

意味理解過程による機械翻訳方式およびその問題点*

横田将生 田町常夫
(九州大学)

1. まえがき

機械翻訳の方式として中間言語方式とトランスファ方式が考えられる。中間言語方式は各国語の処理は独立に行うことができ、入力言語の構造に影響されない目的言語への変換が可能である。しかしながら、この方式は中間言語の設計など技術的に困難が多いため特定の言語対に依存するが比較的達成が容易と思われるトランスファ方式が今日主流をなしている。筆者らはここ数年、中間言語の開発およびそれに基づく自然言語の体系的な意味理解の研究を行い⁽¹⁾、気象情報理解システム (ISOBAR)⁽²⁾⁽³⁾ および自然言語理解システム (IMAGES-I)⁽⁴⁾ を作成した。これら二つのシステムは、入力を中間言語で意味解釈し、その解釈結果より入力文に対する言い換え文を出力するという機能を有している。これは入出力の言語が異なれば翻訳であり、すでに ISOBAR では気象通報世界という限定された領域での日英翻訳を試みている⁽⁵⁾。IMAGES-I は ISOBAR よりも一般的なシステムであり、現段階では英語を入出力言語としているが、将来は複数国語間での翻訳を予定している。本稿では、これら二つのシステムを作成した経験を基に中間言語を用いる機械翻訳の一方式およびその問題点について論じることとする。

2. システム構成および翻訳過程

図1は、ISOBAR および IMAGES-I による翻訳過程を概念的に示したものである。入力言語文は、まず構文解析

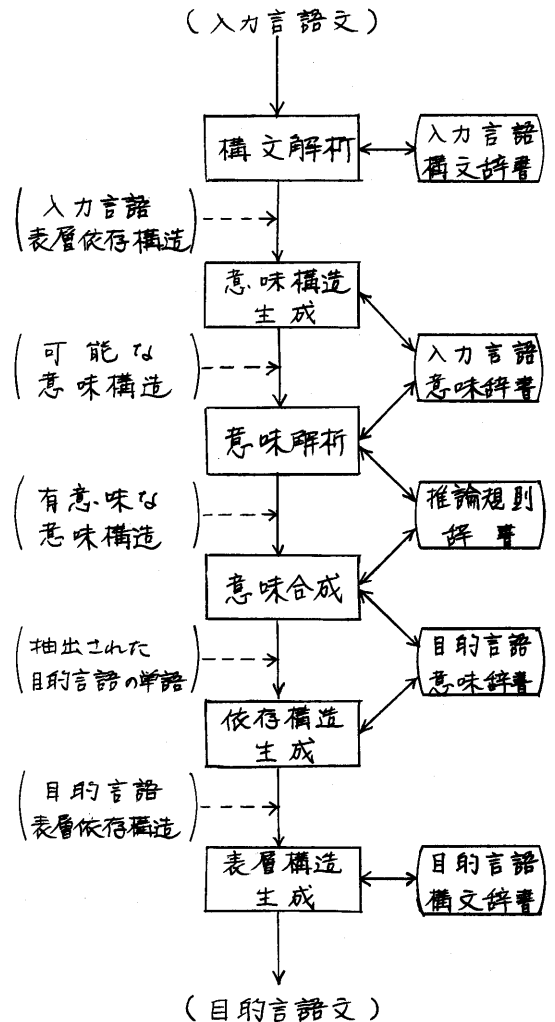


図1 中間言語による意味理解過程を経由する機械翻訳

*A method for machine translation via understanding process and its problems
by Masao YOKOTA and Tuneso TAMATI
(Interdisciplinary Graduate School of Engineering Sciences, Kyushu University).

過程で単語分割、語群形成処理を受け、表層依存構造に変換される。つづいて、意味構造生成過程で入力言語意味辞書を参照することにより、意味構造(中間言語表現)に変換され、意味解析過程に渡される。この一連の処理において、構文解析過程では、未知語を含む文、語群形成規則または依存構造規則を満たさない文は全て棄却される。また意味構造への変換が不可能な表層依存構造や、意味解析の結果、無意味と判定された意味構造も棄却される。

意味合成過程では、入力言語文の意味構造および推論規則の適用により生じた意味構造が、目的言語意味辞書に従って、単語化され、依存構造生成過程において目的言語表層依存構造に変換される。そして最終的に目的言語構文辞書により、英語表層構造が生成される。その際、入力言語文の構造に関する情報は全く参照されないもので、一般には paraphrase 関係にある複数個の目的言語文が生成される。ここでの過程は入力言語文を意味構造に変換する過程は全く可逆的であり、構文および意味辞書は、意味解析に用いるものを逆に解釈することによりそのまま使用する。

3. 意味記述方法の概要⁽¹⁾⁽²⁾

表層依存構造および意味構造は各単語の意味辞書に従って相互に変換される。各単語の意味は概念部と結合操作部により表現され、前者は論理式、後者は依存関係にある単語の概念部同士を結合させる操作手順(逆に解釈すれば概念構造から依存構造を生成させる操作手順)により成っている。ここでは、この両者に関して簡単に説明する。

3.1 概念部の構成

概念部は原子軌跡式(α)を同時的連言(Π)、継時的連言(\bullet)および記号論理における結合子(連言(\wedge),

選言(\vee)など)で結合したものととして表現される。ここで α は式(1)に示す述語式、またはそれを略記した形式(2), (3)で与えられる。この式は、図2に示すような属性空間(A)における線分軌跡を指示している。この軌跡は事物 x の作用により事物 y の属性が時間的または空間的に変化あるいは持続($a_{t_i}, a_{t_j}, P_I, P_{II}$ で表示)することを意味する。

$$\alpha \equiv f(x, y, a_{t_i}, a_{t_j}, A, P_I, P_{II}) \quad (1)$$

$$\equiv (x, y, a_{t_i}, a_{t_j}, A) \quad (2)$$

$$\equiv \begin{bmatrix} x \\ y \\ a_{t_i} \\ a_{t_j} \end{bmatrix} A \quad (3)$$

ここで $A = (A, P_I, P_{II})$, $t_i < t_j$ である。また同時的連言および継時的連言は次の式(4), (5)で定義される。

$$\alpha_1 \Pi \alpha_2 \equiv \alpha_1 \wedge \alpha_2 \wedge (t_1 = t_3 \wedge t_2 = t_4) \quad (4)$$

$$\alpha_1 \bullet \alpha_2 \equiv \alpha_1 \wedge \alpha_2 \wedge (t_2 = t_3) \quad (5)$$

但し,

$$\alpha_1 \equiv (x_1, y_1, a_{t_1}, a_{t_2}, A_1) \quad (6)$$

$$\alpha_2 \equiv (x_2, y_2, a_{t_3}, a_{t_4}, A_2) \quad (7)$$

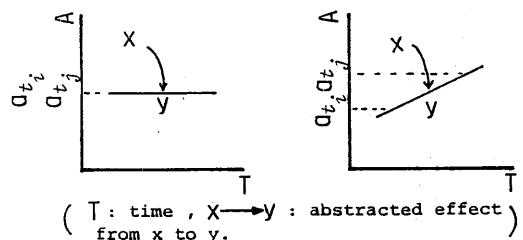


図2 属性空間における原子軌跡

4. 各処理過程の検討

本方式の特長は冒頭で述べた通りであるが、まだ検討すべき問題点が多くある。以下では、主要な問題点に関して具体的に説明し検討を加える。

4.1 表層依存構造解析に関して

例文 E-1 において、‘橋の上で’および‘川で’は等しく‘走らせる’に依存するが、前者は‘太郎’、後者は‘模型船’の場所を指示していると考えられ、このままでは中間言語表現に変換する際に不都合が生じる。

(E-1) 橋の上で太郎は、いつも模型船を、ゆっくり川で走らせる。

このような例は使役型の動詞(～せる、～させる)などで多く見られ、E-2もそのような例である。

(E-2) いやいや上官は兵卒をゆっくり行かせる。

この例で‘いやいや’は‘上官’、‘ゆっくり’は‘兵卒’の様子を指示している。このような場合、E-1のような言い換えに等価な処理が必要である。

(E-1') 橋の上で太郎は、いつも模型船に、ゆっくり川で走らせることをさせる。

次の E-3, 4 は通常よく見られる動詞の複合形も含んでいる。

(E-3) 太郎は菓子を買って食べた。

(E-4) 太郎は氷を溶けて飲んだ。

これら二つの場合、通常、後の動詞の主語や目的語は前の動詞のものに一致すると考えられる。しかしながら、やはり中間言語表現に変換するとき、E-4 の場合は不都合が生じる。なぜなら‘氷を飲む’というのは意味的に異常だからである。このような例は前の動詞が‘質的变化’を伴った事象を指示している場合に見られる。この場合、‘質的变化’の結果生じたもの、すなわち‘水(溶けた氷)’を中間言語表現中に陽に入れてやる必要がある。

4.2 意味的異常の検出に関して

現段階で考えられる意味的異常に関しては以下のようなのがある。

i) 概念要素の多重結合

(E-5) To Mary Tom gives Jim
the book.

ii) 概念要素の種類が不整合

(E-6) The color of the book is square.

(E-7) The color of the book is rain.

(E-8) The red of the book is

(E-9) The color of shape is

iii) 結合に必要な概念部分がない

(E-10) Tom stays from New York.

(E-11) The water melts.

iv) 事物が言及される属性を損たない

(E-12) The paint is silent.

(E-13) He stays here swiftly.

(E-14) Diving is red.

v) 事物が同時に同一の属性に関して異なる値を持つ

(E-15) The hot water is cold.

(E-15) The boy in Tokyo is in Osaka.

(E-16) Slowly, the dog runs fast.

(E-17) The man rises downward.

vi) 現実世界の事実に反する

(E-18) Alligators wear shoes.

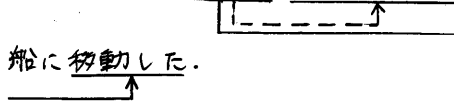
(E-19) Cheese is made from beer.

以上の種類のうち、i) ~ v) は比較的容易に検出できるが、vi) に関しては現実世界における個々の膨大な事例をシステムが知識として持っておく必要がある、非常に困難である。

4.3 曖昧さの除去に関して

構造的曖昧さの除去に関しては、E-20 のように意味的異常性の検出により、あるいは E-21 のように同義性(同一の中間言語表現になること)の検出により除去することができるとあるが、E-22, 23 のような意味的な曖昧さの除去に関しては、本質的に文脈処理が必要であり、そのためには推論処理の強化が不可欠である。

(E-20) 太郎は素早く停止している

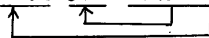


(E-21) Tom with the stick follows Jim.



(E-22) The bill is large.

(E-23) Tom follows Jim with the stick.



4.4 推論処理に関して

現在、次の E-24 から E-25, 26 を導出するような推論規則をシステムに知識として入れている。

(E-24) Tom goes with the book.

(E-25) The book moves.

(E-26) Tom carries the book.

このような推論規則を翻訳システムに採用した場合、目的言語に例えば 'with' に対応する概念が存在せずに 'carry' に対応する概念が存在する場合等には有効であるが、そうでない場合には、多くの無駄な翻訳結果を出力することになりかねない。

また二重否定に関する定式(すなわち $\sim\sim A \equiv A$)などの採用も翻訳を意味乾燥にしてしまう可能性がある。

4.5 目的言語の単語抽出に関して

入力言語の意味解釈結果である中間言語表現から目的言語の単語概念を検出して単語化する場合、目的言語の意味辞書の概念部を参照することにより、まず述語(動詞)概念を抽出する。その際、入力言語の構造に関する情報を参照してはいないので、E-27の入力言語の意味構造から "be red", "be slow" などを最初に抽出することが適切、目的言語文として最も適切と思われる E-28 の他に E-29, 30 などの多くの paraphrase を生成する。

(E-27) 赤い車がゆっくり走る。

(E-28) The red car runs slowly.

(E-29) The car which runs slowly is red.

(E-30) The red car which runs is slow.

このような無駄と思われる出力をささない、あるいは出力順序の優先度を小さくするためには、少なくとも、入力言語の主述語(主動詞, main verb)および主語の概念が意味構造のどこを占めているかの情報を残しておく必要があると思われる。

4.6 代名詞化処理に関して

代名詞化処理はトランスフォーマ式においてもかなり困難な問題の一つである。E-3, 4の英訳は下の E-31, 32 がそれぞれ適当であると思われる。

(E-31) Taro bought the cake and ate it.

(E-32) Taro melt the ice and drank it.

E-31は単語抽出過程で 'ate' の目的語がすでに 'bought' の目的語として単語化されている 'cake' と同一個体ということと比較的容易に 'it' と代名詞化できるであろう。しかしながら E-32 の場合は、E-33 の波線部分(ある種の文脈)の代名詞化であり、意味理解処理を行ったために、かえって翻訳が複雑になる例の一つである。

(E-33) Taro melt the ice and drank the water which was the melt ice.

4.7 表層構造生成に関して

最も重要なことは翻訳した結果の目的言語文が可能な限り曖昧さも含まないことと思われる。翻訳という立場では曖昧な入力言語文は曖昧な目的言語文に翻訳するといった考えもありうるが、E-1を次の英文 E-34 に翻訳した場合などのように、曖昧さの等価性が問題となろう。

(E-34) Taro always runs the model ship slowly on the bridge in the river.

この問題は訳出語彙に深くかかわることによって本方式に限ったことではない。つぎに重要なことは、表層依存構造

が異っていても表層構造が一致する場合(これは構造的曖昧さも含んだ目的言語文を出力する時に起こり、E-21, 23などがその例である)には重複出力をさせないことであろう。

その他、この過程では、トランスファオ式と共通の問題がかなりあると思われる。

5. 格文法による意味処理の可能性

以上の章では、中間言語による意味理解過程を經由する機械翻訳の一方式とその問題点を検討した。現在、機械翻訳における重要課題は意味処理をどの程度、どのように導入するかということである⁽⁶⁾。ここでは、現在、意味処理の標準的な手法となっている格文法による方法すなわち動詞の格要素に意味的標識を付けるやり方でどの程度の処理が可能であるかを考察してみる。

5.1 意味的異常性の検出

4.2で述べた意味的異常性のうち、E-5, 11のようなものが検出可能である。E-5は同一格要素の多重結合を禁止することにより、E-11は動詞の必須格および名詞に付された意味標識を利用する選択制限(selection restriction)により検出できる。しかしながら、この場合、E-4も異常として次のE-35を意味的に適格であるとする可能性がある。

(E-35) 太郎は水を凍らして沸した。

格文法では、必ず存在しなければならぬ必須格と存在しなくてもよい任意格という二通りの格がある。しかしながら、任意格の中には、意味的に存在してはならない格(これを禁止格と呼ぶことにする)がある。例えばE-10の'from'に導かれる'始点格'と呼ぶべき格が'stay'の禁止格の一つである。このような禁止格を各動詞に関して設定すれば、更に意味的異常性の検出能力を高めることができるが、

格の質および量が問題となる。

5.2 意味的曖昧性の検出

格構造は本質的には表層依存構造に格標識を付けたものであるから、E-23のような意味的曖昧さは表示できるが、E-20, 21も曖昧なものとする可能性がある。E-22のような単語概念による意味的曖昧さは意味標識の違いにより表示が可能であるが、意味標識の質および量が問題となる。

5.3 同義性の検出

本来、格文法は意味表現の正確性に関しては期待できない。ただE-36, 37のように態が異なるだけの文同志、あるいは、E-38, 39, 40のように前もって同義とわかっている文型を同一の格構造で表現するのは可能であろう。

(E-36) The cat bit the mouse.

(E-37) The mouse was bit by the cat.

(E-38) 太郎はリンゴを喜ぶ。

(E-39) 太郎はリンゴで喜ぶ。

(E-40) 太郎はリンゴに喜ぶ。

しかしながら、E-38, 39, 40のような例に関しては、格標識の質が個々の動詞に強く依存することになる。また、この'リンゴ'にどんな名称の格を与えようとE-41に示す受託を得るのに余り役に立つとは思えない。

(E-41) Taro is pleased with apples.

6. おまじ

本稿で述べた中間言語による機械翻訳の研究はまだ緒についたばかりで、これからまた多くの問題が明らかになってくるであろう。

文 献

- (1) 横田, 田町: "視覚化された概念モデルに基づく自然語の意味解釈について", 信学論(D), 63-D, 5 (1980).
- (2) 横田, 田町: "言語-図形の対応構造に基づく気象通報の意味解釈について", 信学論(D), 63-D, 6 (1980).
- (3) E. Kawaguchi, et al.: Proc. of 6th IJCAI, 1 (1979).
- (4) 横田, 田町: "自然言語理解システムIMAGE-S-Iにおける意味処理について", 旭総理工研報, 2-2 (1981).
- (5) 横田, 他: "ISO BARによる日英機械翻訳について", 電気通信学会技術研究報告, AL81-44 (1981).
- (6) 長尾, 江井: "機械翻訳の現状", 電気通信学会誌, Vol. 64, No. 5 (1981).