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 (I)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 cource, 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.I 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 EXEO in Fig.I has EXECUTE as its concept. Its actor is HUM11, its object is HUM9, and the place where the event occured is LOCO. IR-FROM is a causal relationship which means that UND1 caused the event of EXEO.

```
CONCEPT  M-MOCK-TRIAL
ACTOR    HUMII=
           CONCEPT  TERRORIST
           ORG    OBJ3=   CONCEPT  TERRORIST-ORG
                          MEMBERS  HUMII
           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  UND1
                ACTOR    HUMII
                IR-FROM  UND1
SCENE4   EXE0=
           CONCEPT  EXECUTE
           ACTOR    HUMII
           PLACE    LOC0=  CONCEPT  CITY
                           NAME    SAN PEDRO PERULAPAN
           OBJECT   HUM6
           IR-FROM  UND1
SCENE1   UND1=  CONCEPT  UNDESIRABLE-ASSISTANCE
                JUDGER   HUMII
                OBJECT   OBJ5=  CONCEPT  GOVERNMENT
                ACTOR    HUM6
SCENE3   TRY0=  CONCEPT  TRY
                ACTOR    HUMII
                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 pedoro 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 Pig.l.

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 inseted between them. For example, in Story 1,

(UND1 node, ACC1, TRYO. sorekara EXEO)

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 bocomes 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.l gives a temporal order with other scenes in the MOP; SCENEl, SCENE2,etc. are in a temporal order. For example, in the MOP of M-POLICE-CAPTURE, SCENEl 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.
TRAO : A POLICEMAN TRANSPORTED A CRIMINAL.
ESCI : A CRIMINAL TRIED TO ESCAPE.

The input story has a structure such as,

KIL2 during TRAO. KIL2 when ESCI.

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

TRAO -» KIL2 and ESCI --> KIL2.

The temporal relation between TRAO and ESCI 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 ESCI occured is confirmed to be the destination or on the way of TRAO, then the generator can infer that ESCI occured during or after TRAO occued. 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 TRAO, 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 ESCI's concept and it has an antecedent, DISABLE-CONT. Therefore, ESCAPE satisfies the properties of DISABLE-CONT. After checking the correspondence between ACTOR of TRAO and OBJECT of ESC1 and also OBJECT of TRAO and ACTOR of ESCI, the generator infers that the temporal order of the events should be TRAO → ESCI. The whole structure of Story 3 thus becomes

(TRAO toki, ESCI 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 approxmation 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 ACC1 are same as those in TRAO. In the Japanese translation, not only is the subject omitted in TRAO, 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." Circular's 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.

Author Index

Abe, N. 1113
Adorni, G. 72
Airenti, G. 78
Allard, D. 789
Allen, B. 340
Allen, J. 741
Ambler, A. 842
Anderson, J. 57
Anlck, P. 653
Antoniou, G. 916
Appelt, D. 573, 595
Araya, A. 439
Asada, M. 1032
Axel, L 1128

Bajcsy, R. 1128
Baker, J. 122
Ballard, D. 1068
Bara, B. 78
Barnard, S. 1062
Barnden, J. 280
Barnes, D. 824
Barnett, J. 774, 786
Barstow, D. 1170
Baskin, A. 256
Bear, J. 696
Bell, W. 87
Berthod, M. 1065
Berwick, R. 459, 710
Bhanu, B. 1136
Bibel, W. 920
Bien, J. 675
Binford, T. 343
Blasius, K.-H. 936
Bledsoe, W. 892
Blicher, A. 1046
Blum, R. 268
Boden, M. 1195
Boggs, W. 653
Bolle, R. 966
Bolles, R. 1014, 1116
Bonnet, A. 185
Bonollo, U. 151
Borning, A. 167
Bradshaw, G. 465
Brady, M. 969
Brooks, R. 799, 845
Brown, C. 1068, 1081
Brown, J. 1158
Bruss, A. 1053
Busemann, S. 643
Bylander, T. 218
Byrd, R. 704

Cammarata, S. 767
Cantone, R. 207
Cappedi, A. 333
Carbonell, J. 653, 668
Castaing, J. 930
Cernuschi-Frias, B. 966
Chamot, D. 1195
Chandrasekaran, B. 218
Cheeseman, P. 198
Chin, D. N. 388
Christaller, T. 588
Church, K. 636
Clarkson, A. 259
Clipplnger, J. 730
Cloarec, J.-F. 385

Cohen, D. 17, 774
Cohen, P. 212, 355
Colmerauer, A. 487
Colombetti, M, 78
Comino, R. 663
Conery, J. 539
Cook, S. 1195
Cooper, D. 966
Copp, D. 116
Corkill, D. 748
Corner, D. 842
Covvey, H. 242
Cowie, R. 952
Cudelou, J.-F. 385
Cuena, J. 246
Curtiss, M. 1081

Dahan, C. 185
Danlos, L 606
Davis, E. 295
Dean, J. 194
DeJong, G. 462
deKleer, J. 1158
Dershowitz, N. 940
Diamond, M. D. 1003
Dilger, W. 321
DiManzo, M. 72
Dodson, D. 536
Doshita, S. 482
Doyle, J. 309, 349
Duchin, F. 1195
Dyer, M. 75

Eder, E. 920
Edman, A. 553
Elfes, A. 830
Ellman, J. 600
Elshout, J. 97
Emde, W. 455

Fang, J.-Q. 1035
Farihas-del-Cerro, L. 926
Faugeras, O. 982, 996
Feuer, A. 550
Filman, R. 365
Finin, T. 176
Fischler, M. 1010, 1014
Fisher, R. 989
Forbus, K. 315
Fox, M. 158
Frey, W. 693, 727
Fribourg, L. 923
Friedland, P. 108
Fronhoefer, B. 920
Fusaoka, A. 405

Gabriel, R. 845
Gallaire, H. 1192, 1199
Ganapathy, S. 1003
Gazdar, G. 556
Gehrke, M. 721
Gemello, R. 663
Genesereth, M. 450
Georgeff, M. 151
Gergely, T. 877
Ghallab, M. 789
Gini, G. 821
Gini, M. 821
Giunchiglia, F. 72

Glicksman, J. 1078
Gray, P. 21
Grechanovsky, E. 807
Greiner, R. 450
Gresse, C. 37
Grinberg, M. 355
Grover, M. 436
Guiho, G. 1
Guida, G. 663
Guoning, S. 164

Habel, C. 455
Hagert, G. 81
Hahn, U. 337
Hallam, J. 792
Jannah, M. 1116
Hansson, A. 81
Hardy, N. 824
Hardy, S. 533
Hayes, P. 668
Hearne, J. 362
Hebert, M. 996
Henschen, L 25, 902
Herman, G. 1128
Herman, M. 1108
Hoeppner, W. 588
Horaud, P. 1116
Hsiang, J. 940
Huang, T. S. 1035
Huertas, A. 1099
Hunt, G. 444
Hyvonen, E. 671

Ichinose, T. 1125
Ishizaki, S. 613
Isomoto, Y. 173
Itho, F. 1113

Jacobs, P. S. 610
Jaffar, J. 500
Jameson, A. 616, 643
Jansweijer, W. 97
Jayaraman, B. 10
John, C. 520
Johnson, P. 29, 236
Jones, K. 621
Jones, M. 678
Josephson, N. 940
Joshi, A. 603
Jouannaud, J.-P. 909

Kaindl, H. 760
Kakusho, O. 173
Kalman, J. 933
Kanade, T. 1108
Kanayama, Y. 834
Kandrashina, E. 346
Kasif, S. 544
Kass, M. 1043
Kassirer, J. 49
Kay, A. 1195
Kehler, T. 108
Kellogg, C. 170
Kender, J. 1104
Kibler, D. 415, 539
Kim, J. 190
King,J. 1199
Klrchner, H. 909

Kiremidjian, G. 259
Kittredge, R. 657
Klahr, P. 859
Kleinosky, P. 158
Kodratoff, Y. 930
Kohli, M. 544
Konollge, K. 377
Konst, L 97
Koomen, J. 741
Kornfeld, W. 514
Kosaka, A. 482
Koskenniemi, K. 683
Kowalski, R. 1192
Kuipers, B. 49
Kundel, H. 1128
Kuroe, S. 1108

Laird, J. 771
Lamping, J. 365
Lander, W. 207
Langley, P. 94, 419, 465
Lassez, J.-L 500
Laumond, J.-P. 839
Lawton, D. 1027
Lebowitz, M. 63
Lee, M. 824
Lehnert, W. 713
Lenat, D. 167, 259
Lesser, V. 748
LeVeen, R. 1128
Lieberman, M. 212
Lipkis, T. 330
Lloyd, J. 500
Long, W. 230
Lou, C. 222
Loveland, D. 182
Lowenfeld, S. 158
Lozano-Perez, T. 799
Lukaszewicz, W. 352
Lux, A. 1121

Mackworth, A. 1187
Maida, A. S. 382
Malik, J. 343
Marburger, H. 588, 643
Markus, A. 66
Marrone, M. 207
Marsland, T. 763
Martin, P. 573
Martins, J. 370
Masui, S. 233
Mauldin, M. 653
Mays, E. 650
McArthur, D. 767, 859
McCarty, L 287
McCune, B. 194
McCune, W. 902
McDermott, D. 45, 302
McDermott, J. 100, 233
McDonald, D. 167, 730
McKeown, K. 582
Meersman, R. 325
Mel'Cuk, I. 657
Mellish, C. 533
Michalski, R. 256
Miller, F. 116
Miller, P. 239
Minker, J. 544

Miranker, D. 850
Mitchell, T. 1139
Mittal, S. 218
Mlzoguchl, R. 173
Moen, J. 236
Moffat, D. 21
Montalvo, F. 365
Moore, R. 272
Moretti, L. 333
Morlk, K. 588, 647
Mostow, J. 40
Motoda, H. 225
Mudler, J. 633
Mueckstein, E.-M. 660
Mukai, K. 547
Mylopoulos, J. 242,1199

Nadathur, G. 603
Nagel, H.-H. 945
Nalto, S. 717
Nakagawa, S.-l. 639
Naor, J. 1086
Narain, S. 859
Naqvi, S. 25
Narasimhamurthl, N. 1003
Nebel, B. 588
Neumann, B. 724
Nevatia, R. 1099
Newell, A. 771
NIlsson, M. 374
Nilsson, N. 1195
Nishlda, T.-A. 482
Nomura, H. 717
Norvlg, P. 624
Novak, H.-J. 724

Ohlbach, H. 916
Ohlsson, S. 426
Ohsuga, S. 391
O'Leary, M. 588
O'Kane, M. 630
O'Keefe, R. 479
Orgren, P. 179
Osherson, D. 566 .

Pacini, G. 862
Pearl, J. 190
Peleg, S. 1086
Pentland, A. 973, 1062
Pereira, F. 573
Peterson, J. 933
Phillips, B. 690
Pinsker, E. 807
Pipitone, F. 207
Plaisted, D. 940
Ponce, J. 982
Pople, H. 108
Popplestone, R. 842
Porter, B. 415
Potmesil, M. 1089
Prade, H. 130
Prazdny, K. 1050

Quinlan, J. 137
Quinqueton, J. 476

Rada, R. 780
Rady, M. 707
Reboh, R. 108, 145
Rector, A. 536
Reddy, U. 10
Relmer, U. 337
Reiter, R. 1199
Remy, J.-L 909
Rendell, L 469
Reyle, U. 693, 727
Rleger, J. 1027
Rissland, E. L 90
Rohrer, C. 727

Rollinger, C.-R. 358, 455
Roos, J.-L. 250
Rosenberg, S. 108, 215
Rosenschein, S. 1180
Rullent, C. 663
Russ, T. 230

Sabah, G. 707
Sage, S. 94
Sakurai, T. 507
Sallantin, J. 476
Samuel, A. 1152
Sauers, R. 110
Sato, M. 507
Schmolze, J. 330
Schoppers, M. 783
Scott, P. 432
Sedlmeyer, R. 29
Seki, H. 405
Shafer, S. 959
Shapiro, D. 194
Shapiro, E. 529
Shapiro, S. 370
Shaw, D. 850
Shepherd, B. 473
Sher, D. 1081
Shibahara, T. 242
Shieber, S. 699
Shimazu, A. 717
Silver, B. 429
Simon, H. A. 465
Sisto, L. 663
Sloman, A. 44
Smith, C. 179
Smith, D. 32
Smith, R. 122, 855
Smith, S. 422
Sobel, A. 233
Somalvico, M. 663
Souvignier, V. 1121
Steeb, R. 767
Steele, G. 845
Steels, L. 395
Steinacker, I. 627
Stevens, K. 1057
Stob, M. 566
Stolfo, S. 116,850
Stuth, B. 1068

Takahashi, K. 405
Talukdar, S. 830
Tanaka, T. 263
Tarnlund, S.-A. 553
Taylor, C. 167
Tenenbaum, J. 1023
Terzopoulos, D. 1073
Thompson, W. 29, 236
Thorpe, C. 959
Tomita, F. 1132
Tong, R. 194
Tropf, H. 1094
Trost, H. 627
Tsai, R. 1038
Tsotsos, J. 242

Udupa, J. 1128
Utgoff, P. 447

Valtorta, M. 162, 777
VanAssche, F. 325
van den Herlk, H. 252
Vershinln, K. 877
Vesonder, G. 116
Vinchesl, C. 333
Vogt, R. 432

Wahlster, W. 588, 643
Walker, A. 526
Wallace, J. 84
Walsh, R. 110
Walter, I. 1094
Walther, C. 882
Wang, J. 222
Webber, B. 650, 1199
Wehrli, E. 686
Weide, G. 398
Weinstein, S. 566
Wesley, L. 203
Weyer, S. 167
Wielinga, B. 97
Wilensky, R. 402
Wilkins, D. 733
Williams, T. 179
Wilson, W. 520
Witkin, A. 1019, 1023
Witkowski, C. 827
Wolf, G. 1128
Wolf, H. 1010
Womann, W. 321
Wood, S. 69
Woods, W. 47
Wos, L 867
Wright, J. 340
Wysotzki, F. 409

Xinsong, J. 164
Xiwen, M. 398

Yachida, M. 1125
Yamada, N. 225
Yin, B. 814
Yu, C. 164
Yuille, A. 969

Zhang, B. 757
Zhang, L 757
Zielinski, J. 116
Zytkow, J. 465