意味情報を用いた日本語の生成
――イェール大学における自然言語処理――

石崎 俊（電子技術総合研究所）

1. まえがき

本報告は筆者が1981年9月から米国イェール大学コンピュータサイエンス学科に1年間滞在し、機械翻訳プロジェクトの一部として作成した日本語生成システムの概要をまとめたものである。

この機械翻訳プロジェクトでは、R. C. Schankの概念依存理論（Conceptual Dependency Theory）[1]に基づいた概念表現（Conceptual Representation）を中心に置いている。すべての対象言語を解析して概念表現で表わし、次に、そこから出発して任意の対象言語を生成することが目標である。

概念表現は、人間の基本的な日常行動を基礎とし、いくつかの行動をまとめて表現したＭＯＰｓ（Memory Organization Packets）[2]を用いる。このような表現形式は特定の言語に依存しないことが期待され、機械翻訳プロジェクトの実施によって見通しが得られると考えられている。

イェール大学の機械翻訳プロジェクトは現在発展中で、英語，フランス語スペイン語，ドイツ語および日本語を対象とし、近く中国語を加える予定である。このプロジェクトはSchank教授を中心に、外国からの客員研究員と外国語を話す大学院生で構成されている。スペイン語入力，英語および日本語出力のシステムが現在動いている。

分析対象は新聞から取り出した20近くのテロリズムに関する記事である。各記事は1～3個の文章からなり、分析によって記事に含まれる事象(Event)の間の意味的関係が得られる。

日本語生成システムは、このような意味構造を持った概念表現から出発して、いくつかの事象の生成順序や適切な接続詞を決定し、概念（Concept）に基づいた訳出語の選択を行って、比較的自然な文章が生成できるように構成されている。[3]

2. 概念依存理論とＭＯＰｓ

概念依存理論は、人間のすべての日常行動（Action）は、いくつかの基本的な行動の組み合せによって表現できると主張する。

PTRANSは位置の移動を表わし、goやputが例である。ATRANSは所有の移動を表わし、例としてgiveやbuyがある。その他MBUILD，INGEST等、合計11種が基本的行動である。

行動の基本形式は ACTOR ACTION OBJECT DIRECTION (INSTRUMENT) で与えられる。一方状態の基本形式は OBJECT (is in) STATE (with Value) で与えられる。このような形式で表現することを概念化（Conceptualization）と呼び、それぞれ行動の概念化，状態の概念化という。

ところで、これらの基本的な行動を組み合せた概念表現よりも高次の意味表現形式として，MOPsが提案されている。

MOPは、ある目標を達成するためのいくつかのSCENEから成り立ち、その中に一つの主となるSCENEがある。ここでSCENEとは、一つの共通の目標を持つ、いくつかの行動をまとめて表現した記憶形態である。

たとえば，M-AIRPLANE（飛行機に乗るMOP）は、チェックインのSCENE、待ち合室のSCENE，飛行機の移動のSCENE等から成り立ち、各SCENEは、いくつかの行動を含んでいる。

3. 概念表現

日本語生成システムが出発点として用いる概念表現は、パーザーがスペイン語の新聞記事を分析して抽出したものである（第8節14図参照）。各記事に対する概念表現は木構造で表わされる。木構造の節（Node）は名詞句と事象（Event）から成り、各節は概念（Concept）を持つ。たとえば、ストーリー1における一つの事象節EXE0は、図1に示すように、概念は"処刑する"（Execute）であり、主語がHUM11，目的語がHUM9、場所がLOC1、この節の上位節はM-M0であることを示している。HUM11, HUM9, LOC1は名詞節である。

```
EXE0   ACTOR      HUM11
       OBJECT     HUM9
       PLACE      LOC1
       SCENE-OF   M-M0
       CONCEPT    EXECUTE
```

図1. 事象の表現例

一方、日本語生成システムは概念の集合を構造化したデータベースと、日本語に関する辞書をあらかじめ持っている。概念構造は図2のように三分類される。

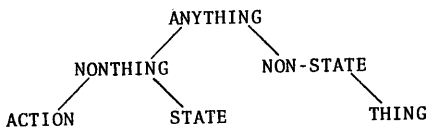

図2. 概念の分類

ACTIONは動詞に対応し、図3のような構造を持っている。

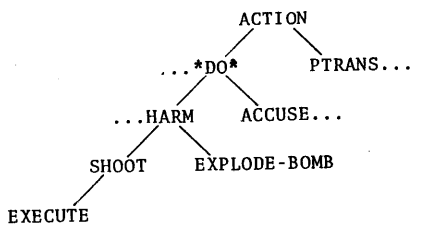

図3. ACTIONの構造

一般に木構造の上下には包含関係があるが、まだ完全に整理された段階ではない。たとえば*DO*は図4のように表現されており、上位概念にACTION、下位概念にACCUSE、HARM等がある。またTEMPLATEで*DO*のACTORはPERSONであると規定している。

```
*DO*   LEVEL      3
       GEN-MOPS   (ACTION)
       SPECS      (ACCUSE HARM VIOLENT-ACT ...)
       TEMPLATE   (*DO* ACTOR PERSON)
       REC-TYPE   CONCEPT
```

図4. *DO*の概念表現

THINGは図5に示すように名詞の集合である。この概念構造は分析対象をテロリズムに限定しているため、比較的単純化されている。

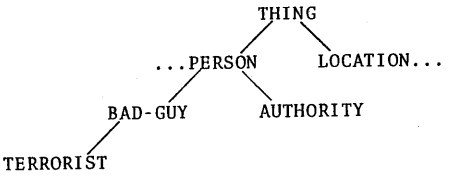

図5 THINGの構造

日本語辞書では、動詞や名詞は基本形が項目として挙げられていて、品詞名や概念表現における対応語が表示されている。動詞の場合は語尾活用の型が指定され、修飾語に付随する助詞が特殊な場合には指定される。名詞の項目では、特殊な複数形の場合にその指定がある。

4. 日本語文章の生成

4-a. 生成の手順

パーザーが分析して得た概念表現にいくつかの事象が含まれる場合には、この生成システムは、それらの事象を発生した時間順序に並べることによって日本語生成文の全体構造を与える。

英語等では、注目する事象がまず主節として生成され、修飾節は後に置かれるのが普通である。

ところが日本語では、修飾節は主として主節の前に置かれる。そして、各事象は発生時刻の順に並べて生成すると、自然な日本語の文章になる場合が多い。そこで本システムでは分析の結果得られた事象を、MOPsや因果関係を用いて時間順序に並べ，適切な接続詞を付加する。次に各事象を順番にDeapth-firstで生成する。

4-b. MOPsを用いた事象の順序付け

図6に警察の逮捕のMOPを示す。SCENE1は加罪の記述であり、それがSCENE2の捜査を引き起こす。その結果としてSCENE3の逮捕が行なわれる。このMOPを用いれば、それぞれの概念がCRIME、POLICE-SEARCH，ARRESTであるような事象は、ACTORやOBJECT等をチェックした後で時間順序付けが行なわれる。

少し複雑な場合として、図7のストーリー10があり、次の事象が与えられる。

KID3（ゲリラが実業家を誘拐した）
DEM1（ゲリラが金を要求した）

パーザーによって次の三つの関係を得る。

(M-E6 KID3 SCENE1)
(M-E6 GET1 SCENE2)
(GET1 DEM1 SCENE1)

M-E6の概念は恐喝のMOPで、SCENE1で人質を手に入れて，SCENE2で金を得ようとする。GET1の概念の中に、SCENE1として（金を）要求する事象があるので、KID3 ⟶ GET1を置きかえて、KID3 ⟶ DEM1を得る。SCENEの間に特別な因果関係がないので、接続詞として"soshite"を選び、出力日本文の構造を

KID3. "soshite" DEM1.

として図7を得る。

4-C. 意味的関係を用いた事象の順序付け.

パーザーは、MOPsの枠組みで表現できない事象には意味的関係を与える。二つの事象をE1，E2として、

LEAD-TO ：E1の後E2が生じた
RESULT ：E1の結果，E2が直接生じた
DURING ：E1の時，E2が生じた.

```
M-POLICE-CAPTURE

ACTOR     AUTHORITY
OBJECT    BAD-GUY
SCENE1    (CRIME ((ACTOR . OBJECT) (IR . SCENE2)))
SCENE2    (POLICE-SEARCH ((ACTOR . ACTOR) (OBJECT . OBJECT)
                          (LEAD-TO . SCENE3)))
SCENE3    (ARREST ((ACTOR . ACTOR) (OBJECT . OBJECT)))
GEN-MOPS  (M-PUNISH)
```

図6　警察の逮捕のMOP

A Spanish industrialist Salvador Beneitez Nieto was kidnapped and then assassinated by suspected lef ist guerrillas according to Guatamalen police. Nieto was the owner of the Panificadora Europa and had been kidnapped by an unknown group on November 18 who demanded a large ransom from Nieto's family.

terososhiki ni zokusu saha shugi no gerira to omowareru terorisutotachi wa, juichigatsu 18 nichi ni youroppa ni sumu supein kokuseki no jitsugyouka no sarubadoru-benaitesu-nieto o yukaishita. soshite, sono geriratachi wa, takusanno kane o youkyushita.

図7. ストーリー10の入力の英訳（上）と日本語出力（下）

GOAL：E1を目標としてE2が生じた.
PRECONDITIONS：E1はE2が生じるための前提条件である。

これらの関係では事象の間の時間順序が定まっており、パーザーの分析結果から一連の事象列を得ることができる。

しかし、それだけでは順序が定まらない場合がある。図8のストーリー3で、

KIL2 (警官が犯罪者を殺した)
TRA0 (警官が犯罪者を護送した)
ESC1 (犯罪者が逃げようとした).

この時、パーザーが与えた結果は

TRA0 ⟶ KIL2 , ESC1 ⟶ KIL2

であり、TRA0とESC1の時間関係は不明である。従って概念表現の中で推論しなければならない。

TRA0の概念から、ACTORがOBJECTをCONTROLすることが導かれ、一方、ESC1の概念からDISABLE-CONTROLが導かれるので、CONTROLの状態からDISABLE-CONTROLという行動が生じたと推論される。ここで、TRA0でのACTORとOBJECTがESC1のOBJECTとACTORにそれぞれ対応することがチェックされる。この結果、図8に示すように、

TRA0 "toki," ESC1 "node," KIL2.

という全体構造がまる。

4-d 事象の生成

各事象は次の三つの関数によって順番に生成される。

(PICK-SUBJECT)
(PICK-WORD-ORDER-SLOTS)
(PICK-VERB)

(PICK-SUBJECT)は、その事象のACTORスロットを取り出し、修飾句を含めて生成する。そのスロットがNILの時は受身形になる。主格を表わす助詞"は"と"が"を厳密に区別して使用するのは大変難しい。[4] 本システムでは簡単に、主節の主格に"は"を使い、従属節には"が"を使用した。

主節と従属節の主語が同一の場合には、主節の主語を省略した。たとえば図9のストーリー18で、全体の構造は

CUOO "tameni," VIOO. "soshite," GRPO.

で与えられ、CUOOとVIOOの主語が同じOBJOのため、VIOOの主語は省略された。

(PICK-WORD-ORDER-SLOTS)は動詞を修飾する語句を生成する。修飾句の順序と助詞は次の様に決め、

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

動詞によって異なる場合には、日本語辞書で特に指定した。たとえば、図10で、辞書の中の"boukousuru"は、

A convict Roger Fidel Marales Gonzalaz was killed by the patrolman who was driving him here from Tierra Azul.
The convict tried to escape by jumping from the vehicle, but when he did the patrolman fatally shot him, according to a responsible police source.

keikan ga, tiera-azuru toiu machi kara hanzaisha no roja-fideru-mararesu-gonzaresu o gosoushita toki, sono hanzaisha ga, nigeyoutoshita node, keikan wa, sono hanzaisha o koroshita.

図8 ストーリー3の入力の英訳(上)と日本語出力(下).

200 civil guards apparently following the orders of lieutenant colonel Tejero occupied the session of congress today and started to shoot at the deputies and ministers who were in the chambers in a coup attempt.

chuusa no hikiiru soshiki ga, kuudeta-o-okosu tameni, gikai no honkaigijou o senkyoshita. soshite, sono chuusa wa, sono honkaigijou de giintachi o juugekishita.

図9 ストーリー18の入力の英訳(上)と日本語出力(下)

WORD-ORDER の項で特に (OBJECT ni) と指定している。

```
(boukousuru    INF-EB
               ACTOR        boukousuru
               *INFL        SAHEN
               WORD-ORDER (ACTOR (OBJECT ni))
               DEF          HARM)
```

図10　日本語辞書における表現例

4-C. 名詞句の生成

名詞句を始めて生成するときは

(PICK-NOUN-MODIFIERS)

(PICK-NOUN)

を使用し、2回目以降の生成では、

(PICK-PRONOUN)

を用いる。

(PICK-NOUN-MODIFIERS)によって生成する修飾句の順序は次のように指定し、MONTH DAY RESIDENCE ···· TYPE NUMBER AGE NAME STATUS

この他の修飾句は節の中で表現された順番に従う。

(PICK-NOUN)は、その節の概念を生成する。たとえばストーリー1のLOC0では san-pedoro-perurapan toiu machi が生成される(図11)

```
(LOC0      NAME      (SAN PEDRO PERULAPAN)
           CONCEPT   CITY)
```

図11　(PICK-NOUN)のための例

また、ストーリー9のHUM23では、bokujou o shoyusuru isha no gaburiera-forunegura が生成され(図12)、概念のPERSONは訳出しない方が自然である。このようにNAMEと,TYPE, RANK 等が性質リストにある場合は、概念を生成しない。

```
(HUM23     POSSESSION   RANCH
           PROFESSION   DOCTOR
           NAME         (GABRIEL FORMEGRA)
           CONCEPT      PERSON)
```

図12　概念を訳出しない例

(PICK-PRONOUN)は同じ名詞句を2回目以降に生成する時に起動される。本システムでは、指示代名詞の使用を避け、(sono isha)のように"sono"を付して繰り返して表現する。ただし、NAMEの項があるときは、TYPEやPROFESSION等の項に"sono"を付けて生成する。たとえば図13で、HUM0は1回目の生成で hanzaisha no fideru-gonzaresu となるが、2回目以降は sono hanzaisha となる。

```
(HUM0      TYPE         CRIMINAL
           NAME         FIDEL-GONZALEZ
           CONCEPT      BAD-GUY)
```

図13　(PICK-PRONOUN)のための例

5. 検討

この日本語生成システムは、15の新聞記事に適用した結果、細かい点は除いて、ほぼ良好な日本語文章を生成したと考えられる。MOPsは、日本語の生成に対して変更する必要なく使用できたので、特定の言語に依存しないことが期待される。

入力文章は、テロリズム関係に限定しており、また、日本語辞書も必要な項目だけを作成してある。入力文章の範囲を拡大すれば、色々な訳語による使い分けの必要性が増してくるが、概念やMOPsを積極的に利用して訳語を選択する方法が考えられる。

本資料は米国で作成した日本語生成システムの要約であるが、電総研で、このシステムを改良、拡大していく予定である。また、日本語パーザーの開発も興味深い。

6. 謝辞

本システムを作成するにあたり、Schank教授を始めイェール大学の関係諸氏に多大な協力を得たことに感謝する。また、多くの励ましと有益な討論を戴いた当所推論機構研究室田中室長に感謝する。

7. 文献

(1) Schank, R.C., "Conceptual Information Processing", North-Holland (1975)

(2) Schank, R.C. "Reminding and Memory Organization. An Introduction to MOPs", Res. Rep. #170, Comp. Science Dept. Yale Unive.

(3) Ishizaki, S., Lytinen, S.V., Littleboy, D., "Generation of Japanese Sentences from Conceptual Representation — Inference Using MOPs —", Res. Rep., Comp. Science Dept. Yale Unive. (To be published).

(4) 田中他, 「自然言語処理技術と言語理論」電総研調査報告205号(1981)

(5) Lytinen, S.V. and Schank, R.C., "Presentation and Translation", Res. Rep. #234, Comp. Science Dept., Yale Unive.

8. 日本語文章生成例

図14にストーリー1の概念表現を示した. 図15と16には、それぞれ, ストーリー1とストーリー9の入力の英訳と日本語出力を示した。

```
M-M0 =
 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 UND1 =
                     CONCEPT UNDESIRABLE-ASSISTANCE
                     JUDGER  HUM11
                     OBJECT  OBJ5 =
                               CONCEPT GOVERNMENT
                     ACTOR   HUM6
           ACTOR   HUM11
           IR-FROM UND1
 SCENE4  EXE0 =
           CONCEPT EXECUTE
           ACTOR   HUM11
           PLACE   LOC0 =
                     CONCEPT CITY
                     NAME    SAN PEDRO PERULAPAN
           OBJECT  HUM6
           IR-FROM UND1
 SCENE1  UND1
 SCENE3  TRY0 =
           CONCEPT TRY
           OBJECT  HUM6
           ACTOR   HUM11
```

図14 ストーリー1の概念表現

At least 60 peasants were executed by a firing squid of men wearing olive-colored uniforms in San Pedro Perulapan, about 25 kilometers east of San Salvador, authorities there said.
According to the same sources, the victims were tried and then executed in the town plaza by guerrillas who accused them of collaborating with the government.

sukunakutomo 60 nin no noumintachi ga, seifu ni kyoryokushita node, terososhiki ni zokusu oribuiro no fuku o kita geriratachi wa, sono noumintachi o saiban-ni-kaketa. soshite, sono geriratachi wa, san-pedoro-perurapan toiu machi de sono noumintachi o shokeishita.

図15 ストーリー1の入力の英訳(上)と日本語出力(下)

Members of a guerrilla group, Ejercito popular de Liberacion, killed seven people and injured five others during an assault Saturday on a ranch located between Antioquia and Cordoba. According to sources 15 well-armed men tried to kidnap Dr. Gabrial Formegra, owner of the ranch la Florista but were confronted by ranch employees.

ejerushito-popyura-de-riberashion toiu terososhiki ni zokusu buki o motta 15 nin no terorisutotachi ga, bokujou o shoyusuru isha no gaburieru-forumegura o yukaishiyoutoshite, doyoubi ni sono isha no iru bokujou o senkyoshita. sono toki, sono terorisutotachi wa, 7 nin no hitotachi o koroshita.

図16 ストーリー9の入力の英訳(上)と日本語出力(下)