

中間概念として認識構造を用いた多言語翻訳方式

川辺 諭 宮崎正弘

新潟大学大学院自然科学研究科

1 はじめに

時枝誠記の言語理論を発展的に継承した三浦つとむの文法理論では、人間の言語活動は、「対象-認識-表現」といった過程的な構造によって運用される(言語過程説)としている。本方式では多言語間機械翻訳機構における中間概念として、この認識構造を採用する。認識構造の記述体系を、特定の言語に依存しないものとして準備することにより、トランスファー方式のように特定の言語対の翻訳機構をいちいち実装する必要がないといった利点がある。また、それぞれの言語が持つ特有の言い回しに対応するため、認識構造をほぼ同等の表現内容を持つ、似通った認識構造へ変換するリボルバー変換機構を準備する。リボルバー変換機構を用いることにより、より自然な目標言語表現の生成が可能となる。なお本機構の実装には、論理型言語である prolog を用いている。

2 認識構造

本節では多言語翻訳方式において、言語同士を媒介する認識構造の記述形式について述べる。

2.1 属性

対象世界における実体の持つ属性であり、[属性, 属性値]といった形式で記述される。知覚属性や抽象属性は、形容詞や、数詞や冠詞などの限定詞を用いて表現される。また、実体同士のさまざまな関係を表現する関係属性は、言語表現では前置詞、後置詞といった機能語などで表現される。

2.2 実体

実体は属性の集合として辞書に記述される。また記述量を削減するために、実体同士に上位下位といった関係を記述している。

```
% 実体辞書での記述例
% subst(クラス, 属性リスト).
subst(c_orange, [[color, orange], [form, ball]]).
% supsub(上位クラス, 下位クラスリスト).
supsub(c_fruit, [c_orange, c_orange, c_lemon]).
```

認識構造における実体は、対象世界における事象をどのように捉えるかといった話者の判断が加わっている。これらの実体は、[クラス|付加属性]といった形式で記述される。

```
% 認識構造中での実体の記述例
[c_orange, [taste, sweet], [size, big]]
```

% 対応する言語表現 (参考)

‘‘a sweet big orange’’ 「甘くて大きなオレンジ」

2.3 グリスト

グリストは、格要素のリストである。格要素としては動詞、主格、対格や、時間格、場所格、道具格といったものを準備している。

% グリストの例

```
[verb,verb,v_eat],
[*sb1,agt,c_person_he],
[*sb2,obj,[c_orange,[taste,sweet],[size,big]]]
```

% 対応する言語表現 (参考)

‘‘he eats a sweet big orange’’ 「彼は甘くて大きなオレンジを食べる」

2.4 フレーム属性

フレーム属性は、格フレーム自体が持つ時制や様相などといった性質である (表 1)。

表 1: フレーム属性

フレーム属性	属性値	内容
機能	叙述, 疑問, 命令など	談話中で格フレーム持つ機能
時制	現在, 過去, 未来	事象が生じた時間
定言	肯定, 否定	事象に対する話者の判断
概言	数値	話者が判断した事象の生起確率
相	開始, 進行, 完了	時の流れに対する事象のあり方
社会	可能, 必要, 許可など	主に主格の社会的属性の表現

2.5 認識構造を用いた翻訳

以下は、認識構造から英語、日本語、中国語文が生成される様子である。

```
@recog,
  [@caseframe,1,
    [<caselist>,
      [*verb,verb,v_watch],      英語 : I watched the movie.
      [*sb1,c_personI],          ---→ 日本語 : わたしは映画をみました.
      [*sb2,c_movie]],           中国語 : 我看了电影.
    [<fattrblist>,
      [func,declare],
      [time,past]]].
```

本方式ではこのような認識構造を中間概念とし、各言語の生成部、解析部を作成する。

3 リボルバー変換機構

言語表現においては、言語や状況、話者によって対象となる事象の捉え方の違いがあるため、ひとつの事象から異なった認識構造が生成されることが少なくない。本方式ではこのような認識構造の揺れに対応するために、あらかじめ類似の認識構造同士を変換するためのテンプレートを準備しておき、言語や状況などに応じて認識構造の変換を行なう。この機構をリボルバー変換機構と呼んでいる。

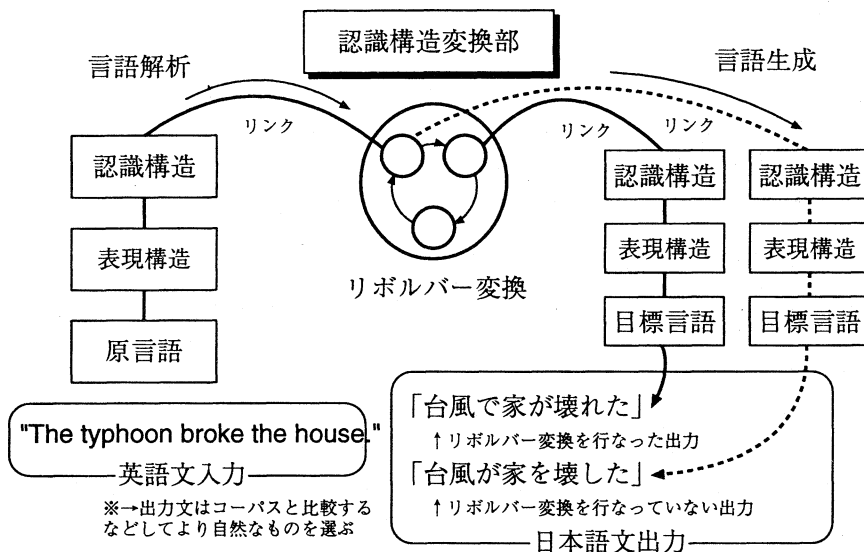


図 1: リボルバー変換機構

例えば “The typhoon broke the house.” という英語表現は “The typhoon” という無生物が主語（動作主）となっている。これに対して「台風で家が壊れた」という日本語表現は、「台風」が原因格となっている。＜台風で家が壊れる＞といった事象に対して、それぞれの言語特有の捉え方（認識）を行なっており、複数の認識構造が作成されてしまい、認識構造を仲介した多言語翻訳に支障をきたす。リボルバー変換を行なうことにより、こういった認識構造の揺れに柔軟に対応する。（図1）。

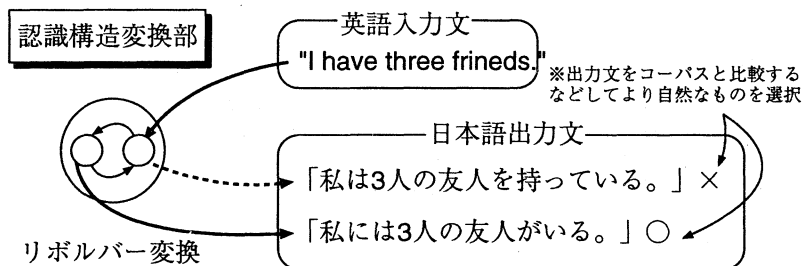


図 2: 存在と所有表現の変換

また、日本語表現における「私には3人の友人がいる」のような人などの存在を表す表現は、英語表現では“I have three friends.”というように所有を表す動詞 have を用いる。このような場合は、動詞データとリボルバー変換テンプレートとをリンクしておき、対格となっている実体(名詞)の属性(この場合は<人>であるか)を判断してリボルバー変換を行う(図2)。変換の結果、複数の文が生成される場合は、目的言語において既存のコーパスと比較するなどしてより自然なものを選択することを考えている。

この他認識構造の大域的な変換としては、「<対格>の<サ変動詞>をする」といった日本語表現を“<動詞><対格>”といった英語表現に変換するサ変動詞変換や、「<対格><動詞><否定様相>」といった日本語表現を“<動詞> no <対格>”といった英語表現に変換する no-名詞変換などを準備する。

また、名詞句変換などの句や語といった局所的なレベルでの変換も行なう。下の例は英語表現では形容詞相当で扱われている“three”といった数詞が、日本語表現では独立した格要素として扱われる場合である。

‘I have three books.’	↔	「私は本を三冊持っている」
[*verb,verb,v_buy],		[*verb,verb,v_buy],
[*sb1,agt,c_personI],		[*sb1,agt_personI],
[*sb2,obj,c_book,[num,3]]		[*sb2,obj,c_book],
		[*sb3,num,c_3]

本方式では、このようなりボルバー変換機構を統語レベルに応じて準備する。リボルバー変換を行なうことにより、こなれた目標言語表現を生成することが可能となる。

4 おわりに

認識構造を中間概念とした多言語翻訳機構を提案した。中間概念を媒介することにより、特定の言語対のための翻訳機構をいちいち準備する必要がないといった利点がある。また、リボルバー変換機構を準備することにより、より自然な言語表現の解析、生成が可能となる。

現在本方式を用いた翻訳機構の実装を進めており、今後の課題としては、

- 各言語の解析、生成部の実装
- 中間概念辞書、各言語辞書の拡充
- リボルバー変換テンプレートの拡充
- リボルバー変換機構のアルゴリズムの検討

などがあげられる。

参考文献

- [1] 時枝誠記: “日本語文法口語篇” 岩波全書 (1950).
- [2] 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).