

英日機械翻訳における内部表現の変換

高松 忍

(大阪府立大学)

西田 富士夫

(工学部)

1. まえがき

本稿は、英日翻訳のトランスラ過程において、主題や叙述の流れを保存して原言語の内部表現を目標言語の適正な内部表現に構造変換する手法について述べている。

一般に、英語と日本語のように語族が離れている場合には、同じ意味内容のものでそれを表現する格構造が異なる場合が多い。すなわち、格構造において、格とその格に入る項の妥当な意味カテゴリーの組合わせが両言語で異なる場合が多い。

又、英語は動作主や所有者などの個体を指向する言語であり、一方、日本語は状態や動作の事象全体を指向する言語であると言われる。具体的には、英語は HAVE 型、DO 型言語であり、日本語は BE 型、BECOME 型言語である。

更に、統語構造の根本的な違いとして、動詞と目的語の語順の違いがある。このような基本的語順の違いから、関係節などの他の様々な統語構造の違いが生じる。

ここでは、以上のような英日間の格構造及び統語構造上の違いに着目して、両言語間の構造変換規則を記述することを試みた。

本方法における構造変換を、単文レベルの変換と複文レベルの変換に分けて述べる。単文レベルでは、所有(HAVE)表現から存在属性(BE)表現への変換、動作(DO)表現から状態変化(BECOME)表現への変換などについて述べ、複文レベルでは、関係節や不定詞句などの変換について述べる。

2. 内部表現

英日両言語を通じて、文及び名詞句の内部表現をそれぞれ、式(1)及び式(2)のような格構造表現で記述する。

$$(K_0-C_0:t_0, K_1-C_1:t_1, \dots, K_n-C_n:t_n) \quad (1)$$

$$t'_0(K'_0-C'_0:t'_0, K'_1-C'_1:t'_1, \dots, K'_m-C'_m:t'_m) \quad (2)$$

ここに、 t_0 は文における受けの述語を、 t'_0 は名詞句における受けの名詞語を示す。 $t_j (j=1, \dots, n)$, $t'_j (j=1, \dots, m)$ はそれぞれ、 t_0 , t'_0 を修飾する係りの句や節の内部表現を示し、一つの語の式(1)又は式(2)の形をとる。 $K_j-C_j (j=0, 1, \dots, n)$, $K'_j-C'_j (j=0, 1, \dots, m)$ はそれぞれ、 t_j , t'_j の格ラベルとその意味カテゴリー名の組を示す。又、記号“x”は名詞句の格構造における受けの名詞語 t'_0 の位置を示す。格ラベルと意味カテゴリー名は、文献(3)と同じものを用いる。

又、英文の解析と内部表現の構成、及び構造変換後の内部表現からの日本文の生成は、文献(3)の手法による。

3. 単文レベルの変換

(a) 所有表現から存在属性表現への変換

所有と存在の表現に関して2つの言語類型がある。一方は、所有述語で存在の概念をも表現する言語(HAVE言語)であり、他方は、存在述語で所有の概念をも表現する言語(BE言語)である。英語は前者の言語であり、日本語は後者の言語である。日本語の所有表現は、“HUMAN が手に PHYSICAL OBJECT をもつ”など用法が限られ、多くの他の英語の所有表現は、日本語では存在や属性の表現で表わすのが自然である。これより、式(3)、式(4)の変換規則に示すように、英語の所有表現を存在表現や属性表現に変換する。

ここに、式(3)、式(4)の規則はそれぞれ、表1、表2に示すような意味カテゴリー C_1, C_2 の組に対して適用される。又、二重下線は、その格の項が主題であることを示し、すは、任意のいくつ

かの項とそれらの格ラベルをまとめて示す(空の場合も含む)記号である。(以下、同様)

(PRED-POSSession:持つ, OBJ-C₁:t₁,
PARTICipant-C₂:t₂)

⇒ (PRED-EXIStence:在る, LOC-C₁:t₁, OBJ-C₂:t₂) (3)

表1 所有表現から存在表現への変換

C ₁	C ₂
HUMAN	HUMAN ∪ NATURAL phenomenon ∪ ABSTRACT OBJECT
NON-LIVING object ∪ PHYS-LOC	PHYS-OBJ
ABSTR-OBJ	ABSTR-OBJ

(PRED-POSS:持つ, OBJ-C₁:t₁, PARTIC-C₂:
t₂ (PRED-ATTRibute ∪ RELation:t₃, OBJ-C₂:*, 世))

⇒ (PRED-ATTRUREL:t₃, LOC-C₁:t₁,
OBJ-C₂:t₂, 世) (4)

表2 所有表現から属性表現への変換

C ₁	C ₂
PHYS-OBJ	PHYS-QUANT
LIVING object	BODY-PART

[例1]

(1) "This CRT terminal has double-faced sockets"
(PRED-POSS:持つ, OBJ-NON-LIVING:CRT端子 (OBJ:*,
DET:この), PARTIC-PHYS-OBJ:ソケット (OBJ:*, ATTR:2面型))

⇒ (PRED-EXIS:ある, LOC-NON-LIVING:CRT端子 (OBJ:*,
DET:この), OBJ-PHYS-OBJ:ソケット (OBJ:*, ATTR:2面型))
"このCRT端子には2面型ソケットがある"

(2) "This magnet has high coercive force"
(PRED-POSS:持つ, OBJ-PHYS-OBJ:磁石 (OBJ:*, DET:
この), PARTIC-PHYS-QUANT:保磁力 (PRED-ATTR:
高い, OBJ-PHYS-QUANT:*))

⇒ (PRED-ATTR:高い, LOC-PHYS-OBJ:磁石 (OBJ:*,
DET:この), OBJ-PHYS-QUANT:保磁力)
"この磁石は保磁力が高い"

(b) 動作表現から状態変化表現への変換

動作と状態変化の表現に関し、(i),(ii)に示すような2つの言語類型がある。

(i) DO型: AGent t₁が OBJECT t₂を ACTION t₀する

(ii) BECOME型: CAUSE t₁により OBJ t₂が STATE t₀になる
公知のように、英語は典型的なDO型言語であり、一方、日本語はBECOME型

言語である。日本語では、NON-LIVINGをAGentとする(i)の動作表現の用法は限られ、多くの場合、これを(ii)の状態変化表現で表わす。これより、式(5)、式(6)の変換規則に示すように、英語の動作表現を状態変化表現に変換する。

ここに、式(5)、式(6)の規則はそれぞれ、表3、表4に示すような意味カテゴリー-C₀, C₁, C₂の組に対して適用される。さらに、式(5)の場合は、もとの格ラベルAGentが格ラベルK₁に書き換えられ、式(6)の場合は、述語のMODAL格に値mが付け加えられる。又、式(5)右辺のt₀^{*}は、(I) t₀の派生自動詞語(動かす→動くなど)又は(II) t₀+VOICE:passiveである。(I)の場合、t₀^{*}は、式(5)の規則名と共にt₀の辞書に記載しておく。

(PRED-C₀:t₀, AG-C₁:t₁, OBJ-C₂:t₂, 世)
⇒ (PRED-C₀:t₀^{*}, K₁-C₁:t₁, OBJ-C₂:t₂, 世) (5)

表3 動作表現から状態変化表現への変換

C ₀	C ₁	C ₂	K ₁
Physical TRANSfer ∪ ATTR-TRANS ∪ PRODUCTION	NON-LIVING ∪ PHYS-QUANT ∪ EVENT	PHYS-OBJ ∪ PHYS-QUANT	CAUSE
Mental-TRANS ∪ PERCEPTual ACTION	ABSTR-OBJ ∪ EVENT	ABSTR-OBJ ∪ EVENT	SOURCE
⋮	⋮	⋮	⋮

(PRED-C₀:t₀, AG-C₁:t₁, OBJ-C₂: (PRED-C₂:
t₀^{*}, AG:t₁^{*}, OBJ:t₂^{*}, 世))

⇒ (PRED-C₂:t₀^{*}, MODAL:m,
CAUSE-C₁:t₁, AG:t₁^{*}, OBJ:t₂^{*}, 世) (6)

表4 動作表現から状態変化表現への変換

C ₀	C ₁	C ₂	m
CAUSE	NON-LIVING ∪ ABSTR-OBJ ∪ EVENT	ACTION	—
ENABLE			capable
PREVENT			impossible

[例2]

(1) "This feedback increases the input resistance of the amplifier"

(PRED-ATTR-TRANS:増大, AG-EVENT:フィードバック
(OBJ:*, DET:この), OBJ-PHYS-QUANT:入力抵抗 (OBJ:
*, DET:def, POSSESSOR-PHYS-OBJ:増幅器 (OBJ:
*, DET:def))

⇒ (PRED-ATTR-TRANS: 増す, CAUSE-EVENT: フィードバック
(OBJ: *, DET: in), OBJ-PHYS-QUANT: 入力抵抗 (OBJ: *,
DET: def, POSSESSOR-PHYS-OBJ: 増幅器 (OBJ: *,
DET: def)))

"このフィードバックにより増幅器の入力抵抗が増す"

(2) "An insulating layer prevents impurities from diffusing"

(PRED-PREVENT: 妨げる, AG-NON-LIVING: 絶縁層
(OBJ: *, DET: indef), OBJ-ACT: (PRED-ACT: 拡散する,
OBJ-PHYS-OBJ: 不純物))

⇒ (PRED-ACT: 拡散する, MODAL: impossible,
CAUSE-NON-LIVING: 絶縁層 (OBJ: *, DET: indef),
OBJ-PHYS-OBJ: 不純物)

"絶縁層により不純物が拡散できない"

(C) 受動表現から能動表現への変換

日本語では、AGentとOBJectの格をもつ動作文において、AGent格の語が表層上なくとも、多くの場合、述語がAGent-HUMANをとらって能動表現を用いるのが自然である。これより、式(7)の変換規則に示すように、英語の受動表現を能動表現に変換する。

(PRED-ACT: to, VOICE: passive,
OBJ-C₁: t₁, 他, [AG-HUMAN])

⇒ (PRED-ACT: to, OBJ-C₁: t₁, 他) 又は
(PRED-ACT: to, VOICE: active,
OBJ-C₁: t₁, 他, [AG-HUMAN]) (7)

ここに、角括弧はその格の語が表層上省略されていることを示す。又、t₁は、toの派生自動詞語を示す。

[例3]

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

(PRED-PERCEP-ACT: 見出す, VOICE: passive,
MODAL: capable, OBJ-PHYS-QUANT: ベース電流 (OBJ: *,
DET: def), [AG-HUMAN], CAUSE: (PRED-PERCEP-ACT: 知る, VOICE: passive, OBJ-PHYS-QUANT: コレクタ電圧 (OBJ: *, DET: def), [AG-HUMAN]))

⇒ (PRED-PERCEP-ACT: 見出す, VOICE: active,
MODAL: capable, OBJ-PHYS-QUANT: ベース電流 (OBJ: *, DET: def), [AG-HUMAN], CAUSE:

(PRED-PERCEP-ACT: 分かる, OBJ-PHYS-QUANT: コレクタ電圧 (OBJ: *, DET: def)))

"コレクタ電圧が分かるので、ベース電流を見出すことができる"

4. 複文レベルの変換

(α) 関係節を含む文の変換

英語は動詞-目的語の語順をとり、一方、日本語は目的語-動詞の語順をとる。このため、制限的關係節を含む文において著しい語順の違いが生じる。従って、英文の關係節をそのまま制限的に訳すと、語順の逆転が起り、理解しにくい訳文となる場合がある。一方、關係節を継続的に訳すと、主題が著しく変化したり、制限的意味に関して論理的に誤った訳文となる場合がある。これより、制限的關係節をそのまま制限的に訳すか、あるいは継続的に訳すかの問題が生じる。

ここでは、叙述の流れ通りに継続的に訳しても以上のような問題があまり生じないパターンを上げ、これらに対して変換規則を与えた。以下に、これらのパターンを示す。

(i) 主節が原因や手段を表わし、關係節がその結果を表わす。

(ii) 主節がBE(コプラ)文、所有文や構成關係文であり、そのPARTICIPANTやOBJectが關係節の先行詞語である。

(iii) 關係節の中に、さらに關係節や副詞節などの従属節が多重に含まれる。

(i)~(iii)はそれぞれ、(i')~(iii')のような理由による。

(i') 因果關係などは時間的に生起する順に継続的に訳すとわかり易い。(ii') 継続的に訳しても継続文は前の主文の主題を補う文となり、主題の変化が目立たない。(iii') 日本語では、先行詞語の前に多重の従属節がきて直感的に極めてわかりにくい。

(i),(ii)に対応する変換規則をそれぞれ、式(8),式(9)に示す。ここに、t

は先行詞語を示し、下線部はその関係節の内部表現を示す。又、PRECEDEは変換された重文の内部表現における先行文の格を示し、角括弧内は、訳文の生成において代名詞化又は省略することを示す。

($PRED: t_0, \#$, $K_1: t(PRED: t'_0, \{ \text{CAUSE}, \text{SO} \}: *, \#')$) $\Rightarrow$ ($PRED: t'_0, \{ \text{CAUSE}, \text{SO} \}: t, \#'$, $PRECEDE: (PRED: t_0, \#$, $K_1: t)$) (8)

($PRED-C_0: t_0, \underline{K_1-C_1: t_1}, K_2-C_2: t(PRED: t'_0, K'_1: *, \#')$) $\Rightarrow$ ($PRED: t'_0, [K'_1: t_1], \#'$, $PRECEDE: (PRED-C_0: t_0, \underline{K_1-C_1: t_1}, K_2-C_2: t)$) (9)

ここに、式(9)の規則は、表5に示すような意味カテゴリー- C_0 、格ラベル K_1 、 K_2 の組に対して適用される。但し、 $C_0 = BE$ の場合、 C_2 は C_1 の上位の意味カテゴリーでないという制約がある。

表5 関係節を含む文の変換

C_0	K_1	K_2
POSSession $\cup$ COMPOSITION $\cup$ BE	OBJ	PARTIC
EXISTence	LOC	OBJ

[例4]

(1) "The control module converts the pattern into indexing signals that position the X-Y table"

($PRED-ATTR-TRANS$: 変換する, AG : 制御モジュール ($OBJ: *, DET: def$), OBJ : パターン ($OBJ: *, DET: def$), GO : インデックス信号 ($PRED-PTTRANS$: 位置決める, AG : *, OBJ : X-Yテーブル ($OBJ: *, DET: def$))) $\Rightarrow$ ($PRED-PTTRANS$: 位置決める, $VOICE: passive$, $CAUSE$: インデックス信号, OBJ : X-Yテーブル ($OBJ: *, DET: def$), $PRECEDE: (PRED-ATTR-TRANS$: 変換する, $VOICE: passive$, $CAUSE$: 制御モジュール ($OBJ: *, DET: def$), OBJ : パターン ($OBJ: *, DET: def$), GO : インデックス信号))

"制御モジュールによりパターンがインデックス信号に変換され、インデックス信号によりX-Yテーブルが位置決めされる"

上例では、主節と関係節の各単文に対し、前節式(5)の規則による変換を行な、た

後、式(8)の規則による変換を行なっている。

(2) "Transistors Q_7, Q_8, Q_9 and Q_{10} form a differential amplifier in which Q_7 and Q_8 operate in the common-collector mode"

($PRED-COMPOS$: 構成する, OBJ : トランジスタ (Q_7, Q_8, Q_9, Q_{10}), $PARTIC$: 差動増幅器 ($OBJ: *, DET: indef$) ($PRED$: 動作する, $LOC: *$, AG : (Q_7, Q_8), $MANN$: 共通コレクタモード ($OBJ: *, DET: def$))) $\Rightarrow$ ($PRED$: 動作する, (LOC : 差動増幅器), AG : (Q_7, Q_8), $MANN$: 共通コレクタモード ($OBJ: *, DET: def$), $PRECEDE: (PRED-COMPOS$: 構成する, OBJ : トランジスタ (Q_7, Q_8, Q_9, Q_{10}), $PARTIC$: 差動増幅器 ($OBJ: *, DET: indef$)))

"トランジスタ Q_7, Q_8, Q_9 と Q_{10} は差動増幅器を構成し、 Q_7 と Q_8 は共通コレクタモードで動作する。上例では、式(9)の規則による変換を行っている。

(3) "The result is a measure of temperature that is sent to a control circuit that maintains proper temperature"

($PRED-BE$: である, OBJ : 結果 ($OBJ: *, DET: def$), $PARTIC$: 尺度 ($OBJ: *, DET: indef$, $POSSESSOR$: 温度) ($PRED$: 送る, $VOICE: passive$, OBJ : *, GO : 制御回路 ($OBJ: *, DET: indef$) ($PRED$: 維持する, $OBJ: *, PARTIC$: 温度 ($PRED$: 適切な, $OBJ: *$))) $\Rightarrow$ ($PRED$: 送る, $VOICE: passive$, (OBJ : 尺度), GO : 制御回路 ($OBJ: *, DET: indef$) ($PRED$: 維持する, $OBJ: *, PARTIC$: 温度 ($PRED$: 適切な, $OBJ: *$)), $PRECEDE: (PRED-BE$: である, OBJ : 結果 ($OBJ: *, DET: def$), $PARTIC$: 尺度 ($OBJ: *, DET: indef$, $POSSESSOR$: 温度)))

"結果は温度の尺度であり、適切な温度を維持する制御回路に送られる"

上例では、式(9)及び(iii)に対応する規則により、多重の関係節の変換を行っている。

(b) 不定詞句を含む文の変換

後置の不定詞句の意味として、 $PURPOSE$ と $RESULT$ がある。しかしながら、主節が状態変化($BECOME$)表現で、 $OBJECT$ が $NON-LIVING$ などの無意志物なら、合目

的動作はとらず、不定詞句の格はRESULTに定まる。これより、主節がこのような条件を満たす場合には、式(10)の変換規則により、主節-不定詞句の順に叙述の流れ通り訳すのが自然である。

(PRED: to, 主, {PURPOSE}, (PRED: to, 主'))

⇒ (PRED: to, 主, PRECEDE: (PRED: to, 主)) (10)

ここに、下線部は後置の不定詞句の内部表現を示す。

【例5】

"When the speed exceeds a safe level, centrifugal force unfolds the petals to block air flow"

(PRED: 開く, AG: 遠心力, OBJ: 花弁 (OBJ: *, DET: def), {PURPOSE u RESULT}: (PRED: ふさぐ, OBJ: 空気流, [AG: 遠心力]),

TIME: (PRED: 越える, OBJ: 速度 (OBJ: *, DET: def), PARTIC: 安全レベル (OBJ: *, DET: indef)))

⇒ (PRED: ふさぐ, VOICE: passive, OBJ: 空気流, [CAUSE: 遠心力], PRECEDE: (PRED: 開く, CAUSE: 遠心力, OBJ: 花弁 (OBJ: *, DET: def), TIME: (PRED: 越える, OBJ: 速度 (OBJ: *, DET: def), PARTIC: 安全レベル (OBJ: *, DET: indef))))

"速度が安全レベルを越えるとき、遠心力により花弁が開き、空気流がふさがれる"

上例では、主節と不定詞句の各単文に対し、前節式(5)による変換を行なった後、式(10)による変換を行なっている。

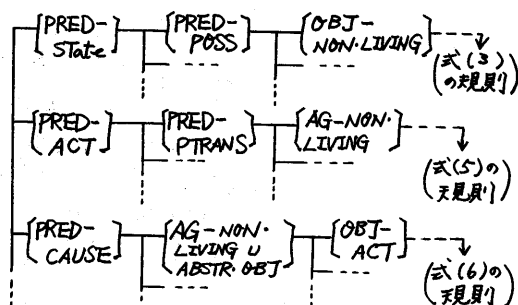
5. 実験

上記の変換規則を用いた構造変換プログラムをLISP言語で作成し、これを英日翻訳システム⁽³⁾に組み込み、翻訳実験を行なった。英文の解析によりえられた内部表現の変換は、まず基本となる単文レベルの変換を行ない、続いて複文レベルの変換を行なう。

変換規則の記述形式は、(〈適用条件〉〈変換操作〉) である。ここに、〈適用条件〉

は、述語の意味カテゴリーやキーとなる格ラベルとその格に入る語の意味カテゴリーなどの条件を記述する。〈変換操作〉は、〈適用条件〉が満たされた場合に行なうべき構造変換や格ラベルの置換などの操作列を記述する。本システムでは、適用可能な変換規則の探索を能率的に行なうため、〈適用条件〉部を図1に示すように階層化している。

図1 単文レベルの変換



以下に、(a)入力英文、(b)解析後の内部表現、(c)構造変換後の内部表現、(d)出力日本語を示す。

【I】では、3節式(4)の規則により、所有表現を属性表現に変換し、【II】では、3節式(7)の規則により、受動表現を能動表現に変換している。又、【III】では、3節式(5)の規則により、主節と関係節の各単文を状態変化表現に変換し、次いで4節式(8)の規則により、主節と関係節間の変換を行なっている。

【I】

(a) (THE REFERENCE MEANS HAS A TEMPERATURE VOLTAGE COEFFICIENT SIMILAR TO THE TEMPERATURE COEFFICIENT OF THE BATTERY)

(d) 参照手段は 温度電圧係数が バッテリーの 温度係数と 類似する

【II】

(a) (IF THE VOLTAGE IS APPLIED IN THE REVERSE DIRECTION, THE BARRIER HEIGHT IS INCREASED AND LITTLE MAJORITY CARRIER FLOW OCCURS)

(d) 電圧を 逆方向に 印加すれば 障壁高さが 増加し、多数キャリアの流れは 殆ど 起らない

[Ⅲ] (a) (CURRENT IN EXCITING COIL GENERATES A
MAGNETIC FIELD WHICH MOVES THE ARMATURE
AGAINST THE FORCE OF FLUID)

(b) ((C2386
(生 VT KA Z
(RN C2386 (生 VI KA Z)))
(M= (ASP IMP) (TNS PRE))
(AG
((C1510 (電流 N))
(LOC ((C14160 (励磁コイル N))))))
(OBJE
((C1510 (磁界 N))
(DET ((C*** (INDEF)))))
(PMOD
((C21510
(動か VT D5 K
(RN C21510 (動 VI D5 K)))
(M= (ASP IMP) (TNS PRE))
(AG ((C1510 (磁界 N)))
(OBJE
((C1463 (電機子 N))
(DET ((C*** (DEF)))))
(PARTIC-0
((C11400 (力 N))
(QUA*OBJ ((C1510 (流体 N))))
(DET ((C*** (DEF)))))

(c) ((C21510 (動 VI D5 K))
(M= (ASP IMP) (TNS PRE))
(CAUSE-T ((C1510 (磁界 N)))
(OBJE
((C1463 (電機子 N)) (DET ((C*** (DEF)))))
(PARTIC-0
((C11400 (力 N))
(QUA*OBJ ((C1510 (流体 N))))
(DET ((C*** (DEF)))))
(PRECEDE
((C2386 (生 VI KA Z))
(M= (ASP IMP) (TNS PRE))
(CAUSE-T
((C1510 (電流 N))
(LOC ((C14160 (励磁コイル N))))))
(OBJE
((C1510 (磁界 N)) (DET ((C*** (INDEF)))))
SS)

(d) 励磁コイルの電流により 磁界が
生じ、磁界により 電機子が
流体の力に対抗して 動く

6. おすび

英日間の構造変換には、以上述べた他に、"many"や"some"などの数量形容詞語の副詞化、述語化や、It~that~文の変換などがあるが、ここでは省略した。

構造変換で重要な問題は、変換すべき適用条件すなわちコンテクストを明確にすることである。今後、より適切な訳文を生成するため、変換規則の拡充、精緻

化ならびに体系化が必要である。

文献

- (1) M. Nagao et al "A machine translation system from Japanese into English - another perspective of MT system -" COLING-80 (1980) 414-423.
- (2) F. Nishida et al "English-Japanese translation through case-structure conversion" COLING-80 (1980) 447-454.
- (3) 高松・西田 "動詞・フーシと格構造に基づく英日機械翻訳" 信学論(D) J64-D, 9 (1981) 815-822.