

012003 結合価パターンを用いた動詞句の翻訳可能性の調査

計算機工学講座 池原研究室 石上 真理子

1 はじめに

言語表現の非線形性の問題に対処する機械翻訳の方式として、文型パターンを用いる方式が提案されている[?]。この方式では、線形要素は変数で表し、変数に対する訳出表現を組み合わせて全体の英文を作成する。すなわち「線形要素の翻訳が、代入された日本語だけで可能である」という仮説がある。

本研究では、変数の中でも動詞句の翻訳に焦点を当て、動詞句の翻訳可能性について調査する。調査のねらいは、次の2点である。

1. 動詞句の翻訳において、上述の仮説の検証
2. 単文の翻訳では、動詞句の訳語選択に結合価パターンが有効である[?]が、動詞句の翻訳についての有効性の検証

そこで、本研究では、一般の翻訳ソフト「翻訳の王様」、ALT-J/E、本研究で試作した翻訳プロトタイプ「queen」の3つを用いた訳文の比較を行う。

2 結合価パターンを用いた翻訳

2.1 結合価パターン

結合価パターンは、用言と格要素（名詞+格助詞）の意味的關係を記述したものである。[?]は、約15,000件の日英文型パターン対を集録している。例を以下に示す。

N1(3主体)がN2(3主体)を援助する N1 help N2
N1(838食料)が口に合う N1 taste good

2.2 翻訳プロトタイプ「queen」

ALT-J/Eは、[?]を用いた翻訳システムである。しかし、本研究では、文型パターン翻訳に適した訳出に向けた今後の改良を狙い、[?]を参考に翻訳プログラムを試作する。

queenは次の点を考慮して作成した。

結合価パターンとの照合条件 格要素の省略、格要素の順序変更、修飾語句の挿入、任意追加の格要素を認める。意味属性制約、必修の述語付属語の条件を検査する。

照合結果の選択 慣用句表現の優先選択、カバー率による選択を行う。

名詞変数への訳語の挿入 既存の日英辞書引きプログラムを用い、入力動詞句と適合する変数に訳語を代入する。適合しない残った変数については、目的語ならば残すが、補語ならば消去する。それは、目的語は文の骨格を左右するが、補語はしないからである。英語の動詞句の出力 主語に該当する変数を削除し、句を形成する。

本プロトタイプでは、主動詞と格要素の訳出構造をとらえることが目的である。名詞訳語の選択、任意の格要素（時間、場所、手段など）の訳語挿入、副詞の挿入、冠詞・数の問題は扱わない。

3 実験

3.1 目的と方法

queen, ALT-J/E, 「翻訳の王様」の動詞句の翻訳の性能評価を目的とする。

実験に使用するデータは、既存の動詞句日英対訳データから、ランダムに取り出した325句とする。ただし、英語句の構成が2単語から14単語までのものを、それぞれ25句ずつ取り出した。

本研究の評価基準を以下に示す。

評価 : 主動詞と主名詞が理想解と完全一致。

評価 : 理想解と異なる構文だが、意味は合っている。

評価 : 基本的な構造は良いが、得たパターン、または得た英文の字面の一部を変形する必要がある。

評価× : 訳出が間違っている。

評価- : 出力パターンが一つもない。

ただし、線形翻訳で追加できる前置詞句は問わない。

評価方法として、再現率・適合率を用いる。定義を以下に示す。

再現率 R : 全入力事例数における、出力パターンが一つでもある事例数 N の割合

適合率 P : N における、総合評価が または である事例数（スコア）の割合

3.2 結果

queenの実行結果を表??に、queen, ALT-J/E, 「翻訳の王様」の R と P を表??にそれぞれ示す。

表 1: queenの実行結果

入力動詞句	彼を恐れる。	
英語理想解	fear him	
出力	当たったパターン	出力結果
	N1 fear N2	fear him

表 2: queen, ALT-J/E, 「翻訳の王様」の R と P

	queen	ALT-J/E	「翻訳の王様」
R	89.5%(291/325)	100%(325/325)	100%(325/325)
P	75.9%(221/291)	70.1%(230/325)	75.7%(246/325)

表??より、queenは、ALT-J/E, 「翻訳の王様」と同程度の性能を持つことが分かった。

4 考察

構成単語数での結果は、単語数多くなるにつれて、スコアが低くなる傾向があった。しかし、結果には揺らぎが見られた。それは、入力動詞句の目的語の有無に左右される。

元々、パターンは変数で表された目的語を含む。queenでは、変数として残ってしまった目的語は文の骨格を左右するので残している。入力動詞句に目的語がない場合、目的語の変数が残ってしまうため、基本的な構造は良くても得たパターンの字面の一部を変形する必要がある。この評価になってしまう。

5 おわりに

実験結果より、動詞句の翻訳に結合価パターンは有効であり、線形要素の翻訳が、代入された日本語だけで可能であることが分かった。しかし、入力動詞句に目的語がない場合において精度が落ちることも分かった。今後は目的語がない場合についても対処する。

参考文献

- [1] 池原, 阿部, 徳久, 村上: 非線形な表現構造に着目した重文と複文の日英文型パターン化, 自然言語処理, Vol.11, No.3, pp.69-95, 2004.
- [2] 金出地, 徳久, 池原, 村上: 結合価文法による動詞の訳語選択能力の評価, 自然言語処理, Vol.11, No.3, pp.149-164, 2004.
- [3] 池原, 他7名: 日本語語彙大系, 岩波書店, 1997.
- [4] 徳久, 村上, 池原: 文型パターンパーサの試作, 言語処理学会第10回年次大会発表論文集, pp.608-611, 2004.
- [5] 石上, 徳久, 村上, 池原: 結合価パターンを用いた動詞句の翻訳可能性の調査, 言語処理学会第11回年次大会, 2005(発表予定).