phrase is generated by the following functions when the noun is processed for the first time, (GEN-NOUN-MODIFIERS), (GEN-NOUN). The second time a noun is needed, it is generated by (GEN-PRONOUN).

For example, when CRIMINAL FIDEL GONZALEZ is generated for the first time, (GEN-PRONOUN) is used the next time. The function generates THE CRIMINAL instead of pronoun "he". Because in Japanese, especially in newspapers, pronouns corresponding to "he", "she" etc. are rarely used. There is a tendency to use more specific words shown above than such pronouns.

An order of the modifier generation for a noun is specified as, TYPE NUMBER AGE NAME STATUS etc. Other modifiers not specified here are generated according to the order in the node representation and they precede these modifiers.

D. ELLIPSES

Subjects in main clauses are usually omitted when their subordinate clauses have the same subjects in Japanese. In Story 1 shown in Fig. 1, the subject and the object in event ACCL are same as those in TRA0. In the Japanese translation, not only is the subject omitted in TRA0, but also the object is as shown in Fig.2. An object is not omitted if the two events have no common subject.

V CONCLUSIONS

This generator works well in translating conceptual representation into Japanese sentences during the processing of 15 terrorism stories. Conceptual representation is expressed to be language independent if the culture or the customs behind the language are not very different from those of the representation. In those stories, the MOPs are not changed for the generation.

The problem of focus is not discussed in this paper. Word order and nested structure should be taken into account if this problem is treated by the generator. But the temporal ordering of events and other methods proposed here are necessary for generation regardless of the focus problem.

ACKNOWLEDGEMENTS

The author first thanks Professor R.C. Schank for his guidance and encouragement. And he also thanks S.L. Lytinen for many of the functions used in the generator. Additional thanks to D.Littleboy, L. Birnbaum and MT Project members for their important suggestions and cooperations, and to Dr. H. Tanaka who supported this work at ETL.

REFERENCES

- [1] Lytinen, S.L. and R.C. Schank, "Representation and Translation." Research Report #234, Dept. of Computer Science Yale Univ. 1982.
- [2] Schank, R.C. "Conceptual Information Processing." North-Holland 1975.
- [3] Schank, R. C. "Language and Memory." Cognitive Science Vol.4 No.3 1980.
- [4] Tanaka, H. "Natural Language Processing and Linguistics." Circulars of ETL No.205 1981.
- [5] Lebowitz, M. "Generalization and Memory in an Integrated Understanding System." Research Rep. #186 Dept. of Computer Science Yale Univ. 1980.
- [6] McGuire, R. Programs for Generation on Dec 20, Dept. of Computer Science Yale Univ.