

逐次変換方式による韓日翻訳ツールの高度化と評価

中村 順一 各務 宏昭
九州工業大学 情報工学部

1 はじめに

韓国語は、語順が日本語とほとんど同じであり、また日本語の助詞に相当するものが存在する。そのため、構文・意味解析を行わなくても、単語や助詞の単純な置換だけで、ある程度日本語への翻訳が可能である。そして、このような韓日両国語間の類似点に着目した機械翻訳システムの研究・発表が多く行われている [1] [2]。しかし、両国語間の機械翻訳システムを実用化し、高品質な翻訳結果を得るためには、訳語、助詞等の選択がやはり必要であるため、深い意味解析を避けることができない。この点が実用システムを作成する上では問題となっている。

ここでは、韓日両国語間の語順等の類似点を利用し、訳語選択をユーザに任せる手法を考える。つまり、翻訳過程でユーザとシステムが協力して翻訳するワードプロセッサ風の機械翻訳システムを作成することにした。翻訳の際に、ユーザが行なうことは、入力と訳語選択であり、システムは単語切りや、辞書引き等を行なう。そこで、我々は、韓日両国語間の語順などの類似点を利用し、かつ深い意味処理が必要な訳語選択はユーザに任せる対話的な逐次変換方式による韓日翻訳ツールに関する研究を行なっている [3, 4, 5]。なお、このツールは最低限、韓国語辞書を引くことができる、日本語を母国語とする人間を対象ユーザとして想定している。

これまでに作成したツール [3, 4, 5] は、ユーザが入力した文節を韓国語の文法規則に従い単語に区切り、単語毎の日本語訳の候補をユーザに表示し、ユーザが訳語を選択する、という逐次的な訳語決定の方法をとっていた。しかし、この方法では、前文脈からの情報が少ない場合に、助詞等の訳語選択が困難になるという問題や、既存の韓国語テキストファイル等を翻訳する場合には入力作業に無駄が生じるという問題が起きていた。そこで、これらの問題を解決するため、訳語選択を一時的に保留し、翻訳を先に進め、後文脈を見てから最終的に訳語を選択したり、韓国語テキストファイルを直接翻訳できるようなユーザインタフェースに改良し、ツールの高度化を図った。本稿では、新たに作成したシステムのユーザインタフェースについて述べ、そのシステムを用いた評価実験結果について報告する。

2 ユーザインタフェースの改良

改良後のツールの基本的な翻訳例を図1に示す。まず、翻訳をしたい文の先頭にカーソルを移動し、Ctrl-SPC-keyにより範囲指定を行なう(画面1)。そして、翻訳したい文の最後にカーソルを移動させ?-keyで変換をする(画面2)。これで、範囲指定した部分の日本語訳が得られる(画面3)。ここで得られた文を見ると、“研究開発費で”の助詞“で”の訳が不自然であると考えられる。そこで、ユーザは、→(右矢印)-keyにより文節間の移動を繰り返し、“研究開発費で”に文節を移動させる(画面4)。Meta-SPC-keyにより語尾変換を行なうと“研究開発費へ”という、より自然な日本語訳が得られる(画面6)。

訳語選択を一時的に保留する為に、未確定の文節に対して、(1)解析結果、(2)辞書登録形、(3)テキスト上や辞書中でのマーカーの位置情報、(4)日本語訳、これらの情報を保存するリストを用意した。そして、文節間を移動する毎に、リスト中より文節にそれぞれの情報を割り当て、訳語選択の保留を実現した。

3 評価実験

また、本ツールの有効性を調査するために、実際に本ツールを用いて韓国語の経済誌「エコノミスト(321号, 324号)」¹から抜き出した600文について、前半300文を、韓国語の文法知識をある程度有し、韓国語の辞書引

¹<URL:http://168.126.70.2/econo/96_1/eco321a.html>, <URL:http://168.126.70.2/econo/96_1/eco324a.html>

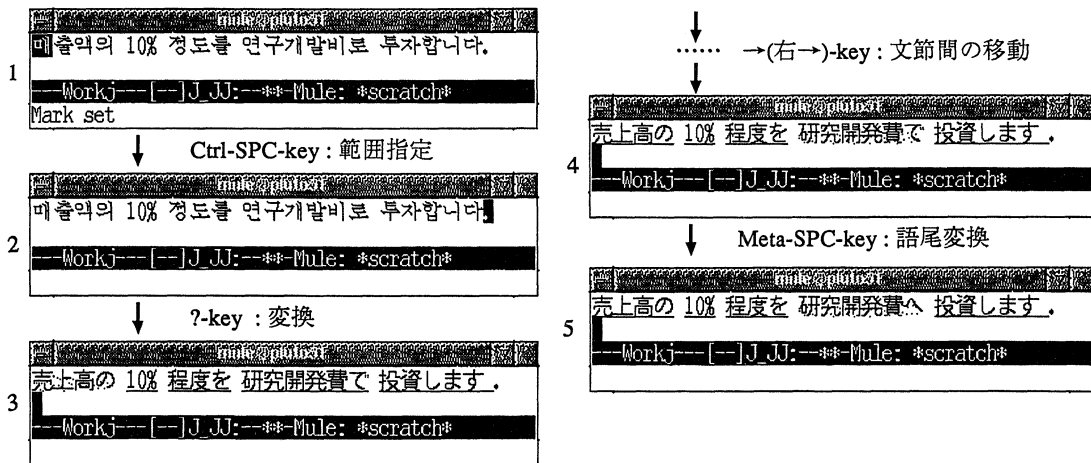


図 1: 翻訳例

きを容易に行なえる翻訳者 A が、後半 300 文を韓国語の文法知識に乏しく、韓国語の辞書引きに多少困難を要する翻訳者 B が翻訳実験を行なった。なお、辞書の初期状態は、ハングル基本単語辞典 [6] に従い 1504 語用意した。そして、翻訳結果に対して、翻訳の質、翻訳の質と翻訳作業の関係の 2 つの観点から評価を行なった。翻訳の質に関しては、「理解容易性」 [7] と「翻訳容易性」の 2 つの評価基準を設けた。

3.1 評価基準

翻訳された結果は必ずしも正しい限らない。そこで、翻訳結果の評価の方法が問題となる。本ツールでは、評価は「理解容易性」 [7]² と「翻訳容易性」の二つの観点から行なった。「理解容易性」とは、本ツールにより得られた訳文がどの程度理解できるかどうかを表すものである。また、本ツールは翻訳を対話的に行ない、訳語選択等はユーザーが行なう方式をとっている。そのため、得られた翻訳結果の「理解容易性」が同レベルでも、翻訳に要する作業が文によって変わってくると考えられる。そこで、ユーザーが訳語選択、あるいは、辞書登録時の辞書引きの作業をスムーズに行えたかどうかを表す「翻訳容易性」を評価基準の 1 つとして考案した。以下にその基準の概略を示す。これらは、ユーザーの主観により規定した。

- 理解容易性(得られた訳文がどの程度理解できるか) [7]
 - A: 文意は明瞭で疑う余地はない。
 - B: 文意は明瞭で理解できるが、文法、用語用法、文体上に多少の問題が見られ、A には至らない。
 - C: 全体的文意は把握できるが、文法、用語用法、文節切りの問題のため、細部理解に自信がもてない。
 - D: 文法、用語用法、文節切りの問題のため、相当な思考の後わずかに文意を仮定できるか、もしくはほとんど把握できない。
 - E: まったく理解できない。
- 翻訳容易性(ユーザーが訳語選択、あるいは、辞書登録時の辞書引きの作業をスムーズに行なえたか)
 - A: すべての単語の意味選定は容易であり、作業は、極めて単純である。
 - B: 殆んどの単語の意味選定は容易であり、作業は、単純である。
 - C: 幾つかの単語の意味選定が困難であり、作業は、多少困難である。
 - D: 殆んどの単語の意味選定が困難であり、作業は、困難である。

² [7] では「理解容易性」と「忠実度」の 2 つの観点を使用しているが、本ツールは翻訳の方式上、忠実の評価ができないので、ここでは、「理解容易性」だけを使用した。

表 1: 韓日翻訳ツールの翻訳の質の評価

		翻 訳 者 A							翻 訳 者 B						
		翻 訳 容 易 性							翻 訳 容 易 性						
		A	B	C	D	E	合計	割合	A	B	C	D	E	合計	割合
理 解 容 易 性	A	107	39	6	0	0	152	50.7	86	33	11	0	0	130	43.3
	B	21	23	8	0	0	52	17.3	21	24	12	1	0	58	19.0
	C	20	37	15	0	0	72	24.0	10	41	45	4	0	100	33.0
	D	0	7	11	0	0	18	6.0	0	0	4	6	0	10	3.3
	E	0	0	6	0	0	6	2.0	0	0	0	2	0	2	0.7
合 計		148	106	46	0	0	300		117	98	72	13	0	300	
割 合		49.3	35.3	15.3	0	0			39.0	32.7	23.7	4.3	0		

E: 作業がまったく不可能といえる程, 困難である。

韓日翻訳ツールの翻訳結果として満足できるものは, 理解容易性が A ~ C の文および翻訳容易性が A ~ C の文である。

3.2 翻訳の質の評価

以上の評価基準に従った, 翻訳者別の評価結果を表 1 に示す。翻訳者 A に関しては, 満足できる翻訳結果, つまり, 理解容易性が A ~ C の文および翻訳容易性が A ~ C の文は, 300 文中 276 文, 92% の結果が得られた。また, その中で, 翻訳が容易に行なえたもの, つまり, 翻訳容易性が A ~ B の文は, 300 文中 247 文, 82% であった。

翻訳者 B に関しては, 満足できる翻訳結果は, 300 文中 283 文, 94% の結果が得られた。また, その中で, 翻訳が容易に行なえたものは, 300 文中 215 文, 72% であった。

翻訳が困難だったもの, つまり, 翻訳容易性が D 以下の文は, 翻訳者 B の行なった 13 文であり, これらには, 韓国の企業名, 地名, 固有名詞等, 韓国の知識がなければ翻訳ができないものが多かった。

3.3 翻訳の質と文節数, 移動・変換・登録回数の関係

図 2, 3 に理解・翻訳容易性と文節数, 移動・変換・登録回数の関係のグラフを示す。どちらのグラフも理解・翻訳容易性が悪くなるにつれてグラフが拡大されている。これは, 理解・翻訳容易性が悪くなるにつれて, 移動, 変換等の操作が増加した事を表わしている。また, 翻訳者 B のグラフの方が, 翻訳者 A のグラフに比べ大きく広がっている。これは, 翻訳者 B の方が, 翻訳者 A に比べ, 翻訳において移動等の操作が多かった事を表わしている。

図 2 のグラフに関して, 翻訳者 B の理解容易性 E のグラフが翻訳者 A のものとは異なった傾向があるのが見られる。これは, 翻訳者 B が理解容易性 E の文に対しては, 翻訳を諦めて, 移動, 変換等の作業を殆ど行なわなかった為だと考えられる。また, 1 文あたりの移動・変換・登録回数を総合的に判断して, 本ツールで翻訳を行なった場合, 翻訳作業がそれ程困難ではないことが確認された。

4 おわりに

本稿では, 逐次変換方式による韓日翻訳ツールの改良による高度化と, 改良後のツールの評価実験について述べた。韓国語の文 600 文に対して, 翻訳した結果, 満足できた結果は, 600 文中 559 文, 93% の結果が得られた。また, 本ツールによる翻訳作業もさほど困難でない事が確認され, 本ツールの韓国語翻訳に於ける有効性が確認できた。しかし, 今後の課題として以下のようなものが残されている。まず, 韓国語文節内の形態素処理機能の強化や, ツール自体の補助機能の追加である。今後, これらの問題点を解決しつつ, ツールの使用経験を積み重ねながら, さらに改良, 評価を行なっていく予定である。

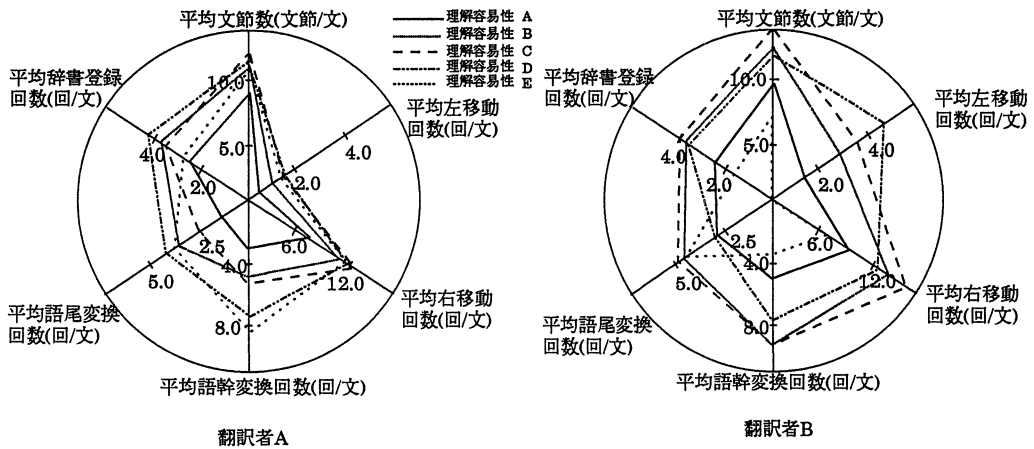


図 2: 理解容易性と文節数, 移動・変換・登録回数の関係グラフ

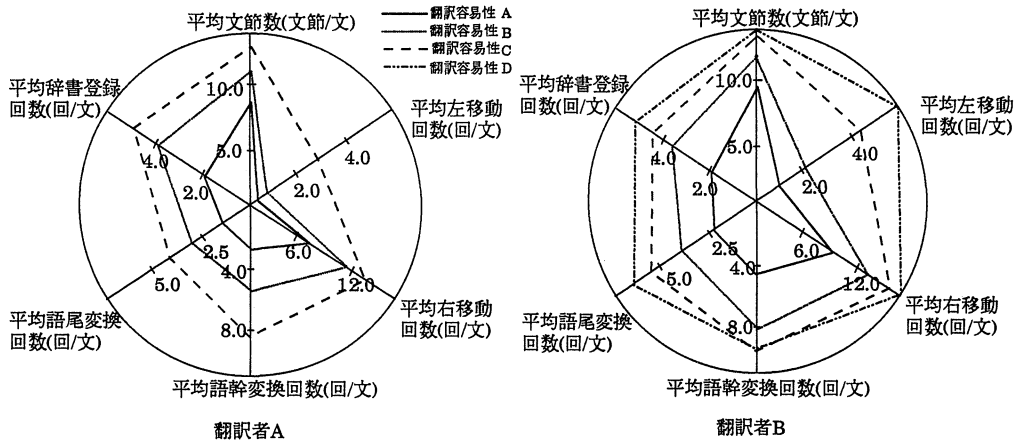


図 3: 翻訳容易性と文節数, 移動・変換・登録回数(回/文)の関係グラフ

参考文献

- [1] 李, 中嶋, 安居院: “助述表現の意味対応による日韓述部機械翻訳システム”, 情報処理学会論文誌, Vol.31, No.6, pp.801-809 (1990).
- [2] 李, 小沢: “韓日機械翻訳のための音韻表現形式による用言の活用処理”, 情報処理学会論文誌, Vol.33, No.12, pp.1565-1577 (1992).
- [3] 各務他: “逐次変換方式による韓日翻訳ツールの試作”, 信学技報, Vol.94, No.292, pp.47-54 (1994).
- [4] 各務他: “逐次変換方式による韓日翻訳ツールの評価”, 言語処理学会第一回大会, pp.285-288, (1995).
- [5] Hiroaki, et al.: “Interactive Korean to Japanese Translation Tool Using Sequential Selection Method”, PACLING95 (1995).
- [6] 金, 韓: ハングル基本単語辞典, 白帝社 (1993).
- [7] 中村: “日英/英日抄録文翻訳システム Mu”, bit 別冊 機械翻訳, pp.143-150, (1988).