

## 日英翻訳システムALT-J/Eの設計思想

6J-1

池原悟、宮崎正弘、東田正信

NTT電気通信研究所

## 1. はじめに

構造言語学が言語の形式に着目したのに対して、変形生成文法は表現の内容に着目した立場を取っているが、対象を内容と捉えるにとどまり、本来一体であるべき内容と形式を分離した二元論的立場から言語現象の説明を行っている。これに基づく最近の機械翻訳の研究においても、言語間で内容(対象)が同一とする立場から言語共通の意味構造を仮定した翻訳方式が考えられているが、翻訳の品質を向上させるには原言語と目的言語の認識構造の違いを考慮する必要がある。本論文では、言語過程説の立場から言語による話者の認識の枠組の違いを考え、構文と意味を一体化した多段翻訳方式を提案する。

## 2. 言語の過程的構造と翻訳

## 2.1 従来の翻訳方式の基本的問題

言語は対象/認識/表現の3段階からなる過程的構造をもつ(文献1)。認識は直接的には個人に依存するが、広くは社会的背景に依存しており、それぞれの言語圏で自然発生的に成立した社会的規範と言うべき文法を介して表現に対応づけられる。従って認識は個人的なものであるがその枠組は社会的なものであり、日本語と英語の質的な差がここにある以上、翻訳においてもこの違いを考慮する事が必要と言える。これに対して生成変形文法では対象が直接表現に結びつく考えを取っており、従来の機械翻訳においても、対象が言語に依存しない点を捉えて、言語に共通の意味構造を仮定したピボット方式が提案されているが、言語のもつ認識の枠組みの違いを考えると、共通言語方式では翻訳の品質を向上させる事は困難であると言える。

## 2.2 主体と客体に関する認識

言語表現に表される認識を大別すると主体(話者)に関するものと客体に関するものに分かれるが、主体も客体化して捉えられる場合があるため、表現は一般に主体のあり方を直接表現する主体的表現と客体化された対象のあり方を示す客体的表現に分けることができる。日本語は謬着言語の特質として主体的表現は助詞、助動詞などの付属語で表され、客体的表現は名詞、動詞などの自立語で表されるため、それぞれ語としての対応関係が明らかであるが、英語は屈折言語の特質として、主体的表現が語の屈折の中に埋もれる事が多いため、日英の表現の意味を語の対応関係で捉える事は困難である。従って、日英翻訳においては、日本語に表された主体と客体のあり方を分離して解析し、まず客体に関して、認識の普遍性、個別性、および抽象化の程度を対応させた英語の表現を生成したのち、主体の客

体に対する時間的、空間的位置や態度などを反映した英文を生成していく事が必要と考える。

## 2.3 統語構造と意味構造の不可分性

主体的表現と客体的表現の解析を進める上で重要な点は統語構造の持つ意味の扱いである。概念構造変換方式は言語による意味表現の違いを認めた点でピボット方式より進んでいるが、本来不可分であるべき言語の形式と内容を分離し、深層と称する意味から表層と称する形式を導こうとしたもので、概念構造への変換の過程で表現のもつ意味が解体され失われていくため、目的言語で正しく再現できなくなる点に問題がある。言語の意味は表現とそれに結びつけられた認識との関係であり(文献2)、認識が構造をもつ以上、構造と意味は不可分の関係にある。古来、統語構造は語が文や句の中で占める位置やそこで果たす役割のこととされているが、それは対象世界の実体と属性の関係などが認識を介して抽象化され表現されたもので、本来の意味の側面と言える。言語は慣用的な表現の集まりと言っても良いとする考えは構造と意味の一体性を言ったものである。表層と異なる箇所意味を仮定する方式では構造のもつ意味の欠落を防ぐことはできない。従って、統語構造のもつ意味を見ないで、部分の意味から全体の意味を合成しようとする原子論(顕微鏡理論)的方法だけでは、係り結びや接続、慣用表現はもとより、比較的単純な日本語においても訳文の品質を上げる事は困難で、表現の意味を表現の構造と一体化して掘り取る事が必要と考えられる。

## 3. ALT-J/Eの翻訳方式

前節の議論をふまえ、本論文では

- ①主体と客体に関する認識構造の抽出
- ②意味と構造の一体的扱い
- ③解析と変換の融合

の3点を特徴とする図1、2の多段翻訳方式を提案する。

## 3.1 日本語解析(文献3)

(1)主体的表現と客体的表現の分離と平文の生成

人力文は形態素解析の結果にもとづく係り受け構造の解析によって用言の支配範囲を決定し、基本的に一つの用言に支配される範囲の表現を単文として抽出する。単文は、主体的表現と客体的表現に分けて解析される。前者では、助動詞を中心に様相(法)、時制が解析されるが、英語では活用語尾によってアスペクト、態も表現されることを考え、ここで両者の解析も行う。副助詞で示される客体的取り上げ方、その他、様相と時制が重った形式名詞を含む表現などからの情報も抽出する。これら主体的表現によって

Design Concept for Japanese to English Translation System ALT-J/E

Satoru IKEHARA, Masahiro MIYAZAKI, Masanobu HIGASHIDA

NTT Electrical Communications Laboratories

表された話者の認識情報を取り出したあとの文を平文と称し、客体的表現として扱う。

(2) 日本文の等価的変換

平文には自立語の他に本来主体的表現の一部である格助詞及び格助詞相当語が含まれる。これは客体のあり方に関する話者の認識を含めて文意を汲み取り英語に変換する事を狙った為であり、平文が英語への変換の基本単位となる。ところで、日本語の格助詞の多彩な使い分けによって表現される話者認識の微妙さに対応する英語の表現を見出す事が困難な場合も多い。そこで、日本文の等価的変換として、日本文と英文の文型パターンを参照しながら、英文として訳し分けられる範囲に格助詞及び格助詞相当語の用法を縮退させる。

3.2 日英の表現の変換 (文献4)

(1) 主体的表現の対応づけ

様相、時制に関する解析結果(現在各々約50種、40種に分類される)に英語の様相と時制を対応させる。副助詞の示す情報は様態を示す情報として抽出されており、英語の様態と対応させる。その他、否定、比較、態、命令、使役、感嘆などに関する情報はそのまま、英文生成側へ持ち越させる。

(2) 客体的表現(平文)の多段変換

文の構造の持つ意味を含めて文意を掘り取って翻訳するには、慣用的な表現を一まとまりとして抽出して英文に対応させた(慣用表現変換)のち、意味を担う単位をより大きな単位で捉えて英文に対応させる事が望ましい。そこで、日本文の文型を個々の用言の用法に着目して英語文型に対応させた意味的結合値パターン変換規則を用いて単位文のもつ上位の構造から順に英文への変換を行う。単位文は最上位の述語が単一の構造の文を言う。以上の2段階の変換

からもれた表現の変換を行うため第3の変換として、汎用の文型パターンを用いた汎用表現変換の方式を組み込んでいる。この変換を受けた表現はきめが粗く、より縮退した英文となるため、翻訳の品質は前2段階の変換の品質に負うところが多い。前2段階での訳し分けのきめ細かさは適応するパターン数、及びパターンに結びつけられる体言の意味属性の分解能に依存するため、本システムでは慣用表現と結合値パターンについて各々1万件以上と、それを体言に結びつける約2800種の単語意味属性からなる意味属性体系を準備しており、パターン数はさらに増加させることにより翻訳の品質をさらに向上させる予定である。本方式ではパターンは表現に対して個別的に準備していく事が可能であり、パターンのデバッグと増強に当たって他のパターンとの相互矛盾を考える必要が殆ど無いため、システムのチューンアップが容易である。

3.3 英文生成(文献5)

(1) 客体的表現(素文)の生成

前述の3段階の変換によって得られた単位文に対する英文構造を組み合わせて平文に対応する英文(素文と言う)を生成する。訳語は個別パターンのもつ意味属性の制約条件と名詞句のもつ意味的係り受け構造(上述の意味属性体系が用いられる)などから絞り込まれる。

(2) 素文に対する主体的表現の付加

各素文に対してまず使役、受身、命令、3人称依頼、勧誘などの様相に対応する文の枠組みが決められ、英語基本文型が作成される。英語基本文型はその後その他の様相、時制など、対応する英語側の概念に基づいて変形され、英文が生成される。英語側の変形の処理としては否定変形、不定詞挿入、時制処理、助動詞挿入、副詞挿入、疑問変形などがあり、様相、時制の重なりなどの処理もこれらの処理を複合させて実行される。

4. あとがき

ALT-J/Eの設計思想とそれに基づく翻訳方式として、主体的表現と客体的表現の個別的扱いと3段階の変換からなる多段変換方式の基本的考え方を述べた。本システムは現在第一次の評価実験を終え、方式全体の強化と不足機能の追加を行っているが、本論文で述べた基本方式により、翻訳の品質が従来の方式に比べて向上する事が確認されている。

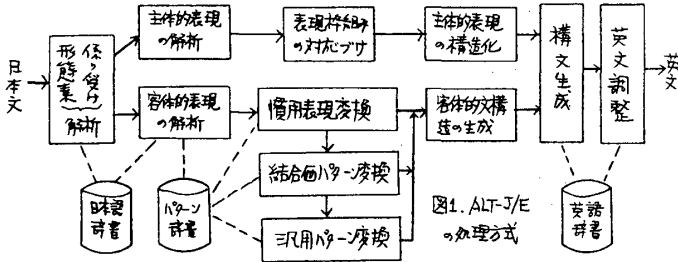


図1. ALT-J/Eの処理方式

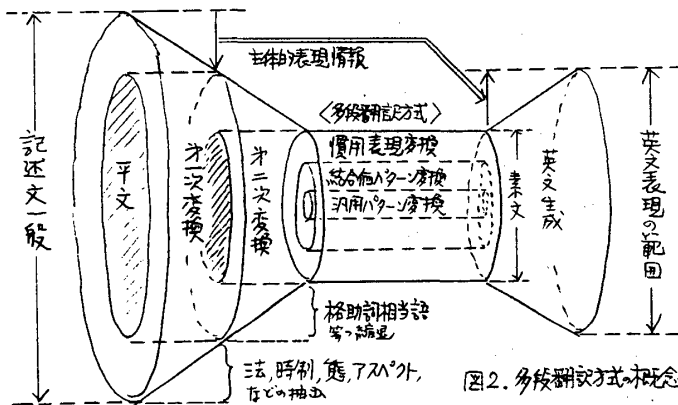


図2. 多段階翻訳方式の概念

<参考文献>

- (1) 時枝誠記「国語学原論」岩波書店、昭16
- (2) 三浦つとむ、「認識と言語の理論I~III」草野書房、1967
- (3) 宮崎他、日英翻訳システムALT-J/Eにおける日本語解析技術、第33回情処全大(1986)
- (4) 林他、日英翻訳システムALT-J/Eにおける日本語解析技術、第33回情処全大(1986)
- (5) 岡本他、日英翻訳システムALT-J/Eにおける英文生成技術、第33回情処全大(1986)