

日英機械翻訳における差分概念に着目した 訳語選択用情報の獲得法

3W-3

杉浦徹哉

宮崎正弘

新潟大学大学院自然科学研究科

1 はじめに

機械翻訳において原文の意味・ニュアンスを正しく伝え、文脈や状況にふさわしい訳語をいかに決定するかが大きな問題となっている。

従来、ある名詞に対して複数の訳語候補があり目的言語において同一の意味属性を持つような語の訳語選択において、名詞自身の意味的特徴の準備がなされていない場合が多い。

本稿では、そのような名詞を訳出する際、目的言語において同一の意味属性を持つ複数の訳語候補間において、その概念の差分から訳語選択用情報を獲得する方法を提案し、その有効性を論じる。

2 名詞の意味属性

本稿で用いられる「意味属性」とは、日英機械翻訳システム ALT-J/E[1](以下、ALT-J/E) で用いられている「一般名詞意味属性」[2]である。ここで意味属性は、意味属性番号により約 2700 の属性に分類され、その単語が一般名詞的に使用される際の用法を、12 段からなる木構造として体系的にまとめられている。

翻訳において意味属性を用いることで、文章中の名詞がどのような語義を持つかを推定することができ、意味的な側面から訳語の選択を可能とする。また、ALT-J/E の格フレーム情報などを併用

することで、より精度の高い意味属性を推定、もしくは決定することができる。

3 対象とする名詞

3.1 意味属性の一致する名詞

翻訳対象の名詞に対して一意に意味属性が決定しても訳語選択ができない場合がある。表 1 に示すような、複数の訳語候補に同一の意味属性が記述されているときなどである。このような名詞が ALT-J/E の日英対照辞書では約 1500 語みられるが、名詞の語義が共通の意味属性に決定されたとき、訳語選択が失敗するといったケースが機械翻訳でしばしばみられる。

表 1: 訳語が共通意味属性をもつ名詞

見出し語	訳語	意味属性 [番号]
権利	claim	権利 [1166]
	right	権利 [1166]
底	bottom	底 [2625]
	bed	底, 川 [2625,496]
卵	egg	卵, 肉・卵 [670,843]
	roe	卵, 魚介類 [670,842]

3.2 差分カテゴリを持つ名詞

ここでは、訳語候補間で同一の意味属性を持つ名詞の中で、共通する意味属性以外の差分カテゴリをもつ名詞 400 語を対象とした。表 1 では「卵」のような意味属性番号を持つ語である。見出し語「卵」は 670 番を共通の意味属性として持ち、訳

語「egg」「roe」それぞれに差分カテゴリ 843 番と 842 番を持つ。これらの差分カテゴリを利用することで訳語候補間の語の特性を導き、名詞の訳語選択用情報を得る方法を提案する。

4 訳語選択情報の獲得

4.1 獲得法

差分カテゴリから訳語選択用情報を獲得する手法を例をまじえて述べる。

上述した一般名詞「卵」は共通の意味属性「卵」を持つ。ここで差分カテゴリの情報を日本語語彙大系 [2] から抽出すると表 2 のようになる。

表 2: 「卵」の差分カテゴリ

訳語	意味属性 [番号]	意味属性を持つ単語
egg	肉・卵 [843]	うずら、鶏...
roe	魚介類 [842]	魚、鮎、鮭...

この表から訳語「egg」は鳥類や獣をさす語の場合に選択され、「roe」は見出し語が魚介類であるときに有効であることが判断できる。このように原文中での語彙が共通の意味属性で解釈されたときでも、文脈解析などで見出し語がさす対象とこの情報とを比較することで、その訳語を正しく選出することが可能となる。

また木構造からなる意味属性体系から情報を得ることができる。共通意味属性「害」をもつ見出し語「害」の差分カテゴリは、訳語「harm」が 743 番、「damage」が 1180 番である。この関係を図 1 に示す。

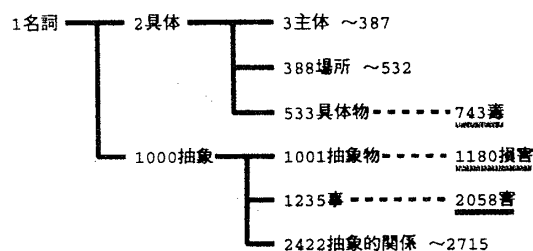


図 1: 意味属性体系

この図から、「harm」の差分カテゴリは「具体」の下位概念であるのに対し、「damage」は「害」と同じ「抽象」の下位に属する。このことから、見出し語「害」が意味属性「害」で解釈されたならば、差分カテゴリの意味的近さから訳語「damage」が優先度が高いことが推測できる。

4.2 評価

これまで訳語選択が失敗していた差分カテゴリを持つ名詞 400 語に対し、対訳例文中の名詞を対象として訳語選択を机上で行なったところ、およそ 50 % の語に有効な差分情報を準備することができた。

5 おわりに

訳語候補間において、共通する意味属性および他の意味属性を持つ一般名詞において、その差分概念から訳語候補に訳語選択用情報を抽出し、語の訳し分けおよび文脈解析への足掛かりとする手法を提案した。

今後の課題として、これらの情報を利用した訳語選択機構を、現在試作中のマルチコーパスを用いた日英多段用例翻訳システム [3] に組み込み、その有効性について評価すること。本稿で対象外とした意味属性の一致する名詞の扱いなどがある。

謝辞

日英機械翻訳システム ALT-J/E の日英対照辞書、および名詞意味属性体系データを提供して下さいた NTT コミュニケーション科学研究所の関係各位に深謝する。

参考文献

- [1] 池原, 宮崎, 横尾: “ALT-J/E における翻訳辞書の構成と意味解析型翻訳機能”, 電子情報通信学会春季全国大会, No.SD-5-2(1991).
- [2] 池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林: “日本語語彙大系 (全 5 巻)”, 岩波書店 (1997).
- [3] 池田, 宮崎: “マルチコーパスを利用した多段用例翻訳方式”, 情報処理学会第 53 回全国大会, 4L-9(1996).