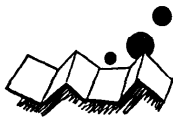


## 解説



# 機械翻訳における中間表現の変換 と文の合成手法†

西田 富士夫<sup>††</sup> 高松 忍<sup>††</sup>

### 1. はじめに

自然言語は同じ事実や関係を述べるにも各言語により独自の構造をもっており、一段階で変換することは難しい。それでソース言語Sからターゲット言語Tへ変換する場合、語順などの文法的構造の特徴をとり去った中間表現に変換してからT言語に翻訳する中間言語方式が現在では多く用いられている<sup>8)~11), 14), 21)</sup>。すなわちS言語を構文的、意味的に解析してS言語の中間表現をつくり、これよりT言語への語彙変換および構造変換を行ってT言語にふさわしい中間表現に変換し、これよりターゲット文を合成する。

中間言語方式において解析、変換、合成の処理を上述のように分離して行う方式をトランスファ方式と呼び、多言語間翻訳に有用で、かつ、体系的な方式である。この方式では単語辞書はS辞書、S-T変換辞書、T辞書を本来別々に備えることになるが、実用上はこれらの辞書から各ターゲット言語ごとに1個の辞書を合成し、2国語間翻訳のようにこれを利用する方が効率が高いと考えられる。またこのような合成辞書により2.2節のイディオム処理のように、S言語の解析自身もT言語を特定することにより効率化することができ、このため中間言語方式にはトランスファ方式のいろいろな変形が存在する。

中間表現は文中の中心動詞や中心名詞と他の語句間の修飾限定関係、あるいは意味的な依存関係を表すものが大部分である。基礎となる体系や記法により中間表現（または内部表現）にはいろいろなものが提案されているが、ここでは格構造を用いるものについて述べる<sup>16)</sup>。これは文の中の各自立語の前に語の意味的役割を示す格ラベルを前置した表現である。そして意味

的役割を表層的に表す語順や語の活用形や前置詞後置詞パターンなどの表現をとり除き、語彙は別として意味的表現を可及的に言語に独立に表そうとしたものである。言語間の距離が大きい日本語と印欧語間の翻訳には必要な中間表現と考えられる。

文の中間表現は命題部 (proposition) とモーダリティ部 (modality) とからなるが<sup>9)</sup>、いま、ソース言語Sの文の命題部の中間表現を

$$(K_1 - C_1 : t_1, \dots, K_n - C_n : t_n) \quad (1)$$

とする。ここに  $K_i - C_i$  は言語Sにおいて、格ラベルとその格に入るべき語  $w_i$  のカテゴリとの対、 $t_i$  はタームと呼ばれ、原則として  $w_i$  の原形で表し、修飾子をもつときには

$$t(K_{i1} - C_{i1} : t_{i1}, \dots, K_{im} - C_{im} : t_{im}) \quad (2)$$

で表す。 $t$ の右の部分は $t$ の修飾子で、 $t_{i1}, \dots, t_{im}$ の一方は $t$ に等しく、これを'\*'で表す。上式は $t$ が他のタームと括弧内のような関係をもつタームであることを表す。ただし、カテゴリ $C_i$ が自明のときなどには今後簡単のためにこれを省略する。

次に(1)式からこれに対応するターゲット言語Tの中間表現

$$(K_1' - C_1' : t_1', \dots, K_n' - C_n' : t_n') \quad (3)$$

を構成し、これよりTの書きかえ規則などを用いて、Sから(1)式を求める過程と逆の過程によりターゲット文を構成する。

この解説で取り扱うのはまず $t_i$ から $t_i'$ への変換、すなわち語または句レベルの変換で第2節に述べる。第2は文の格構造の変換で第3節に述べる。

さて非生物主語の問題として知られているように<sup>1)</sup>、(1)式において $t_i$ を $t_i'$ に置き換えたものは、必ずしもTにおいて慣用的で理解しやすい表現に置き換えられない。また日英翻訳における冠詞のように(3)式を推測により補填しなくてはならない場合がある。

さらに格構造の変換を必要とする要因がある。それはソース文に含まれる情報は、可及的に同じ順序で

† Intermediate Expression Transfer and Sentence Generation in Machine Translation by Fujio NISHIDA and Shinobu TAKAMATSU (Department of Electrical Engineering, Faculty of Engineering, University of Osaka Prefecture).

†† 大阪府立大学工学部電気工学教室

ターゲット文に再生されることが望ましいということである<sup>2)</sup>。また、これに関連して文の主題となるタームが翻訳過程を通して継承されることが望ましい。

文の主題の定義は現在のところ充分明らかにされていないが、要は文の題目となっている項目や場面などのラベルであり、多くの場合、先行文などで与えられる既知情報の項目である。一方、焦点とは主題またはこれに関連する主要項目の詳細な記述や内容であり、主題に対し新しい情報を付加する<sup>3), 15)</sup>。

英語の場合、主題は主語や場所格などの副詞句として多くの場合文の先頭に与えられ、焦点は文の後部に関係節などの句や節として与えられる。

日本語の場合、文頭に出ている手段格や場所格などの連用修飾句も、見出し的なものであれば主題と考えられ、主題と焦点の位置関係は英語と同様な関係にあることが多い。

ところが日本語と英語とでは文法が異なり格構造を変換しなければ、ソース文の主題の語がターゲット文の主題の位置にくるとは限らない。特に日英翻訳においては英語の主語の位置が半ば固定されているため、直訳では主題を継承した英文がえられないことがある。

このため格構造変換を行い、ソース文の情報の出現順序や主題をになうタームをできるだけターゲット文に継承することが望ましい。

最後に必要な変換は(3)式の間中表現からターゲット文への変換で第4節に述べる。言語によっては英文のようにきびしい文法規則をもつものや、日本語のように省略や名詞句の動詞化などを行わないと読みにくいものがある。

以上について筆者らが英日、日英機械翻訳について行ったこと<sup>16)-20)</sup>を中心に述べる。

## 2. 語句の間中表現の変換

### 2.1 訳語の選定

どの言語でも、これを用いる民族の慣習や文化に従って語の流用と意味の転移が行われ、多くの多品詞語や多義語が存在する。また言語間における訳語の関係は一对多でかつ、非対称である。したがって逐語の処理では翻訳は不可能に近い。

しかしどの自然言語も同じ実世界の物事の、ほとんど同じ認識能力をもつ人間による記述を通して相互に結びついている。したがってそれぞれの専門分野の用語や、物事の各場面ごとに作っておいた一般のフレ

ームとくい違いを生じないように、ソース言語で書かれた文を文脈に従って解釈することにより、おきかえるべきターゲット言語の文をほぼ正しく作ることができるであろう。

この中でもっとも顕著なものが科学技術の各分野の専門用語の処理や対訳であろう。専門用語が指示するものは、複雑な関係やデバイスであっても、一般の用語とは異なった一意的な意味をもつものであり、このため指示しているものの処理や翻訳が容易なものとなっている。しかし専門用語の中には‘operator’と‘運転手(機械)’、‘交換手(電話)’、‘手術者(医学)’、‘演算子(数学)’、…などのように分野ごとに一対一に対応しないものがあるので、文脈から分野を同定するか、前もって分野名が与えられていることが必要である。

一般に用いられる基本的な語句も、なんらかの文脈を参照することにより正しい訳語を選定できることが多い。その一つは述語の格構造と格に入る語のカテゴリ制限を用いるもので、多くのシステムで用いられている。

筆者らは文に現れる主述語の意味カテゴリに基づき、文の記述内容を約20個のフレームに分類した。そしてフレームごとに主述語にかかる語の意味的役割と意味的カテゴリならびにその表層構造を規定する表を、英語や日本語について試作した。このような表を用いてパーズングを行えば、文の中の各語の意味的カテゴリや役割を文脈から能率的に同定することができる。これより適当な訳語を選定することができる。以下に二、三の例を示す。

#### 例1

- (1) a. We stopped to talk.

(AGent- PRED- PURPOSE-  
THINGS Physical ACTION  
TRANSfer)

われわれは 話すために 立ち止まった。

- b. He stopped talking.

(AG- PRED- OBJ-  
HUMAN ENDACT ACT)

彼は 話すのを やめた。

- (2) a. She put some water

(AG- PRED- OBJ-  
THINGS PTRANS PHYSOBJ)

in the jug.

(GOal-  
PHYSOBJ)

彼女は 水を 水差しに 入れた。

- b. He put his feelings

(AG- PRED- OBJ-  
(HUM Mental TRANS MENTOBJ  
in words.  
INSTRument-) MENTOBJ)

彼は 自分の感情を 言葉で 表現した.

上述の方法は言語を統一的な立場からとらえ、辞書を小型化するのにかなり役立つものと思われる。しかし、指定された格のカテゴリに包含される大部分の語句には成立しても、成立しない語句も存在する。すなわち‘碁をうつ’や‘将棋をさす’といった限られた対象にのみ用いられるコロケーションの問題がある。このような場合には、より細分化したカテゴリを用いるか、後述のイディオムのな処理により適当な訳語を選定することが必要である。

田中らは基本動詞‘take’や‘make’などを例にとり必要に応じて主語、目的語、前置詞句など3個程度の語句を参照して、約20通りの訳し分けを行う実験をしている<sup>6),7)</sup>。

訳語選定において重要なのは、上述のような選定を行う基本的な体系に加えて、多義性が少なくソース言語との対応関係が明らかなイディオムやこれに準ずる語句を見出しとして積極的に辞書に導入することである。イディオムとは‘look for’とか‘首をひねる’‘知恵をしぼる’のように、通常、構成要素からイディオム全体の意味を推定することは困難であるが、全体としてはそれだけで意味がほぼ確定しているものを指している。したがって、これらを最長マッチなどにより速やかに同定できるように、単語辞書のイディオム欄に見出しをつけて記載することが望ましい。

一方、‘結果として～になる’⇒‘result in~’や、‘～で役割を果たす’⇒‘play a part in~’や‘～に損害を与える’⇒‘cause damage to~’などは少なくともソース言語では構成単語から全体の意味は明らかである。しかし対訳語をみつける場合、前述のサブフレームによるカテゴリ細分の方法では、場合分けや選定の規準が難しくまた非能率である。このような場合にはイディオムと同じく S-T 変換辞書やS言語の単語辞書(T言語が特定されているとき)のイディオム欄に見出しをつけ、構成単語の一般情報より優先してイディオム情報を用いることが、S-T 変換やS言語の解析の効率の上から望ましい。なお、これらのイディオムや準イディオムは、外部の語句との結合のため、次節に述べるような格構造変換を行わねばならない場合がある。

2.2 語句の格構造の変換

他国語間、たとえば日本語と英語とでは、慣用的に用いられる表現の単位が必ずしも同じでなく、両言語間で中間表現をみだすタームが部分的に対応して存在しない場合がある。このようなときには格構造の変換を導入することが必要である。たとえば‘小包の重さを測る’の中間表現は日本語では

(PRED: 測る, OBJ: 重さ (OBJ: \*,  
POSSessor: 小包)) (4)

となる。しかし英語では‘重さ’を含む述語句‘重さを測る’(weigh)が一つの他動詞(VP6A)として存在し、修飾子のPOSS格‘小包’が主節のOBJ格にくり上がって‘weigh a package’となり、その中間表現

(PRED: weigh, OBJ: package (OBJ: \*,  
DETerminer: a)) (5)

が(4)式に対応する。

これらの語は、辞書の各単語の熟語欄に表-1のように記載する。なお上述のような一つの動詞語に対応する述語句において、述語句の構成要素の名詞語が連体修飾されることが多い。このため修飾子を変数化して次のような一般的な変換公式を用意し、辞書には参照すべき変換公式の番号を記載しておけば格構造の変換を能率的に行うことができる。

TR 1: (PRED: *t*<sub>0</sub>, OBJ: *t* (OBJ: \*, POSS: *n*))  
⇨(PRED: *t'*-*t*<sub>0</sub>', OBJ: *n'*) (6)

TR 2: (PRED: *t*<sub>0</sub>, OBJ: *t* (OBJ: \*, PRED: *adj*))  
⇨(PRED: *t'*-*t*<sub>0</sub>', MANNer: *adv*) (7)

ただし *t'*-*t*<sub>0</sub>'などは *t* と *t*<sub>0</sub> を連ねた熟語に対する訳語を表し、*n*, *adj*, *adv* は名詞、形容詞、副詞の語を表す。このようにして次の訳文がえられる。

表-1 単語辞書の熟語欄

見出し	日本語中間表現	訳語
測る	...	...
重さを—	OBJ-PQUANT: 重さ-を	PRED-THINKACT: weigh, VP 6 A
nの重さを—	OBJ-PQUANT: 重さ-を (OBJ: *, POSS-PHYSOBJ: n-の)	refer to TR 1
降る	...	...
雨が—	OBJ-NATURE: 雨-が	PRED-NATURAL phenomenon. rain, VP 2 A OBJ-CIRCUM: it
adj-雨が—	OBJ-NATURE: 雨-が (OBJ: *, PRED-ATTR: adj)	refer to TR 2

## 例2

(1) 今日は 激しい 雨が降っている。

⇔It is raining heavily today.

(2) その増幅器は 第2検波器のあとに続く。

⇔The amplifier follows the second detector.

## 3. 文節の中間表現の変換

## 3.1 単文レベルの変換

日英両言語の単文レベルでの主な違いは、英語では行為者格や対象格などのタームが広い意味カテゴリーにわたって許されているのに対し、日本語ではかなり狭い範囲に制限されていることである。このため主述語の働き方や、主述語にかかるタームの格が異なってくる。これらはまた所有表現と存在表現あるいは属性表現との差異や、動作表現と状態変化表現との差異などとして現れる。

## (1) 所有表現と存在・属性表現との間の変換

物や属性の所有表現と、物や属性の存在表現とは意味的に等価で、互いに他に変換できることが多い。元来、英語は所有表現向きの言語、日本語は存在表現向きの言語と呼ばれている。英語でも 'There is a book on the desk' にみられるように、互いに永続的な構成関係や所有関係にない物の、場所や物における存在には存在表現が用いられるが、それ以外の物事の存在には広く所有表現が用いられる。これに対し日本語では、手にもてる物や車や家などの所有権の所有などを除いて存在表現を用いることが多い。

次に英語の所有表現と日本語の存在表現との変換規則を示す。

(PRED-POSSession: have, OBJ-C<sub>1</sub>: t<sub>1</sub>,  
PARTIC-C<sub>2</sub>: t<sub>2</sub>) ⇔ (PRED-EXISTence: ある,  
いる, LOC-C<sub>1</sub>: t<sub>1</sub>', OBJ-C<sub>2</sub>: t<sub>2</sub>)  
禁止条件 P (8)

ただし、二重下線は1. 節で述べた主題の項を表す。また、C<sub>1</sub> は物や場所、C<sub>2</sub> は物のカテゴリーのラベルで、t<sub>1</sub>, t<sub>2</sub>, t<sub>1</sub>', t<sub>2</sub>' がこれらのカテゴリーをみたし、かつ、禁止条件をみたさない場合には、矢印の変換を行うべきことを示す。禁止条件Pは、この場合、ソース文が 'a HUMAN has a OBJECT with his HAND' や 'a HUMAN has FOOD' などに相当する場合で、これらの文に含まれるタームの主要な格とカテゴリ対の組などで表す。

次に英日間の属性所有表現の変換規則を示す。英語の属性所有表現

(PRED-POSS: have, OBJ-POBJ: t<sub>1</sub>,  
PARTIC-C: t<sub>2</sub>(PRED-ATTRibute: t<sub>3</sub>,  
OBJ: \*)) (9 a)

と、日本語の属性表現

(PRED-ATTR: t<sub>3</sub>', LOC-POBJ: t<sub>1</sub>',  
OBJ-C: t<sub>2</sub>') (9 b)

または

(PRED-ATTR: t<sub>3</sub>', OBJ-C: t<sub>2</sub>' (OBJ: \*,  
POSSessor-POBJ: t<sub>1</sub>')) (9 c)

は、原則として互いに一方から他方に中間表現を変換する。ただし、C は {Phys QUANT, BODYPART} なるカテゴリーラベルの集合である。

## 例3

(1) This system has two kinds of sensors.  
(PRED-POSS: have, OBJ-POBJ: system(OBJ:  
\*, DET: this), PARTIC-POBJ: sensors(OBJ:  
\*, CHAR: kinds (OBJ: \*, NUM: two)))

⇔(PRED-EXIST: ある, LOC-POBJ: システム  
(OBJ: \*, DET: この), OBJ-POBJ: センサ  
(OBJ: \*, CHAR: 種類 (OBJ: \*, NUM: 2)))

このシステムには2種類のセンサがある。

(2) This magnet has high coercive force.  
(PRED-POSS: have, OBJ-POBJ: magnet.  
(OBJ: \*, DET: this), PARTIC-PQUANT:  
coercive-force (PRED-ATTR: high, OBJ-  
PQUANT: \*))

⇔(PRED-ATTR: 高い, LOC-POBJ: 磁石  
(OBJ: \*, DET: この), OBJ-PQUANT: 保磁力)

この磁石は保磁力が高い。

または(9 c)により、この磁石の保磁力は高い。

## (2) 動作表現と状態変化表現との間の変換

熱が気体の圧力を増やした' と '熱で気体の圧力が増えた' のように、動作表現と動作によってもたらされる状態変化表現とは一対一に対応する。

さて、非生物主語の問題としてよく知られているように、英語では事象や非生物のカテゴリーに属する語も AGent 格にすることができ、これらの語を AG 格とする動作表現を用いて簡潔な文型により事象を記述することが多い。一方、日本語では AG 格の語に人間や生物をえらぶことが多く、特に非生物の語が AG 格として OBJ 格の人間に動作を及ぼす表現を避けている。このため英語では AG 格にする語を日本語では原因格や道具格などとし、動作表現の代りに状態変化表現により事象を記述することが多い<sup>4)</sup>。動作表現

表-2 動作表現と状態変化表現との間の変換

$C_0$	$C_1$	$C_2$	$K_1$
PRODUCTION POSSTRANS	NLOBJ	POBJ	INSTR
PRODUCTION POSSTRANS	ACTION	POBJ	MEANS
ATTRTRANS	NLOBJ EVENT	POBJ	CAUSE
MTRANS PERCEPACT	MOBJ EVENT	THING	SOURCE
USE	PRODUCT	THING	LOC

と状態変化表現との間の主な変換を次式と表-2に示す。

$$\begin{aligned} & (\text{PRED-}C_0: t_0, \text{AG-}C_1: \underline{t_1}, \text{OBJ-}C_2: t_2) \\ & \Rightarrow (\text{PRED-}C_0: t_0', \text{K}_1\text{-}C_1: \underline{t_1'}, \text{OBJ-}C_2: t_2') \end{aligned} \quad (10)$$

上式において  $K_1$  は動作表現における AG 格に対応する状態変化表現の格で、動作表現における主述語、AG 格、OBJ 格のカテゴリ  $C_0, C_1, C_2$  により表-2 のように定める。また動作述語  $t_0$  と状態変化述語  $t_0'$  とは次のように対応させる。

(a) ‘動かす’ と ‘動く’ のように、動作表現では他動詞  $t_0$ 、状態変化表現では自動詞  $t_0'$  をとる。

(b) 対応する自動詞  $t_0'$  がなければ、 $t_0$  の受身形、またはこれと等価な語、たとえば ‘示される’ に対してこれと等価な ‘わかる’ を用いる。

同様な変換に使役述語を用いた動作表現と、モーダリティ (modality) の格の語を付加した状態変化表現がある。

$$\begin{aligned} & (\text{PRED-}C: t_0, \text{AG-THING}: \underline{t_1}, \\ & \text{OBJ-ACT}: (\text{PRED-ACT}: t_2, \text{OBJ-POBJ}: t_3)) \\ & \Rightarrow (\text{PRED-ACT}: t_2', \text{MODALITY}: m, \\ & \text{OBJ-POBJ}: t_3', \text{K-THING}: \underline{t_1'}) \end{aligned} \quad (11)$$

ただし、使役述語  $t_0$  のカテゴリ  $C$  が CAUSE や ENABLE であるとき、 $K$  は INSTR や CAUSE の格をとり、 $m$  は可能 (することができる) や  $\phi$  (付加語不要) の値をとるものとする。

例 4

(1) Comparing these two signals gives a measure of the force.

$$\begin{aligned} & (\text{PRED-POSSTRANS}: \text{give}, \text{AG-ACT}: \\ & (\text{PRED-THINKACT}: \text{compare}, \text{OBJ-MOBJ}: \\ & \text{signals} (\text{OBJ}: *, \text{DET}: \text{these}, \text{NUM}: \text{two})), \\ & \text{OBJ-PQUANT}: \text{measure} (\text{OBJ}: *, \text{DET}: \text{a}, \end{aligned}$$

$$\begin{aligned} & \text{POSS-PQUANT}: \text{force} (\text{OBJ}: *, \text{DET}: \text{the})) \\ & \Leftrightarrow (\text{PRED-POSSTRANS}: \text{与えられる}, \\ & \text{MEANS-ACT}: (\text{PRED-THINKACT}: \text{比較す} \\ & \text{る}, \text{OBJ-MOBJ}: \text{信号} (\text{OBJ}: *, \text{DET}: \text{これ} \\ & \text{ら}, \text{NUM}: 2 \text{個})), \text{OBJ-PQUANT}: \text{目安} (\text{OBJ}: \\ & *, \text{POSS-PQUANT}: \text{力} (\text{OBJ}: *, \text{DET}: \text{そ} \\ & \text{の}))) \end{aligned}$$

これらの2個の信号を比較することによりその力の目安が与えられる。

(3) 受動表現と能動表現との間の変換

動作表現で AG 格の語を明らかにする必要がない場合や、OBJ 格が主題である場合、英語では前述のような格構造変換を行うか、または受動形で表現する。他方、日本語では AG 格の語が欠けても、一般には能動形をとるが、科学技術文の英日翻訳では受動表現のままに他動無意志的に訳出しても自然な場合が多い。

しかし、‘することができる’ や ‘すべきである’ などの様相を表す助動詞を伴う場合や、‘喜ぶ’ ‘知る’ ‘見る’ などの感情や知覚などを表す動詞に対して受動形を用いることはほとんどない。このような場合にはモーダリティの態格を能動で置きかえる。さらに受動形他動詞  $v_i$  に対し、自動詞またはその相当語  $v_i$  が存在する場合には  $v_i$  を  $v_i$  で置きかえる。

例 5

Since the collector voltage is known, the base current can be found.

⇒コレクタ電圧がわかるので、ベース電流を見出すことができる。

3.2 複文の場合

翻訳において単文、複文、重文などは通常それぞれターゲット語の同種の文に変換される。しかし場合によっては複文を重文などに変換することが望ましいことがある。ここでは、このような場合の中間表現の変換について述べる。

(1) 関係節

関係節の翻訳で問題になるのは、関係節が制限的に使われている英日翻訳の場合である。関係節が長く複雑なとき、これを制限的に連体修飾的に訳せば、複雑詳細な関係節の情報が見出し的な主節の被修飾項に先行し、わかりにくい日本語が生成されることになる。一方、英文の情報の出現順序に従って継続的に訳すのは原則として好ましいが、場合によっては次のような不適切な訳文を生成することがある。

(a) He has a daughter who lives in Vienna.

彼には一人の娘がおり、ウィーンに住んでいる。

(b) Madame Curie is a woman who discovered radium.

キュリー夫人は一人の婦人で、ラジウムを発見した。

これらはいずれも制限的に用いられている関係節を継続的にとり扱うことにより生ずる問題である。(a)は‘彼には娘はただ一人しかいない’という誤りを招来する特殊な場合である。(b)は‘a woman’を‘a great scientist’でおきかえたときには継続的に訳しても不自然に感ぜられないことからわかるように、(b)の和訳において‘婦人’が主語に対して新しい情報をもっていないことから生ずる不自然さと考えられる。

そこで制限的に用いられている関係節を連体修飾的に翻訳する一つの規準として

(a) 関係節の長さが短いとき

(b) 主節の述部が主節の主部に対して情報を含まず、主節単独では不自然なとき

とし、それ以外は継続的にターゲット文を生成することが考えられる。

さて、継続的に翻訳する場合、主節と関係節とで主題の違いが目立つことがあるが、次のようにすれば比較的滑らかに関係節を主節につなぐことができる。

いま関係節を含む文の中間表現を

$$(PRED: t_0, K_1: t_1, \dots, K_n: t(R)) \quad (12)$$

とする。ここに  $R$  は関係節の中間表現で、 $t$  はその先行詞である。上式から継続的な表現をうるために(12)式を次のような中間表現に変換する。

$$(PRED: t_0', K_1: t_1', \dots, K_n: t'), \\ \text{COMMENT}:(R') \quad (13)$$

ただし、 $(R')$  は(12)式の先行詞  $t$  を主題化した関係節  $R$  の中間表現である。多くの場合、 $R'$  は  $R$  の中で  $t$  に主題化標識をつけるだけでよい。そして  $R'$  から表層表現に展開するとき、 $t$  の前置や係助詞‘は’の付加を行って  $t$  を  $R$  の表層文において主題化する。しかし、 $t$  が AG 格をもつ  $R$  の OBJ 格のような場合には、 $R'$  を

$$(PRED: \text{である}, \text{OBJ: } \underline{\text{それ}}, \text{PARTIC: } \text{もの} \\ (R)) \quad (14)$$

のように構成し、 $t$ (=‘それ’)を主題化した方が自然な日本語がえられることが多い。

また(13)式の COM は重文において、先行する文に対する  $(R')$  の役割を示す文の役割名で、 $t$  の  $R'$  における格に応じて文と文との接続の形や接続詞を指定す

るものとする。通常は COM から  $(t_0'$  の連用形) + ‘,’ や、 $(t_0'$  の終止形) + ‘が’ + ‘,’ などを生成するが、先行の主節が長文であったり命令文であるときには、 $(t_0'$  の終止形) + ‘.’ を生成するものとする。

例6

(a) These transistors form a bias current source whose current is used to produce the level-shift voltage across the resistance.

これらのトランジスタにより一つのバイアス電流源が構成されるが、その電流はレベルシフト電圧をその抵抗の両端につくるために用いられる。

(b) I handed him an expensive book that she bought in Kyoto yesterday.

私は彼に一冊の高価な本を手渡したが、それは彼女が昨日京都で買ったものである。

(2) 結果格・目的格

関係節の場合と同様に、so~that 節や不定詞句による RESULT 格や PURPOSE 格の句節が長いときには、次のように中間表現を変換して、継続的な表現に展開した方がよい。

$$(PRED: t_0, \dots, \{PURPOSE, DEGR, \dots\}: S) \\ \rightarrow (PRED: t_0', \dots) \{RESULT, CONSEQ, \dots\}: (S') \quad (15)$$

ただし重文の役割名の結果格や CONSEQ からは、 $(t_0'$  の能動連用中止形) + ‘その結果’) や  $(t_0'$  の終止形) + ‘ので’ などを生成する。

例7

例6(a)の後半に(15)式の中間表現の変換規則を適用しその結果を表層表現に展開すれば

・・が構成されるが(を構成し)その電流を用いて(その結果)レベルシフト電圧をその抵抗の両端につくる。

なる訳文がえられる。

なお、日英翻訳においても、日本語がプロセス文の後に結果文などがつづく重文で与えられたときには、(13)式→(12)式や(15)式の逆方向の変換を行って複文の英文に翻訳する方が簡潔でよい場合が多い。

### 3.3 変換手順

ソース文に対して適切なターゲット文をうるために行う中間表現の各種の変換は、中間表現の主要部からタームの部分構造に向かって行う。まず文の文節に対応する中間表現の主要部から、主題の保持やターゲット文におけるタームのカテゴリと格の整合性を考慮して変換を行う。変換規則は

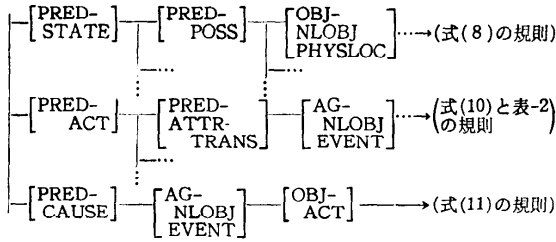


図-1 単文レベルの変換

(適用条件 変換操作) (16)  
 のプロダクション・ルールと同様な形で与える。

ここに適用条件は述語の意味カテゴリや、格ラベルとその格に入るべき語の意味カテゴリなどについて変換を行うべき条件を記述する。変換操作は適用条件が満たされたときに行う格ラベルの置き換えや、部分的な構造変換などの手順を示す。適用可能な変換規則の探索を能率的に行うために、適用条件をたとえば図-1のように階層化する。次に変換をつづけて行う例を示す。

例 8

The electric field becomes high enough to produce a kind of field emission that suddenly increases the number of carriers.

⇒

電界が十分に高くなり、その結果一種の電界放出を生じ、それにより突然キャリアの数が増加する。

この例では、(15)式の規則により、下線部①の不定詞句を継続的な表現に変換し、次に(12)式-(13)式の規則により、下線部②の関係節を継続的な表現に変換している。さらに、(10)式と表-2の規則により、下線部②の動作表現を状態変化表現に変換している。

4. 中間表現からの目的文の生成

$$V_1(K_1-C_1: t_1, \dots, K_n-C_n: t_n) \\
 == V_2(K_1-C_1: t_1) V_3(K_2-C_2: t_2, \dots, K_n-C_n: t_n) \quad (17)$$

のような書きかえ規則を用いて、入力文のターゲット言語での中間表現を格ラベルをキーとして展開することにより、目的文を生成する。ただし、 $V_1, V_2, V_3$  は非終端記号を表す。

筆者らのシステムでは、これらの書きかえ規則を表の形にまとめ、また同じ非終端記号の複数個の書きかえ規則に対して、選択などのため適用条件部を設けている。これらの適用条件は文や述部から主語や目的語

を生成するとき、それらの語の長さにより、主語を修飾する関係節を文末に後置したり、間接目的語と直接目的語の中、長くて複雑な方を後置するような選択規則である。また手順部を設け、書きかえ規則の適用に当たり、タームを表層の文法に適合するように変形する手順を記述している。

ここでは主として日英翻訳における英文生成について述べる。

4.1 述部の生成

英語の述部は、中間表現のモーダリティ部の情報すなわち、テンス、アスペクト、否定、態などの情報を参照して、命題部の PRED 格のタームを変形したり助動詞や副詞などを適当に付加して生成する。

また OBJ 格や CONT 格などのタームの表層表現はこれらがかかる動詞などのガバナの型により、動名詞句、不定詞句、形容詞句、名詞や that 節などの中の二、三の形に限定されることが多い。また格の表層標識である前置詞などもこれらがかかる動詞や中心名詞などの語により特殊なものをとることがある。このため、表-3 のように格ラベルに A. S. Hornby の動詞パターンを対応づけた表と、単語辞書を用いて、タームを指定された語句の形に変換し、必要に応じて適正な前置詞などを付加する。また、主述部が 'suppose' や 'suggest' などの VP 9 で、目的格が不定詞でなく that 節をとるようなときには、that 節の主語を復元したり、述部の時制を主節のそれに一致させる。

このように品詞変換の必要から、日英翻訳の単語のトランスファ辞書には、日本語の名詞語を主見出しとし、これと関連する動詞語、形容詞語、形容動詞語などの見出しを同じ枠内におくと便利である。そしてこれに対応する英語の語句や文法情報を接続する。

例 9

(1) その事実は重要と思われる。

The fact is thought important. <VP 25>

(2) 彼らはその現象をやがて見出すものと推測

表-3 格ラベルと動詞パターンの対応

動詞パターン名	OBJ 格	CONT 格
VP 3 B, VP 9	that 節	/
VP 6 C/D	動名詞句	
VP 17	名詞句	to 不定詞句
VP 22, VP 25		形容詞句
⋮	⋮	⋮

した。

They supposed that they would find the phenomena soon. <VP 9>

(1)では、'think'の動詞パターンがVP 25であるのでCONT格のタームを形容詞に変換する。(2)では、'suppose'の動詞パターンがVP 9であるのでOBJ格のタームをthat節に変換する。

#### 4.2 名詞句の生成

科学技術の術語には、名詞句として関係詞や前置詞を伴わない簡潔な複合名詞的表現が多く用いられる。ここではこれについて述べる。複合名詞は特殊なものもあるが、多くは次のような書きかえ規則により中間表現から生成することができる。

<NP( $p$ (PRED: \*, OBJ:  $t$ , INSIT or MANN or LOC:  $i$ ))>

= <{ADJ( $i$ )|<N( $i$ )}> <N( $t$ )> <VN( $p$ )> (18 a)

<NP( $t$ (PRED:  $p$ , OBJ: \*, INSTR:  $i$ ))>

= <N( $i$ )> <V-EN( $p$ )> <N( $t$ )> (18 b)

<NP( $i$ (PRED:  $p$ , OBJ:  $t$ , INSTR: \*))>

= <N( $t$ )> <{VN( $p$ )|<V-ING( $p$ )}> <N( $i$ )> (18 c)

ここに、 $p, t, i$ は動作、対象、道具などのタームを表わし、NP, N, ADJ, VN, V-EN, V-INGはそれぞれ、名詞句、名詞、形容詞、動名詞、動詞の過去分詞形、現在分詞形を表す。(18 a~c)はそれぞれ、動作、対象、道具を中心語とするタームを複合語に展開するとき、語順と各タームのとるべき語形(品詞)を示しており、これにより多くの専門用語を解析したり合成したりすることができる。

例 10

- |     |          |            |             |
|-----|----------|------------|-------------|
| (1) | デジタル     | LCR        | 測定          |
|     | MANN     | OBJ        | PRED        |
|     | ⇒digital | LCR        | measurement |
| (2) | パターン     | 変換         | 回路          |
|     | OBJ      | PRED       | INSTR       |
|     | ⇒pattern | converting | circuit     |

#### 5. おわりに

本稿では訳語の選定、中間表現の変換、目的文の生成について述べたが、英語名詞の単複、冠詞の選定などの問題は割愛した。文節レベルの中間表現の変換の問題は理論的にも実用的にも重要な問題であろう。中間表現の変換は構文的な構造変換をとめない、訳語の修正と異なり、機械出力の構文的な修正は労力がかかり大きいと予想される。したがって、中間表現の変換

の研究を進め、構文的な post editing はできるだけ避けるのが望ましい。この線に沿って、本稿では二、三の中間表現の変換や展開などについて述べたが、変換規則の適用条件を今後もっと精細化する必要があるであろう。変換規則の不適切な適用は直訳以上に理解し難い訳文を生成することもあると考えられる。

なお意識の必要から、文における主題と焦点との関係について、とりあえず現在なされている常識的な説明を述べたが、文と文との関係などとともに、将来、もっと詳しく研究される必要があるものと思われる。

#### 参考文献

- 1) 国広哲弥：構造的意味論，三省堂（1976）。
- 2) 別宮貞徳：翻訳を学ぶ，八潮出版（1976）。
- 3) R. クワーク他著，池上訳：現代英語文法，大学編，紀伊国屋書店（1977）。
- 4) 池上嘉彦：「する」と「なる」の言語学，大修館書店（1981）。
- 5) Simmons, R. F.: Semantic network: Their computation and use for understanding English sentences, In Schank and Colby (eds.), Computer models of thought and language, W. H. Freeman, pp. 63-113 (1973).
- 6) 田中，井佐原，安川：融合方式による機械翻訳システムの実験，自然言語処理研究 34-2 (1982)。
- 7) 田中穂積：基本動詞「MAKE」を含む文の日本語への訳し分け，自然言語処理研究 43-3 (1984)。
- 8) 内田，増山：機械翻訳における概念変換について，自然言語処理研究 39-3 (1983)。
- 9) 村木一至：概念依存図式と英語文生成，自然言語処理研究 44-3 (1984)。
- 10) 栗崎，平井，北橋：格構造に基づく日本語意味表現からの英文生成，自然言語処理研究 36-3 (1983)。
- 11) 橋本，鈴木，野垣内，櫛：文レベルでのトランスファのための簡易な意味表現についての検討，自然言語処理研究 43-2 (1984)。
- 12) 長尾 真編：日本語情報処理—第 10 章機械翻訳，pp. 288-315，電子通信学会（1984）。
- 13) Nishida, T. and Doshita, S.: An English-Japanese Machine Translation System based on Formal Semantics of Natural Language, COLING-82, pp. 277-282 (1982)。
- 14) 長尾，辻井，中村：日英科学技術文献の速報システムに関する研究の概要，情報管理，Vol. 28, No. 1, pp. 17-32 (1985)。
- 15) Derr, M. A. and Mckeown, K. R.: Using Focus to Generate Complex and Simple Sentences, COLING-84, pp. 319-326 (1984)。
- 16) Nishida, F., Takamatsu, S. and Kuroki, H.: English-Japanese Translation through Case-Structure Conversion, COLING-80, pp. 447-



- 454 (1980).
- 17) Nishida, F. and Takamatsu, S.: Japanese-English Translation through Internal Expressions, COLING-82, pp. 271-276 (1982).
- 18) 高松, 西田: 動詞パターンと格構造に基づく英日機械翻訳, 信学論 (D), Vol. J64-D, No. 9, pp. 815-822 (1981).
- 19) 西田, 高松, 谷: 英日機械翻訳における格構造変換と日本語の生成, 情報処理学会「自然言語処理技術」シンポジウム論文集, pp. 47-56 (1983).
- 20) 高松, 藤田, 谷, 西田: 日英機械翻訳における内部表現の部分的トランスファと英文の生成, 情報処理, Vol. 26, No. 5, pp. 788-797 (1985).
- 21) Boitet, Ch. and Nedobejkine, N.: Recent developments in Russian-French machine translation at Grenoble, Linguistics, Vol. 19, pp. 199-271 (1981).

(昭和60年6月13日受付)