

拡張翻訳テーブルを用いた日韓機械翻訳

5H-9

金 政仁† 金 泰錫‡ 大駒 誠一†

慶応義塾大学 ・ 東義大学（韓国）

1. はじめに

日韓機械翻訳を考えると、まず両国語の特徴に関する研究を先行させるが、多くの研究者らは両国語間の類似性に驚いていた。言葉の順序がほとんど同じであること、語尾の活用による表現の多様化、助詞を用いた語と語の接続関係の表示、漢字で表れる名詞の意味、尊敬語の存在などは機械翻訳の面で見ると、大変有利であり、ほとんどの研究者らは直接翻訳方式を提案している。両国語は単語の入れ替えだけでもかなりの文の意味が通じるほどの類似性を持っており、もう少しの活用語の処理、多訳語の処理を追加することによって、実用化できると期待されている。

本論文は、日韓直接翻訳方式で有効であると判明された「翻訳テーブルを用いた日韓直接翻訳方式」を基本として、その長所を維持しながら短所を最小化した「拡張翻訳テーブルを用いた日韓機械翻訳方式」について記述する。

2. 従来の翻訳テーブル

翻訳テーブルとは、日本語の用言が使われたとき、その用言の活用形態やその後に来る単語の意味によって対応される韓国語をあらかじめ登録したものである。翻訳テーブルは品詞別に分けられており、後に来る単語との接続関係によって、動詞は17個、形容詞は12個、形容動詞は11個の訳語を用意していた。後に来る単語の接続情報としては、品詞、活用形、助動詞の文法的な意味などを分離せず用いていた。表1に日本語の動詞「行く」に対する翻訳テーブルを示す。

助詞・助動詞の場合も翻訳テーブルを用いた訳語の選択処理ができる。助詞の場合は前接した単語の意味素性を、助動詞の場合は用言と同じく後接した単語との接続関係を用いて適切な訳語を選択する。用言の翻訳テーブルとは異なって、訳語の数が決められていないのが特徴である。

既存の翻訳テーブル（TT）を用いた日韓機械翻訳方式では、大きく4つの短所を持っていた。

- 1) 用言の翻訳テーブルには接続関係の記述が固定されていて拡張しにくい
- 2) 前後の接続関係を表す接続情報が少ない
- 3) 活用単語が同じ韓国語に翻訳される場合も複数登録しなければならない（表1の3, 5など）
- 4) 品詞別に翻訳テーブルが存在するので、翻訳処理の一貫性がない

以上の問題点を解決するために、翻訳テーブルの概念を拡張した拡張翻訳テーブルを導入したい。

表1 日本語の動詞「行く」の翻訳テーブル

活用形	活用単語	接続情報による韓国語の対応		
		No	後接单語の意味および文法情報	対応する韓国語
未然形	行か	1	否定	가 ga
	行こ	2	否定以外	가 ga
連用形	行き	3	丁寧, 希望	가 ga
		4	丁寧+過去	갔 gat
		5	丁寧+否定	가 ga
		6	様態	갈 gal
		7	用言	가기 gagi
		8	連用中止形	가서 gase
		9	転成名詞	행 haung
	行っ	10	過去	갔 gat
終止形	行く	11	推定, 伝聞	갈 gal
		12	丁寧	가 ga
		13	接続助詞	가 ga
		14	終了	가다 gada
連体形	行く	15		갈 gal
仮定形	行け	16		가 ga
命令形	行け	17		가 ga

3. 拡張翻訳テーブル

翻訳テーブル（TT）が用言の基本形を基準に考案されたのに対して、拡張翻訳テーブル（ETT）

Japanese-Korean Machine Translation Based on Extended-Translation-Tables.

† Jung-In Kim, Seiichi Okoma : Keio University, Japan

‡ Tae-Suk Kim : Donggeui University, Korea

は、活用単語を基準に作られた。まず、活用単語をエントリにした辞書を用意する。それに加え、エントリが複数の訳語に翻訳可能な場合は、活用語辞書を用いる。日本語の動詞「行く」の拡張翻訳テーブルを表2に示す。

表2 日本語の動詞「行く」の拡張翻訳テーブル

A. エントリテーブル

原形	品詞	エントリ	単語の情報		訳語	マルチ
			活用形	No		
行く	5段動詞	行か	未然1		가	No
		行こ	未然2		가	No
		行き	連用1		가	Yes
		行っ	連用2		갔	Yes
		行く	連体		가는	Yes
		行け	仮定		가	No

B. 活用語テーブル

エントリ	活用形	前接情報			後接情報			訳語
		品詞	活用形	No	品詞	活用形	No	
行き	連用1				助動	終止	12:ます	갑
					助動	連用1	12:ます	갔
					助動	未然2	12:ます	갑
					記号		2:	가서
		体言	LOC					행
行っ	連用2				助詞		22:て	가
					記号		1:.	간다
行く	終止				助動		15:そう	간다
					助詞			가
					助詞		37:ほど	갈

拡張翻訳テーブルの特徴を羅列すると、次のようになる。

- 1) 接続関係による訳語の数を品詞に関係なく自由に変更できる
- 2) 付属語の特殊な接続処理のため、助詞・助動詞・記号に番号を付けて接続関係を表す接続情報として使う
- 3) 代表的な訳語を用意し、同じ訳語に訳される場合の重複登録を避ける
- 4) 品詞に関係なく接続関係を表す項目が一致している

次の例文はTTとETTを用いて翻訳した結果である。TTでは不自然な訳文がETTでは自然に翻訳されていることがわかる。

例文1) 名古屋には行ってもいい。

TT >나고아에는 갔어도 좋다.

ETT >나고아에는 가도 좋다.

説明) 「行っ」が過去の訳語に翻訳されてしまう

例文2) 名古屋へ行くほどに余裕がある。

TT >나고아에 기정도로 여유가 있다.

ETT >나고아에 갈 정도로 여유가 있다.

説明) 後に来る助詞によって訳語が違う

例文3) 行きと帰りは反対の概念だ。

TT >행과 돌아오기는 반대의 개념이다.

ETT >가기의 돌아오기는 반대의 개념이다.

説明) 転成名詞「行き」が目的地を表していない場合、別の訳語が必要である

4. 結論

翻訳テーブルを用いた日韓機械翻訳方式は、活用語処理に有効であった。しかし、翻訳テーブルの拡張性欠如、訳語を選ぶための接続関係の不足、同じ訳語の重複登録、翻訳処理の一貫性欠如などが問題点として表れた。このような短所を解決するため、拡張翻訳テーブル方式を提案し、どの単語にも訳語の数が自由に設定できた。そして、接続関係を細かく表示するため、助詞・助動詞の1個1個を接続情報として用いることで、特殊な助詞・助動詞と接続されている場合も自然な訳語に翻訳できるようになった。また、代表訳語を用意し、訳語の重複登録を避けた。最後に、すべての品詞の翻訳テーブルの構造を同じ形式に統一させ、品詞に関係せず一貫性を持った翻訳処理ができるようになった。すなわち、日韓機械翻訳における拡張翻訳テーブル方式は従来の翻訳テーブル方式が持っていた短所の相当の部分を解決することができた。

参考文献

- 1) 金泰錫, 金政仁, 大駒誠一, 浦昭二: 意味接続関係に基づく翻訳テーブルを用いた日韓機械翻訳における日本語の形態素解析, 情報処理学会43回全大集(3), pp201-202(1991)
- 2) 金泰錫, 浦昭二: 日韓機械翻訳における意味接続関係を用いた韓国語の生成処理, 情報処理学会論文集, Vol.33 No12, pp.1578-1588(1992)