

機械翻訳における中間表現としての認識構造の枠組み

6Y-1

柴田 悟之 宮崎 正弘

新潟大学大学院自然科学研究科

1 はじめに

認識構造は、時枝誠記の言語過程説を発展的に継承した三浦つとむ [1] の説に基づき話者の把握した対象世界を表現したものである。本稿では機械翻訳における中間表現として用いる認識構造の枠組みについて提案し、機械翻訳における有効性について論じる。

2 認識構造について

言語の生成において、話者（書き手）は対象の客観的な姿を概念として捉える（客体的認識）と共に、対象に対する主体的判断（主体的認識）を重ね合わせる。それにより、話者は頭の中で対象のイメージ（認識）を作り上げる。このような認識を構造化したものが認識構造である。

認識構造は客体的表現と主体的表現から構成されている。認識構造が各言語で共通であるとすれば、機械翻訳にとって都合がよい。しかし実際には対象の捉え方が各言語で異なっているため、各言語の認識構造は異なっている。特に客体的表現部分はそれぞれの言語が様々な固有な表現を持ち、その内容も実に多様である。一方、主体的表現部分は、表層的な表現は多様であるが、その内容は話者の主観的判断や感情に関する情報であり、極めて限定されている。したがって、認識構造においては、主体表現の方が客体表現よりも各言語の言語規範の影響が比較的大きくないので、形式の統一を図り易い。

3 中間表現の枠組み

本稿では、日英翻訳を例として挙げる。

A Framework of Recognition Structure
for Machine Translation System
Noriyuki Shibata, Masahiro Miyazaki
Niigata University

3.1 客体的表現

客体的表現は、対象の客観的な姿を概念化したものを表し、各言語により異なる形式で表される。また英語ならば前置詞や冠詞、日本語ならば助詞といった、実体に対する話者の捉え方、つまり主体的表現を表す語も、客体的表現の構成要素となる。ここでは英語文の認識構造記述 [2] で使用されるものを例として示す。

- 動的属性
実体の動作を表す。
- 静的属性
対象の状態・性質を表す。
- 対象__ Agent
話者が注目している、もしくは動作を行う対象の実体。
- 対象__ Object
動的属性の効果が及ぼす目標。
- 推移__ Agent
動的属性の効果が及ぼした後の対象の姿。
- 受益者
動的属性の効果が及ぼした後の対象__ Object の移動先。
- 時間属性
対象、または各属性の存在する時間の情報。
- 場所属性
対象、または各属性の存在する場所の情報。

3.2 主体的表現

多様な表層的表現で表された話者の主体的な判断を表す主体的表現を以下の抽象化された少数の素性を 1

個以上組み合わせることによって形式的に記述する。主体的表現は原言語、目的言語を通じて共通のものとして定義し、かつ翻訳に必要な情報を出来る限り単純な素性のセットに押さえることにより、解析生成処理の便宜を図る。図1に英語文の認識構造の例を示す。将来的には本稿の枠組みに基づき再構築を目指す。

- 判断（肯定／否定）
客体的表現に対する表現主体の直接的な意志判断。
- 時制（現在／過去／未来／不明）
表現主体が表現の焦点をどの時間においているかを表す情報。
- 様相（疑問／命令／推量／宣言／義務／許可／可能／希望／丁寧）
各種機能語に反応して意味を追加する。
- 態（能動／受動／使役／被使役）
主体となんらかの作用との関係を表す。
- 相（未完了／完了／進行／持続／開始／反復／経験）
動作や状態の様子を表す。
- 実体把握（総称／特定／不定など）
客体表現部分の構成要素が表す主体表現であり、実体の捉え方を表す。

1- 認識世界

- |- 客体認識
 - | |- 動作主
 - | | |- 実体把握 -- < 特定 >
 - | | |- 実体概念 -- < boy >
 - | |- 動的属性 -- < eat >
 - | |- 対象
 - | | |- 実体把握 -- < 不定 >
 - | | |- 実体概念 -- < cake >
- |- 主体判断
 - | - 時制 -- < 現在 >
 - | - 判断 -- < 肯定 >
 - | - 態 -- < 能動 >

図1: The boy eats a cake の認識構造の例

4 翻訳のながれ

原言語文を解析し、認識構造を抽出する。各言語間の差異が比較的少ない認識構造の主体的表現部分は原言語、目的言語に共通のものとしてあるため、そのまま目的言語側にわたす。一方各言語間で形式の異なる客体的表現部分は、両言語対間の文型パターン対辞書および用例対を用いて目的言語に変換する。変換された目的言語の客体的表現部分と、原言語からそのままわたされた主体表現部分を融合して、目的言語の認識構造を構成する。こうしてできあがった目的言語の認識構造を用いて目的言語文を生成する。図2に変換処理の簡単な流れを示す。

このような処理を行うことにより、冠詞や助詞などの微妙なニュアンスを翻訳に反映させることができると考えられる。

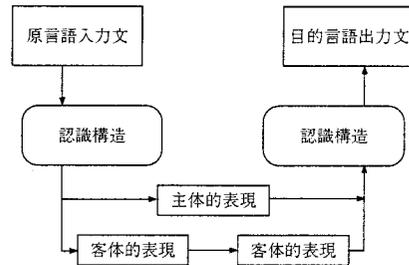


図2: 変換処理の流れ

5 おわりに

本稿では、機械翻訳における中間表現として用いる認識構造の枠組みと、機械翻訳における有効性について論じた。今後、情報の過不足や、複雑な文に対応できるように認識構造の枠組みを洗練する。本稿に示された認識構造の枠組み中間表現として用いた翻訳システムの設計の試作を進めている。

参考文献

- [1] 三浦つとむ:日本語とはどういう言語か、講談社学術文庫(1976).
- [2] 高草木、宮崎:SGLR-plusによる話者の認識構造を抽出する英語文パーザの試作、情報処理学会第58回全国大会、No.1E-8(1999).