6Y-04 結合価文法による動詞の訳語選択能力の評価

金出地真人　池原悟　村上仁一
鳥取大学大学院工学研究科

1．はじめに

機械翻訳において複数の訳語を持つ語が，翻訳の精度を下げる原因となっている．そこで結合価パターン(以下パターン) を用いて訳語を決定する結合価文法が開発されたが，その精度は定量的に示されていない．本研究では，動詞の訳語選択において，結合価文法がどの程度効果があるかを結合価文法を用いた翻訳ソフト ALT-J/E(以下 ALT) を用いて検証する．なお，評価対象の例文には IPAL の動詞に関する例文を用いる．

2．結合価文法による訳語選択

2.1 結合価文法

結合価文法は，用言と格要素 (名詞 + 助詞) の意味的関係を記述したものである．これにより用言と名詞との間に意味的な制約が生まれ，日本語解析で発生する意味上の多義が解消されると考えられる．本研究では，日本語語彙体系 [1] に記載されている「構文意味辞書」のパターンを使用する．

表 1: パターンの例 (かっこ内は一般名詞意味属性名)

N1(人)	が	N2(休暇)	を	送る	N1 spend N2
N1(人)	が	N2(生活)	を	送る	N1 live N2
N1(人)	が	N2(人)	を	起こす	N1 wake N2
N1(人)	が	N2(訴訟)	を	起こす	N1 file N2

2.2 一般名詞意味属性体系

パターンは，格要素の名詞の代わりに一般名詞意味属性で記述されている．一般名詞意味属性体系とは，単語の見方，捉え方に着目して，名詞の意味的用法を整理，体系化したシソーラスである．本研究で使用する一般名詞意味属性体系では，約 40 万語の名詞が，最大 12 段の木構造を構成する 2,710 の意味属性に分類されている．

Evaluation of English Word Translations for Japanese Verbs using Valency Grammar
Masato KANADECHI, Satoru IKEHARA, Jin'ichi MURAKAMI
Tottori University

3．評価実験

評価実験では IPAL の動詞に関する例文を ALT で翻訳し，翻訳結果を○ △ × の三段階で評価する．

3.1 評価基準

評価○　：ALT の英訳文の動詞が，対訳例が用いている動詞と一致する場合．

例 1：二つの川がこの地点で 合う。
対訳 1：The two rivers **join** at this point.
ALT の訳 1：Two rivers **join** in this point.

評価△　：ALT の英訳文と対訳例が異なる動詞を使用しているが，出力された英訳文が文法的に正しく，意味が通じる場合．

例 2：彼は準備を 急いだ。
対訳 2：He **prepares quickly**．
ALT の訳 2：He **hurried** preparation

評価×　：ALT の英訳文が間違っている場合．

3.2 実験結果

実験の結果をデフォルト訳語による評価結果 (最頻度の訳語による翻訳結果) と一緒に表 2 に示す．また参考に市販の翻訳ソフト (H システム) による翻訳結果も示す．

表 2: 実験結果

評価	ALT-J/E	デフォルト訳語	(参考)H システム
○	49% 2572 文	22% 1141 文	27% 302 文
△	40% 2081 文	33% 1740 文	40% 442 文
×	11% 589 文	45% 2361 文	33% 470 文
合計	5242 文	5242 文	1111 文

4．考察

実験の結果 ALT では，デフォルトの訳語に比べ，正解率 (○ と△ の合計) が 34%高かった．このことから結合価文法を用いればより高い精度で訳せることが分かった．しかし一方で 11%の文については ALT でも正しい訳語が出せなかった．そこでその原因を調べ，結合価文法の能力の有効性の限界を調べる．

4.1 正しい訳語が出せなかった原因

正しい訳語が出せなかった原因を評価× となった例文 589 文中 123 文について調査し表 3 に示す．

表 3: 正しい訳語が出せなかった原因

	翻訳失敗の原因	割合
1	パターンが登録されていない場合	21% 26 文
2	パターンの照合に失敗した場合	28% 35 文
3	格要素の名詞が訳せなかった場合	8% 10 文
4	慣用表現が用いられている場合	11% 13 文
5	形態素解析に失敗した場合	9% 11 文
6	係り受け解析に失敗した場合	17% 21 文
7	訳の時制がおかしい場合	1% 1 文
8	例文が二通りの意味にとれる場合	5% 6 文
	合計	123 文

4.2 原因の調査と解決策

4.1 節で挙げた原因を調査し，その解決策を探る．

(1) パターンが登録されていない場合

例文に対するパターンが登録されていない場合，訳語を絞り込むことができないので，正しい訳語が導き出されない．この場合は新しいパターンを追加すれば解決できる．

(2) パターンの照合に失敗した場合

同じ動詞に対してパターンが多く登録されると，例文に対してどのパターンが最も適しているのかの判断が困難になり，誤ったパターンを適用してしまうことがある．この場合は，パターン照合のアルゴリズムの改良により改善される．

(3) 動詞に係る名詞が訳せなかった場合

パターンは動詞とそれに係る語との関係を登録したものなので，例 3 のように動詞に係る名詞の意味 (属性) を取り違えると，例文と正しいパターンを照合できずに誤った訳を導き出してしまう．

例 3 : A 大学は法学部の助手の ポスト が空いている．
対訳 3 : The Low department of A university needs an assistant.
ALT の訳 3 : Assistant's **mail box** to the low department is open the A university .

(4) 例文に慣用表現が用いられている場合

例 4 : 彼は 話の腰を折った．
対訳 4 : He **interrupted a person's speech** .
ALT の訳 4 : He **broke the waist of talk** .

例 4 は慣用表現の文であり，文字通りのそのままの意味ではない．よって，慣用表現，例えば例 4 に対しては『N1(人) が話の腰を折る N1 interrupt a speech』というように意味属性を用いずに個別にパターンを登録する必要がある．

(5) 形態素解析に失敗した場合
(6) 係り受け解析に失敗した場合
(7) 訳の時制がおかしい場合

(5)(6)(7) の場合は訳語選択の問題とは関係ないが，これらの作業が正しく行われないと，結合価文法を使うことができない．

(8) 例文が二通りの意味にとれる場合

例 5 : 職場の不満から彼は家族の者に 当たった。
対訳 5 : He **was hard on** his family because of complaints he had about his job .
ALT の訳 5 : He **corresponded to** the person of his family from the discontent of a place of work .

例 5 は，対訳の通り『彼が家族に当たり散らした』という意味である。しかしパターンに当たる部分『彼は家族の者に当たった』だけに注目すると、対訳の意味とは別に、「彼は (誰かに対して) 家族の関係にある」という意味にも解釈できる。よって『彼は家族の者に当たった』に対応するパターン『N1 が N2 に当たる』は，前後関係から『be hard on ~ 』と『corresponded to ~ 』の二つの訳語を使い分ける必要がある．しかし，そのためには前後関係などを調べる必要があり，結合価文法だけで訳し分けるのは不可能だと言える．

3.3 考察の結果

4.2 節より正しい訳語が出せなかった例文のほとんどが，結合価文法を用いる以前の処理の問題にあることが分かった．また，『 8 ．例文が二通りの意味にとれる場合』の例文以外の問題は，解決される可能性があることもわかった．よって結合価文法を適用できれば，『二通りの意味にとれる文』以外の文に対しては，正しい動詞の訳語を導き出すことが可能と言える．

5 . まとめ

結合価文法の動詞の訳語選択についての効果を調べるため，ALT を使って IPAL の例文を翻訳し，その翻訳結果を調査した．その結果，平易な単文については 89%の文について意味の正しい訳語を導き出すことができた．また間違った訳語を出した例文についても今後さらに翻訳機などの環境が改善されれば，二通りの意味にとれる文以外の 98 ~ 99%の文について結合価文法による効果があると考えられる．

参考文献

[1] 池原，宮崎，白井，横尾，中岩，小倉，大山，林：日本語語彙体系 5 ．構文体系，岩波書店 (1997)