



図3 「払う」の関連語

3. 2. 比較照合実験

実験には、対訳用例データベースとして、市販の英和辞典から抽出した対訳例文(約1.9万文)を利用した。これらの例文は比較的短い文が多く、訳語照合をしやすいだけでなく、語彙的にも、用法的にも広範囲にわたっており、システムの基本チェックを行なうのには向いていると考えられる。

この対訳用例データの英語テキストを英語解析、英日変換した段階で、第3文型(SVO型)と判断された動詞(対象語との共起関係も調査する目的から対象格に名詞[代名詞も含む]をとる場合に限定した)の訳語について訳例テキストと比較照合実験を行なった。今回の実験では、9544の該当事例を検出し、訳例と同様に訳されていると判断された割合(完全一致、異表記・同意語・類義語で照合できたものの割合)は、32.2%(3076/9544例)であった。

また、この結果を、動詞の見出しごとに整理してみると、見出しによって大きな差が生じていることが明らかになった(例えば、"leave"は、訳例と同様に訳されている割合が7.6%(5/61例)であったのに対して、"know"は、83.0%(54/65例)であった)。

4. 不一致表現の分析

訳語の比較照合で、翻訳例とは異なった訳し方がなされた事例については、その原因を分析する必要がある。訳語の不一致表現を自動的に分析する手法は確立していないが、不一致と判断される原因として、

- (a) 翻訳辞書データの不備
 - (共起関係辞書データの不備など)
- (b) 関連語辞書の不備
- (c) 英語解析の失敗(格パターン認識の失敗など)
- (d) 日本語形態素解析の失敗(未知語の存在など)
- (e) 意識されている場合
- (f) 英語と日本語で格依存構造的に大きな違いがある場合

などが考えられる。

ちなみに"take"について不一致表現141例(出現数204)を分析した結果、表1のようになった。

この分析結果のうち、関連語辞書の不備については、辞書データを整備することによって不一致事例から除外することができる。また、翻訳辞書の不備に関しては、次節で述べるように翻訳辞書を更新する必要がある。以上のケースについては、辞書の更新を続けることによって、不一致と判断される用例の割合は減少させることができる。

□ 翻訳辞書の不備	4 6
● 対象格との共起関係情報の不備	1 8
● 副詞によって影響	1 4
● 後方の前置詞句によって影響	1 3
□ 関連語辞書の不備	2 4
□ 日本語形態素解析の失敗(未知語の存在、接続不可の問題など)	6
□ 英語解析の失敗(格パターンの認識の失敗など)	4
□ 代名詞の照応	3
Shake the medicine well before taking it.	
□ 意識および構造が大きく違う	1 7
□ その他	4 1

表1 "take"の不一致表現の分析結果

この段階でも不一致表現と判断された表現に対しては、別の角度からの検討(例えば、辞書データ構造や翻訳手法の再検討)が必要な用例が多く含まれており、その事例データとして利用することが考えられる。

5. 翻訳辞書のチューニング

訳例と異なった訳し方をしている例文について、動詞と対象格の単語の間で共起関係が存在するかどうかを調べるために、動詞と対象格の単語のペア(英語と日本語)で見出しごとに整理してみた。共起関係が存在するかどうかは、人間が判断せざるをえないが、不自然な訳し方がされている用例は、動詞と対象格の単語のペアからだけでも(文全体を読まなくても)効率的に検出できる。この結果を容易に共起関係辞書に反映させるために辞書登録用のデータフォーマットを定め、効率的に翻訳辞書を更新できるようにした。

6. おわりに

対訳用例を利用することによって、翻訳辞書の検証およびチューニングを行なう試みについて述べた。これらの処理によって、共起関係を含めて辞書データの信頼性を向上させることが可能になった。

今後は、さらに、

- 広範囲の分野で、しかも、大量の対訳テキストの収集
 - を行ない、翻訳辞書の更新を図るとともに、より自然な訳文を生成するために、
 - 辞書データ構造の再検討
 - 翻訳方式の複合化
- などについても検討していきたいと考えている。

[参考文献]

[1] 佐藤, 長尾: 実例に基づいた翻訳, 情報処理学会研究報告, NL-70-9(1989)
 [2] 野村, 田中(編): 機械翻訳, bit別冊, 共立出版(1988)