

英日機械翻訳における意味構造変換に関する一考察

片桐 恭弘

野村 浩郷

(日本電信電話公社)

武蔵野電気通信研究所)

1. はじめに

翻訳に対する需要はますます増加しつつあり、これに伴って機械翻訳システムに関して日本のみならず世界各地で盛んに研究が成されている。^{[1]-[7]}しかしながら、従来人手によって成されていた翻訳に関する種々の技術を機械翻訳システムに取り込むという研究方向はあまり見られない。これは一つには、人手による翻訳の対象となるのは詩や小説に代表される文学作品が多く、文体やニュアンス等、高次の要素の保存・移行が翻訳上の主な問題点として取り上げられる傾向にあるのに対し、現状では機械翻訳システムの対象とするのはなるべく単純・定型的で機械的処理に適した文であり、翻訳の目標もまず正しい訳文を得るという点に重点が置かれるため、両者の間に問題意識の差がある点に原因があろう。しかしながら、人手によるにせよ、機械を用いるにせよ、翻訳という共通の作業を行う以上当然両者の間に共通な要素は多く存在するはずであるし、特に機械的処理による翻訳がある程度可能となってきた現状では、より自然で実用的な機械翻訳を実現するために人手による翻訳の技術を参考にし、取り入れることが必要になると考えられる。

以上のような観点から本稿では、まず翻訳過程について考察を加え、次に従来翻訳を行う上で留意すべきであると考えられている諸点を概観し、さらにそれらを機械翻訳処理に取り入れる際の問題点について述べる。

機械翻訳処理の手法としては、現在いわゆる構造変換による方法が多く用いられるようになってきている。これはまず原言語解析によって原文から原言語に則した意味表現を抽出し、これを構造変換処理によって目的言語に則した意味表現に変換し、それから翻訳文を生成するという三段階の処理から成り立っている。このうち、原言語解析および目的言語生成の処理は機械翻訳のみならず他の自然言語処理とも共通する技術であるが、両者を媒介する構造変換処理は翻訳に固有であり、原言語と目的言語とを媒介する機能を果たすものである。ここでは翻訳手法のうちでも特に構造変換処理に関連するものを取り上げる。

2. 翻訳に対する考え方

2.1. 翻訳の過程

ある種類の言語によって情報が伝えられる過程では、発話者は一定の伝達内容(情報)を他者に伝達する意図を持ち、その言語の表現を生成する。この際、発話者は話す場合書く場合を問わず一定の受け手を想定している。このようにして生成された言語表現は、当然その言語の言語形式に従うことによって意味内容を担っている。すなわち、言語表現はその言語の語彙に属する単語をその言語の文法に従って配列して構成されている。言語表現の担う意味内容には、このような言語形式によって決定される部分の他に、言語外の要因すなわち文化に依存する部分もある。これは言わば言語表現の情緒的意味・ニュアンスに相当すると考えることができるが、これは当然言語によって異なるものである。例えば、日本語で「頭を掻く」と言えばはにかみの表現であるが、英語で対応する動作を表わ

す“to scratch one's head”は苛ら立ちの表現となる。翻訳という作業はこのような言語による伝達を一つの言語(原言語)から他の言語(目的言語)へと移しかえる役目を担っている。従って、この過程で翻訳者は図1に示すように、両者の文化のどちらにも所属し、異種言語間での情報伝達を媒介する必要がある。すなわち、翻訳者はまず原言語の文化の枠組の中で原文を理解し、原文発話者の意図を汲みとり、それを目的言語の文化の枠組の中で、目的言語内の受け手を想定して翻訳文として表出しなければならない。

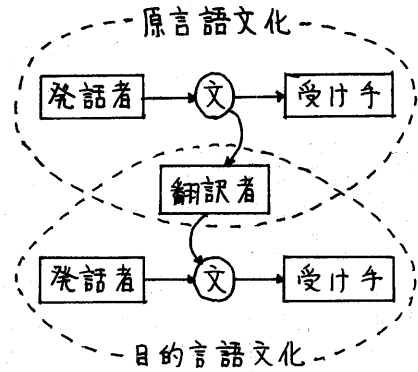


図1. 翻訳の過程.

2.2 翻訳における形式的等価と動的等価^[8]

翻訳者が原文から翻訳文を構成する際の基本的方針には大きく分けて二通りが存在する。それらは原文と翻訳文との形式的等価を目指す方針と、動的等価を目指す方針と呼ぶことができる。前者はいわば直訳・逐語訳に対応し、原文に重点を置く立場である。すなわち、原言語の伝達文の形式と内容を、そのまま目的言語に移すことを目的とする。従って原文中の単語は翻訳文でもやはり単語に、しかも名詞は名詞に、動詞は動詞にというように品詞を保存して翻訳し、文に関しても一文を一文に対応させて翻訳し、むやみに分割したり統合したりはしない。また、慣用的表現や、原言語と目的言語とで文化的背景の異なる表現等も原文に忠実に翻訳し、翻訳文のみで十分意味の伝わらない場合には脚注を付して補うという方針である。

これに対して後者の動的等価を目指す翻訳は、いわば意訳に対応し、目的言語の受け手が翻訳文から受ける効果か、原言語の受け手が原文から受ける効果と同等になることを目標とする。従って原言語の形式は時には無視され、むしろ翻訳文の目的言語としての自然さが重視される。また、文化的背景に関しても、原文が原言語の文化内で果たした機能と同等の機能を翻訳文が目的言語の文化の中で果たすような翻訳が求められる。

これら両者の立場は翻訳の方針の両極端を示すものであって、実際に行われている翻訳は、文の種類や性格に応じてこれら両者の間に分布している。例えば、科学技術解説文は小説等と比べるとより形式的等価に近い翻訳で十分なことが多い。また反対に、表層に現われる音形が重要な意味を持つ洒落のように動的等価を目指す以外にうまい翻訳の方法の無い場合もある。例えば、

Do you know how to change an apple into a pear ?

Add one, and you will have a pain.

うまく翻訳するには、動的等価を目指して

りんごをなしに変えるにはどうすればいい？

食べてしまえばなしになる。

とでもする以外に方法がない。^[10]

翻訳に対する形式的等価の立場と動的等価の立場との比較対照のために一例と

して、聖書の一節をギリシャ語原文から二種類の異なる英語文へと翻訳した例を図2に示す。形式的等価の立場は機械的処理によって実現するという観点から見るとより容易であるが、全くの逐語訳を行ったのでは得られる翻訳文は一般に分かりにくく、慣用的表現に対しては誤訳を生みやすい。これに対して動的等価の立場は理想的に機能すれば良質の翻訳文が得られるが、これを実現するには伝達内容に対するより深い理解が必要とされ、特に機械的処理によってこれを実現するのは困難である。今後は、逐語訳によって得られる不自然な翻訳文や誤まった翻訳文をより良質の翻訳文とするために、人手による動的等価を目指した翻訳の特徴を分析し、順次機械処理に取り入れて行く必要があると思われる。

ヨハネ伝1章6-8節

ギリシア語:

¹ egeneto ² anthrōpos, ³ apetalmenos ⁴ para ⁵ theou, ⁶ onoma ⁷ autō ⁸ Iōannēs;
⁹ houtos ¹⁰ elthen ¹¹ eis ¹² marturian, ¹³ hina ¹⁴ marturēsē ¹⁵ peri ¹⁶ tou ¹⁷ phōtos, ¹⁸ hina
¹⁹ pantes ²⁰ pisteusōsin ²¹ di' ²² autou, ²³ ouk ²⁴ ēn ²⁵ ekeinos ²⁶ to ²⁷ phōs, ²⁸ all' ²⁹ hina
³⁰ marturēsē ³¹ peri ³² tou ³³ phōtos.

(a) ギリシャ語原文。

直訳的移転(第一段階):

¹ became / happened ² man, ³ sent from ⁴ God, ⁵ name ⁶ to-him ⁷ John; ⁸ this /
⁹ the-same ¹⁰ came-he ¹¹ into / ¹² for ¹³ testimony / witness, ¹⁴ that / in-order-that
¹⁵ testify / witness-might-he ¹⁶ concerning / of / ¹⁷ to the light, ¹⁸ that / in-order-
¹⁹ that all ²⁰ believe-might-they ²¹ through ²² him. ²³ not ²⁴ was ²⁵ he / that-one ²⁶
²⁷ light, ²⁸ but ²⁹ that / in-order-that ³⁰ witness-might-he ³¹ concerning / of / ³² to the
³³ light.

(b) 単語の置換。

最少移転(第二段階):

¹ There ² CAME / WAS ³ a man, ⁴ sent from ⁵ God, ⁶ WHOSE ⁷ name ⁸ was ⁹ John;
¹⁰ HE / the-same ¹¹ came ¹² into / ¹³ for ¹⁴ testimony / witness, ¹⁵ that ¹⁶ he ¹⁷ might ¹⁸ witness
¹⁹ of / ²⁰ to the light, ²¹ that ²² all ²³ might ²⁴ believe ²⁵ through ²⁶ him. ²⁷ He ²⁸ was ²⁹ not
³⁰ the light, ³¹ but ³² came ³³ that ³⁴ he ³⁵ might ³⁶ witness ³⁷ of / ³⁸ to the light.

(c) 英語の構文に合致させる。

ASV:

¹ There ² CAME ³ a man, ⁴ sent from ⁵ God, ⁶ WHOSE ⁷ name ⁸ was ⁹ John. ¹⁰ The
¹¹ same ¹² came ¹³ for ¹⁴ witness, ¹⁵ that ¹⁶ he ¹⁷ might ¹⁸ BEAR ¹⁹ witness ²⁰ of the light, ²¹ that
²² all ²³ might ²⁴ believe ²⁵ through ²⁶ him. ²⁷ He ²⁸ was ²⁹ not ³⁰ the light, ³¹ but ³² came ³³ that
³⁴ he ³⁵ might ³⁶ BEAR ³⁷ witness ³⁸ of the light.

(d) 形式的等価を目指した翻訳。

PME:

¹ * A man ² NAMED ³ * John ⁴ WAS ⁵ sent ⁶ BY ⁷ God ⁸ * * ⁹ AS ¹⁰ a ¹¹ WITNESS ¹² * ¹³ to
¹⁴ the light, ¹⁵ so ¹⁶ that ¹⁷ ANY ¹⁸ MAN ¹⁹ WHO ²⁰ HEARD ²¹ HIS ²² TESTIMONY ²³ might ²⁴ believe
²⁵ in ²⁶ the light. ²⁷ THIS ²⁸ MAN ²⁹ WAS ³⁰ not ³¹ himself ³² the light: ³³ * he ³⁴ was ³⁵ SENT
³⁶ simply ³⁷ AS ³⁸ a ³⁹ personal ⁴⁰ WITNESS ⁴¹ TO ⁴² THAT ⁴³ light.

(e) 動的等価を目指した翻訳。

図2. 翻訳における形式的等価と動的等価。
 (文献[8]より引用)

3. 動的等価を目指した翻訳の諸技術

ここでは、英語・日本語間翻訳で、動的等価を実現するために原文と翻訳文とを逐語的に対応させず、形式を変化させている例をいくつか具体的に示す。これらの翻訳の技術を順次機械翻訳処理に取り入れることによってより自然な翻訳が得られることになる。なお、以下に示す例は、主に科学技術解説文(J.A.Feldman, "Programming Languages," Scientific American, no. 12, 1979) および文献[10]から引用したものである。

(a) 多義語の取り扱い

一般に単語とその表わす概念との対応は一対一ではなく、しかもその対応の様子は言語によって異なる。次の例は、行為の過程を表わす概念およびその過程によって得られる結果を表わす概念と、それらを表現する名詞との対応が日本語と英語とで異なる場合を示している。このような場合、訳語を決める際に行為の過程と過程によって得られる結果との区別を明確にする必要がある。

行為の過程	建造	construction
過程によって得られる結果	建造物	
行為の過程	料理	cooking
過程によって得られる結果		dish

(b) 単語の非標準的置換・難訳語

文中に置かれた単語には普通辞書に記述されている訳語をそのまま割り当てたのでは不自然な翻訳文しか得られない場合がある。例えば次の例では"write"をそのまま「書く」と訳しては日本語として不自然になるため「考える」と訳されている。

Most algorithms for solving computational problems were written with the assumption that they would be carried out by a human being.

計算問題を解くためのアルゴリズムは人間によって実行されるという仮定のもとに考えられた。

同様に次の例では"be covered with"をそのまま「覆われる」と訳すと不自然になることが示されている。

The branches of apple trees are almost covered with blossoms.

(?)りんごの木の枝はほとんど花に覆われています。

りんごの木の枝はいっぱい花をつけています。

また次の例文中の"serve"には日本語として適切な訳語を当てるのが難しく、場合場合で異なった訳語が割り当てられる。

For human beings this formula serves as a sufficiently precise algorithm for computing quadratic roots.

人間にとってはこの公式は二次方程式を計算するのに十分正確なアルゴリズムとして通用する。

(c) 慣用表現

慣用表現は言語によって異なるため、それを構成する単語ごとではなく、句全体をひとまとまりとして翻訳する必要がある。

Freeing the programmer from concern with ~

プログラマーは、もはや ~などを心配しなくてもよくなり

(d) 無生物主語

次の例に示すように英語では無生物主語はかなり多用されるが、日本語では行為を表わす動詞に対して能動態表現で無生物の主語を用いることはほとんどない。従って、このような英語文を日本語文に翻訳する場合には格構造を変換する、態を変える等の変換を行って訳す必要がある。多くの場合英語文中の無生物主語は日本語文では例のように道具格あるいは場所格として表現される。

A natural language can express a wide range of concepts.

自然言語によって広範囲の概念を表現することができる。

All programming languages include operations for assigning values of variables.

すべてのプログラミング言語には、変数に値を割り当てるための操作が含まれている。

(e) 言語による冗長度の違い

言語によって表現の冗長度が異なり、原文をそのまま翻訳すると不自然になることがある。例えば日本語では「手で引っ張る」「手でさわる」「足で蹴る」等の表現を用いるが、英語ではこれらの動作は手あるいは足で行うことが分かりきっているとされ、単に "draw", "touch", "kick" と表現される。従って次の例に示されるように翻訳の際にはそれぞれの言語の表現の冗長度に合わせて適宜語句の省略あるいは追加を行う必要がある。

ベルを鳴らすには手で触れればよい。

You only need to touch the bell for it to ring.

(f) 言いまわしの違い

英語と日本語との言いまわしの違いによって直訳では不自然な訳となる場合がある。次の例は、英文を直訳して得られる表現が日本語として不自然であり、言いまわしを変えて翻訳する必要がある場合を示している。

He is a good speaker of English.

(?)彼は英語の良い話し手です。

彼は英語を話すのがうまい。

彼は英語がうまい。

(g) 品詞の非対応による文構造変化

自然な翻訳を目指す場合、常に名詞を名詞、動詞を動詞にと品詞を保存して訳すよりは、品詞にとらわれずに訳した方がよい場合がある。例えば次例に示されるように名詞が動詞あるいは逆に動詞が名詞に翻訳される場合には節構造の変化を伴う。

Since the introduction of the first compilers, ~ .

最初のコンパイラが開発されて以来、~ .

A compiler can be designed to emphasize speed of computation ~

コンパイラの設計で重視されるのは ~ における計算スピードである。

また、次の例のように英語の前置詞に対して日本語の動詞の対応する場合もある。

He is against the plan.
彼はその計画に反対している。

(h) 関係節の翻訳

英語の関係代名詞に導かれる関係節は日本語では連体修飾句に対応するが、名詞を修飾する部分があまり長くなると日本語として不自然となる。

The man whose wife is an American can't speak English.

- (1) 妻がアメリカ人であるその人は英語が話せない。
あの人は奥さんがアメリカ人だが英語が話せない。

(i) 主語の省略

日本語では主語の省略が可能なので、次の例に示されるように英語文の受動態を能動態に変形して訳すことができる。逆に主語の省略された日本語文を英語に翻訳する際には省略された主語を復元することの必要な場合もある。

Symbol $\sqrt{\quad}$ must be put into different form.

平方根記号 $\sqrt{\quad}$ は異なった形に直しておかなければならない。

また、日本語の複文構造では、主節と従属節の主語が一致する構造にし、一方を省略すると自然な表現が得られることが多い。

As he often tells a lie, people can't trust him.

- (1) 彼はよく嘘をつくので、人々は彼を信用できない。
彼はよく嘘をつくので、ひとに信用されない。

(j) 文体

文体に関して留意すべき点としては、まず翻訳文中での訳語の統一が挙げられる。また、日本語では通常動詞が文末にくるため、「～です。」「～ます。」「～である。」の形が頻出し、このため文末が常に同じ形になると文章が大変読みづらくなる。

(k) 叙述の流れ

叙述の流れを自然にするために、複文の翻訳では主節と従属節との順序を工夫する必要がある。次の例は叙述の流れを時間の流れと一致させ、旧情報から新情報へという叙述の方向を維持するために "before" 以下の従属節を主節の後になるように翻訳した例である。

So, they were married. Presently, God sent them a baby girl
— a boy came, too — and seven happy years passed
before misfortune befell them.

ということでふたりは結婚した。やがて女の赤ん坊をさずかり
— 男の子も生まれ — 7年間幸福に暮らしたのだが、そこで
とんだ災難が降ってわいた。

4. 機械翻訳処理に取り入れる際の問題点

前章で述べた動的等価を目指した翻訳の種々の技術は、それぞれ機械翻訳処理に取り入れようとする場合、難しさの程度においてもまた取り入れる必要性においても様々であり様ではないが、例えば無生物主語に伴う格構造の変換等は比較的容易に取り入れられ、自然な翻訳を得るためにも効果的であると考えられる。

これらの技術を機械翻訳処理に取り入れる際の問題点は、従来人手による翻訳で用いられていた技術を機械翻訳の意味構造変換処理に用いる規則の形で明確に記述するという規則の記述の問題と、そのようにして記述された規則を実際の処理で適用する場合の規則適用の制御の問題との二点に分けられる。

従来翻訳の技術の記述は人間による翻訳を前提としていたため、直観的であり、機械で取り扱えるような明確な規則の形とはなっていない。従ってそれらをより明確化し、規則化する必要がある。この場合、規則を機械で取り扱うためには規則の内容のみならず、その適用条件をも合わせて記述する必要がある。

次に、規則の数が増加すると実際に処理を行う場合にはそれらの規則の適用を制御する方法が問題となる。規則の増加に伴う処理手順の可能な組み合わせの増加を防ぐためには、規則をいくつかのグループにまとめる方法が有効と考えられる。このような観点から、意味構造変換処理に用いる変換規則を、単語・単文・複文・文脈のそれぞれのレベルごとにまとめることが有効であろう。前章で示した翻訳の諸技術のうち、(a)~(c)は単語、(d)~(f)は単文、(g)~(i)は複文、(j)~(k)は文脈のレベルにほぼ対応している。

また、規則の記述と適用の制御との両者に関連するか、良質な翻訳文を得るためには、原言語解析によって得られた原言語意味構造から直接意味構造変換を行うばかりでなく、場合によっては解析によって得られた意味表現を一度意味的に

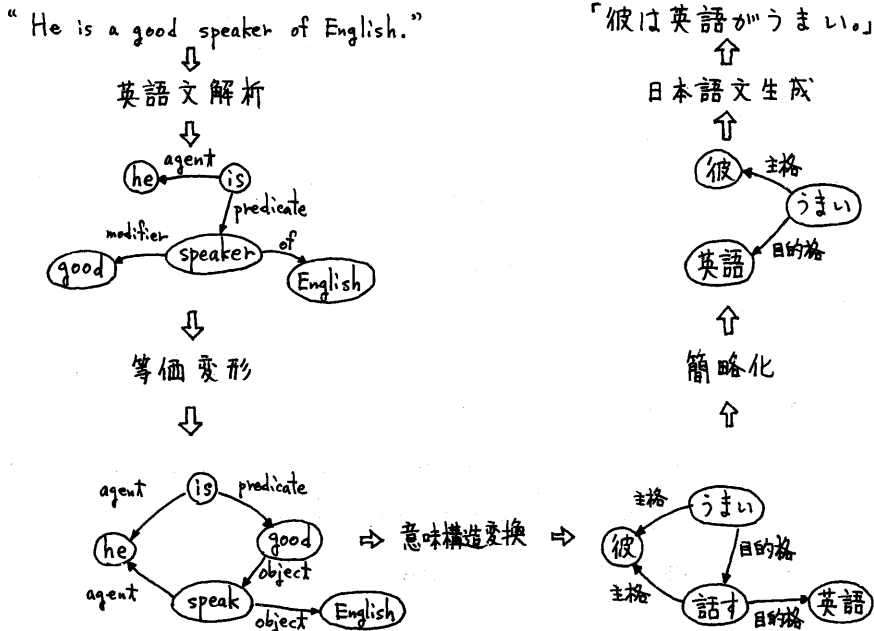


図3. 意味構造変換の前後処理.

等価な形に変形し、再解釈を施してから意味構造変換を行うという処理も必要になる。例えば、前章の(f)に述べた言いまわしの違いによる変形の例に対しては、図3に示すように英語文解析によって得られる意味表現を一度等価な形に変形してから意味構造変換を行うことによって動的等価に対応する日本語翻訳文を得ることが可能となる。このように意味構造変換において意味表現の再解釈を行うには当然意味内容に関連した種々の知識を参照することが必要となる。

5. 結び

構造変換方式による機械翻訳処理の観点から翻訳の過程に対して考察を加え、従来の人手による翻訳の技術、翻訳上の問題点を列挙し、それらを意味構造変換処理に取り入れる際の問題点を考察した。従来翻訳者の翻訳行為を規則の形で明確に記述することはほとんどなされていない。良質で実用的な機械翻訳を実現するにはこれらの規則化が望まれる。また、現状では機械翻訳システムの開発はどうしても trial and error となるざるを得ない。従ってシステムとしては拡張、変更の容易な柔軟な構成をとる必要がある。

参考文献

1. D. Burden, "Natural human languages automatically translated by computer: the SYSTRAN II system," Computer and People, pp. 12-15, 24, 1981.
2. Ch. Boitet and N. Nedobekine, "Russian-French at GETA: Outline of the method and detailed example," Proc. 8-th ICCL, pp. 437-446, 1980.
3. P. Isabelle, et al., "TAUM-AVIATION: Un système pour la traduction automatique de manuels techniques," Proc. 7-th ICCL, 1978.
4. M. Nagao, J. Tsujii, K. Mitamura, H. Hirakawa, and M. Kume, "A machine translation system from Japanese into English," Proc. 8-th ICCL, pp. 414-423, 1980.
5. M. King, "Design characteristics of a machine translation system," Proc. 7-th IJCAI, pp. 43-46, 1981.
6. J. G. Carbonell, R. E. Cullingford, and A. V. Gershman, "Steps toward knowledge-based machine translation," IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-3, pp. 376-392, 1981.
7. 内田裕士, 小部正人, 林達也, "日英機械翻訳システム ATLAS/U," 第23回情報処理学会全国大会講演論文集, pp. 817-818, 1981.
8. E. A. Nida, Toward a Science of Translating, Leiden, E. J. Brill, 1964.
(邦訳 成瀬武史訳, 翻訳学序説, 開文社, 1972.)
9. E. A. Nida, C. R. Taber, and N. S. Brannen. The Theory and Practice of Translation, Leiden, E. J. Brill, 1969.
(邦訳 沢登善仁, 竹川潔訳, 翻訳—理論と実際, 研究社, 1973.)
10. 別宮貞徳, 翻訳の初歩, Japan Times, 1980.