

日韓機械翻訳における拡張翻訳テーブルを用いた 韓国語の生成方法

金 政 仁[†] 大 駒 誠 一^{††}

日本語と韓国語は文字や単語は大きく異なっているが、文法的な面では、多くの類似性を持っている。この類似性をうまく用いることで、日韓機械翻訳における構文解析や意味解析のかなりの部分が省略できる。しかし、類似性を用いる翻訳処理は、用言の語幹と活用語尾を分離して翻訳を行うので、自然な韓国語の生成がうまくできない短所を持っていた。以前に、我々は活用を考慮した訳語の翻訳テーブルをあらかじめ用意することによって、ある程度用言の自然な訳語が得られることを示した。しかし、この翻訳テーブルを用いた翻訳処理には、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現れた。そこで、本論文では、このような短所を解決するため、新しく翻訳テーブルの概念を拡張した拡張翻訳テーブルを用いた日韓機械翻訳方式を提案し、その翻訳実験の結果、かなり自然な韓国語が得られたので報告する。

A Method of Generating Korean Language by Using Extended Translation Tables on Japanese-Korean Machine Translation

JUNG-IN KIM[†] and SEIICHI OKOMA^{††}

Japanese and Korean grammar have many similar features. Using these similarities, we have developed Direct Translation Method by which a deep level analysis of syntax and semantics between two languages can be omitted. Japanese-Korean Direct Machine Translation systems have a significant difficulty in selecting an appropriate target word when a declinable Japanese word is translated into Korean because the stem and ending of such words are separated by the translation system. Previously, a method of using Translation Tables for solving problems with declinable words in advance before translating them into the suitable target words was proposed by Kim. But this method was lacking the expansibility of Translation Tables and relation-connection for selecting a target word and also has such demerits as the repetition of registering target words redundantly and an inconsistency in the translation process. In order to overcome these problems we propose a new Japanese-Korean Direct Machine Translation Method utilizing Extended Translation Tables, and we evaluate the translation results of both methods.

1. はじめに

日韓機械翻訳を考えるにあたって、その両者の文法を比較すると、互いに驚くほどよく似ている。特に、語順がほとんど同じなので、機械翻訳の際大変有利であり、日韓機械翻訳では、この数年の間に多くの文献で直接翻訳方式を提案している^{10)~22)}。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、述部の活用処理や多義語の処理を追加することによって、実用に耐えるシステムの構築ができるかと期待されている。述部の活用に対しては、両国語の述部表現の意味対応による活用語処理¹⁰⁾、韓日機械翻訳に適用した音韻表現形式による述部の活用処理¹⁵⁾、などが提案されているが、韓国語の用言の不規則的な活用をうまく表現しきれなかった。活用語処理に対して、我々は用言の不規則的な活用をルール化することを避ける方策として、訳語をあらかじめテーブルの形に用意して翻訳する方式を提案し¹³⁾、その有効性を確認した¹⁷⁾。一方、多義語の処理としては、語と語の関係を用いた意味解析による単文の多義語処理¹¹⁾、体言の意味素性を羅列した格形式パターンを用

[†] 浦項工科大学校情報通信研究所
POSTECH Information Research Laboratories, Pohang
University of Science and Technology

^{††} 慶応義塾大学理工学部
Faculty of Science and Technology, Keio University

いた動詞の多義性処理¹⁶⁾などが提案され、改善されつつある。

本稿では、従来の翻訳テーブルが持っている短所を改善するため、前後接続単語との接続関係が記述できる拡張翻訳テーブルを用いる方式を提案する。また、この新しい拡張翻訳テーブルを用いて翻訳処理を行い、既存の翻訳テーブルを用いた翻訳処理の結果と比較し、検討する。

2. 日本語と韓国語の比較

日韓機械翻訳において、両国語の類似性を用いるためには、両国語の比較分析を行う必要がある。日本語と韓国語は文法的にはよく似ている言語であるが、日韓機械翻訳を考えると、品詞の不一致、活用型の不一致、助詞の使い方の違いなどの様々な相違点が問題となる。

2.1 両国語間の類似性

日本語と韓国語を言語学の面から比較して、次のような類似点があることが明らかになっている⁷⁾。

- (1) 語順がほとんど同じである。
- (2) 語尾の活用によって用言の表現が多様化できる。
- (3) 漢字で表現された名詞の概念がほとんど同じである。
- (4) 助詞を用いて語と語の接続関係を表示する。
- (5) 単数、複数の区別があいまいであり、冠詞が存在しない。
- (6) 真偽の疑問文に対する答えは、その命題の真偽に依存する。
- (7) 敬語が発達している。

このような両国語の類似性を用いて日韓機械翻訳を行うと、比較的簡単な解析および変換により、品質の良い翻訳システムの構築が可能であると期待される。

2.2 両国語間の相違点

日本語と韓国語は似ている言語であるといわれているが、実際に対応をとってみると、次のように様々な相違点がある^{6)~8)}。

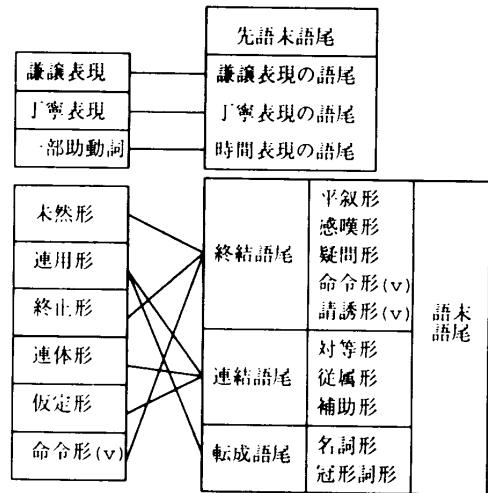
- (1) 品詞が一致していない。

日本語は10種類の品詞(名詞、動詞、形容詞、形容動詞、助詞、助動詞、感動詞、接続詞、連体詞、副詞)であるが、韓国語は9種類の品詞(名詞、代名詞、数詞、助詞、動詞、形容詞、冠形詞、副詞、感嘆詞)である。その対応関係も1対1に対応していない。表1に日本語の品詞とそれに対応する韓国語の品詞を示す。品詞の分類方法は文献によって異なるが、ここでは両国語の学校文法を用いた^{6),8)}。

表1 日本語と韓国語の品詞対応

Table 1 A part of speech in Japanese corresponding to Korean.

日本語の品詞	対応する韓国語の品詞
名詞	名詞, 代名詞, 数詞
動詞	動詞
形容詞	形容詞
形容動詞	形容詞
副詞	副詞
連体詞	冠形詞, 代名詞
接続詞	副詞
感動詞	感嘆詞
助動詞	叙述格助詞, 動詞, 形容詞
助詞	助詞



(a) 日本語の活用形 (b) 韓国語の活用形

図1 日本語と韓国語の活用型比較

Fig. 1 Correspondence of Japanese and Korean conjunction types.

- (2) 用言の活用型が一致していない。

日本語の用言は「語幹 + 語尾」の分け方をしているが、韓国語の用言は「語幹 + 先語末語尾 + 語末語尾」に分けられる。図1に両国語の活用型を比較する。また、活用ルールの場合、日本語の用言は「上下1段」、「5段」の規則活用に、「カ変、サ変」の不規則活用をするのに対して、韓国語では語幹と語尾が一定な形態に結合、活用する規則用言とそうでない11種類の不規則用言に分けられる⁷⁾。表2に韓国語の不規則活用を示す。そして、活用語では語幹と語尾が結合するとき起こる特徴として、日本語には音便「い、っ、ん」が存在するが、韓国語では子音で終わる語幹(閉鎖語幹)には、調音素「으」を挿入する。同時に語幹が陽母音「ㄷ, ㅌ」であるか陰母音「ㄱ, ㅋ, ㆁ, ㆁ」であるかによって副詞形語尾は「아, 어」の2種類の

表2 韓国語の用言の不規則活用
Table 2 Irregular conjugation of Korean declinable words.

活用型	説明	不規則活用の例
ㄷ	語幹末音ㄷが母音で始まる語尾と結合するときㄷに変異する	듣 으니 → 들으니 묻 으니 → 물으니
ㄹ	語幹末音ㄹがㄹㅂㅅㅇで始まる語尾と結合するときㅂㅅㅇに変異する	살 니 → 사니 울 니 → 우니
ㅂ	語幹末音ㅂが母音で始まる語尾と結合するときwに変異する	돕 우니 → 도우니 눅 어서 → 누워서
ㅅ	語幹末音ㅅが母音で始まる語尾と結合するときㅅが脱落する	긋 으니 → 그으니 붓 어서 → 부어서
ㄷ	語幹末音ㅇが어と結合するときㄷに変異する	흐르어서 → 흘려서
ㅇ	語幹末音ㅇが어母音語尾と結合するときㅇに脱落する	쓰 어서 → 써서
우	語幹末音우が語尾어と結合するとき우に脱落する	무 어서 → 퍼서
여	하다動詞に아で始まる語尾が結合すると아가여に変異する	운동하 어서 → 운동하여서
러	이르다動詞の語幹にあ, 어で始まる語尾が結合すると러に変異する	이르 어 → 이르러
거라	命令形語尾어라が거라に変異する	자 어라 → 자거라
너라	命令形語尾어라가너라に変異する	오 어라 → 오너라

表3 助詞「に」の対応表
Table 3 Correspondence table of the particle “NI”.

番号	日本語の例文	役割	対応する韓国語の例文
1	AはBに等しい	比較	A는 B와 같다
2	水が水に変わった	変化	얼음이 물로 변했다
3	彼は社長になった	結果	그는 사장이 되었다
4	映画を見に行く	目的	영화를 보러 간다
5	彼女にあげた	生物	그녀에게 드린다
6	東京に送る	場所	동경에 보낸다
7	先生に差し上げる	尊敬	선생께 바친다
8	彼は文子にあった	対象	그는 문자를 만났다
9	自動車に乗った	乗物	자동차를 탔다

変異形が存在する（母音調和による形態音素の変異）。

- (3) 助詞の使い方が一致していない。

日韓機械翻訳の観点から見ると、「に、で、の、が」など、様々な助詞は1対nの対応関係を持っており⁷⁾、直接対応による訳語の生成ができない。特に助詞「に」は適切な韓国語の助詞に対応しにくい。日本語の助詞「に」を使った例文とその訳文を表3に示す。日本語の助詞「に」は韓国語に翻訳される時、表3のように「와, 에, 에게, (으)러, 데, 를」などの様々な訳語を選択する必要がある。

- (4) 敬語の表現が異なる。

また、両国語には敬語が存在する。尊敬の意味

を含む名詞「先生、あなた」、謙讓の意味を含む名詞「私、家内、小生」などは「선생님, 당신」, 「저, 집사람, 소생」に1対1の対応ができるが、その以外の敬語を表すときの表現に相違点がある。

- 日本語には丁寧の意味を表す接頭語「お、ご」が存在するが、韓国語には存在しない。（ごはん, お魚 ⇒ 밥, 생선）
- 日本語では「お（動詞の連用形）になる」の形で尊敬の意味を表すが、韓国語では尊敬表現の先語末語尾を用いて、「語幹 + 尊敬表現の先語末語尾(시) + 語尾」の形で尊敬の意味を表す。（お読みになる ⇒ 읽으시다）
- 日本語には謙讓の意味を含む動詞「申す、いたす、存ずる、いただく」などが存在する。そして「お（サ変名詞、転成名詞）する」の形で謙讓の意味を表すが、韓国語では「語幹 + 謙讓表現の先語末語尾(옵, 이옵, 시옵, 사옵) + 語尾」の形で謙讓の意味を表す。（お待ちします ⇒ 기다리겠습니다）

このように日本語と韓国語はいくつかの相違点を持っており、両国語の単語を1対1に対応させることは簡単ではない。特に、両国語の活用語の対応は複雑な関係になっており、適切な処理を必要とする。

3. 従来の翻訳テーブル

前後に接続された単語との接続関係による1対1の対応関係を作るために、著者らは日韓機械翻訳において効率が良い活用語処理の方策として「翻訳テーブルを用いた用言の活用語処理」を提案し、その有効性を確認した^{14),17),18)}。翻訳テーブル方式とは、あらかじめ、各単語の活用形に前後の接続関係を考慮した訳語を用意して、選択アルゴリズムを利用して翻訳対象単語の適切な訳語を選択する方法であり、韓国語の述部の生成部分での誤りを少なくするために提案した。この方式は活用語において、両言語の複雑な対応関係を意識せず、活用語の翻訳ができるメリットを持っている。本稿では2種類の翻訳テーブルについて述べるが、用語の混乱を避けるために、これからは従来の翻訳テーブルを「翻訳テーブル」といい、この翻訳テーブルを拡張して新たに提案する翻訳テーブルのことを「拡張翻訳テーブル」という。

3.1 翻訳テーブルを用いた翻訳処理

日韓直接翻訳方式は日本語の単語を韓国語の単語に

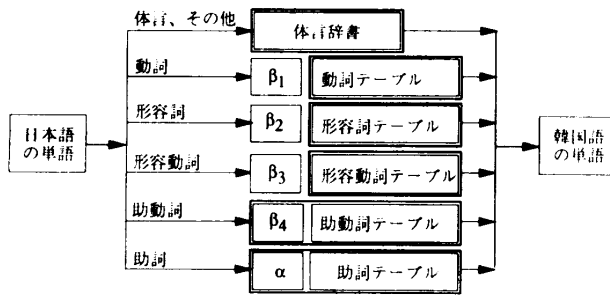


図2 従来の翻訳テーブルを用いた品詞別翻訳処理
Fig. 2 Translation process with translation tables.

1:1 対応させる単純な作業の繰返しで成り立っている。翻訳テーブルは 1:1 の単純な対応が不可能である活用語を処理する場合、前後に接続している単語との接続情報を用いて翻訳可能な複数の訳語から適切な訳語を選ぶことを支援する。図2 は日本語の1つの単語が韓国語の単語に翻訳される過程を表す。翻訳テーブルには複数の訳語が登録されており、前接した単語の文法情報 α と後接した単語の文法情報 β によって適切な訳語を選択することができる。ここで、二重四角形は翻訳辞書を表しており、 β_1 は動詞の17種類の後接情報を、 β_2 は形容詞の12種類の後接情報、 β_3 は形容動詞の12種類の後接情報を表している。そして、 β_4 は助動詞の後接情報であるが、助動詞の後接情報はその数が決められていない。それは、助動詞ごとに接続できる品詞が異なるためである。助詞の場合は前接する単語の情報を参考にして訳語を用意する。前接情報である α も助詞ごとに異なる接続情報を用意する。動詞の接続情報 β_1 は表4の後接单語の文法情報という項目にその内容が記述されている。すべての動詞に対して同じ接続情報 β_1 を準備し、それに合わせて訳語を用意する。動詞の翻訳テーブルのように β_2 , β_3 は形容詞、形容動詞に後接单語の接続情報が定義されており、その訳語は形容詞辞書、形容動詞辞書に各々保存する。翻訳テーブル方式は品詞を基準とし、品詞別翻訳処理のルーチンを用いて翻訳対象の単語を次々に翻訳する。

3.2 翻訳テーブルの例

表4 に日本語の動詞「行く」に対する翻訳テーブルを示す。動詞は6種類の活用型に分けられている。また後接单語の情報によって17種類の接続情報とそれによる訳語を記述した¹³⁾。

実際に翻訳テーブルを用いて訳語を生成してみよう。日本語の文「学校に行きませんでした」という文は、動詞「行く」の連用形が表れており、「行き」の後に「丁寧(ませ) + 否定(ん) + 丁寧(でし) + 過去(た)」の意味の助動詞が接続されている。この場合は

表4 動詞「行く」の翻訳テーブル
Table 4 Translation table of the verb "IKU".

活用型	活用単語	接続情報による韓国語の対応			適用例
		No	後接单語の文法情報	対応韓国語	
未然形	行か	1	否定	가	行かない
	行こ	2	否定以外	가	行こう
連用形	行き	3	丁寧, 希望	가	行きます
		4	丁寧 + 過去	갔다	行きました
		5	丁寧 + 否定	가	行きません
		6	様態	갈	行きそうだ
		7	用言	가기	行きにくい
		8	連用中止形	가서	行き,
		9	転成名詞	행	行き
		10	促音便	갔	行った
		終止形	行く	11	推定, 伝聞
12	丁寧			가	行くでしょう
13	接続助詞			가	行くが,
14	終了			가다	行く.
連体形	行く	15		갈	行く人は
假定形	行け	16		가	行けば
命令形	行け	17		가	行け.

「行き」と隣接された意味が優先され「丁寧(ませ) + 否定(ん)」の後接意味と対応する韓国語「가」に翻訳される。しかし、一般に動詞はもっと複雑な性質を持っており、17個の接続情報で、すべての接続関係を表すことは無理である。

(1) 「行く」の連用形「行っ」は、過去の意味を表す助動詞「た」に加え(表4の10)、補助用言を後に続ける助詞「て」および並列の意味を表す助動詞「たり」と接続することができ、その訳語も「갔, 가」となって、連用形「行っ」の自然な表現ができない。

アメリカに行っ
미국에 갔다
学校に行っってもいい
학교에 가도 좋다
山に行ったり, 海に行ったりする
산에 갔다가 바다에 갔다가 한다

(2) 終止・連体形「行く」の場合は「行くが」, 「行くのは」, 「行くが」, 「行くほど」などの接続ができ、各々に対する訳語「가는 것, 가는, 가, 갈」を用意する必要がある(表4の13)。

東京に行くがいい
동경에 가는 것이 좋다
東京に行くのは私だ
동경에 가는 것은 나다
東京には行くが, 君には会えない
동경에는 가지만 너와는 만날 수 없다

東京に近く行くほど渋滞がひどい
 동경에 가까이 갈수록 정체가 심하다

- (3) 連用形「行き」が転成名詞と解析されるためには、前接した品詞や体言の意味素性が必要である。(表4の9)

そこに行きなさい
 거기에 가세요
 東京行き, 大阪行きなどが書いてある
 동경행, 오사카행등이 적혀있다
 行きと帰りは反対の概念だ
 가기와 돌아오기는 반대의 개념이다
 東京に行き, 彼女と会った
 동경에 가서, 그녀와 만났다

- (4) 連用形「行き」、終止形「行く」などでは同じ訳語が複数登録される場合が多く、翻訳辞書の効率を良くするためには訳語の重複登録を避ける工夫が必要である。(表4の3と5, 12と13)

以上のように従来の翻訳テーブルにはいくつかの翻訳率を下げる原因が残っていた。翻訳テーブル方式での短所をまとめると、次のようである。

- $\beta_1, \beta_2, \beta_3$ の後接情報をシステムの中に記述した。すなわち、拡張性が足りない。助動詞の後接情報 β_4 、助詞の前接情報 α と同じように翻訳辞書の中にデータの形式で登録する必要がある。
- 接続関係を記述するため、単語の情報として品詞、体言の意味素性、用言の活用形、助動詞の文法的な意味などを用いた。適切な接続関係を表すためにより細かい分類をする必要がある。
- 複数の接続関係から同じ訳語に翻訳できることが多い。代表的な訳語を導入して重複定義を避ける。
- 翻訳システムは各品詞別の翻訳ルーチンを持ち、また翻訳辞書の構造も品詞別に異なる。一貫性のある翻訳システムの構築ができない。

以上のような短所の解決策として、動詞の翻訳テーブルを含めて、すべての翻訳テーブルの構造を新たにし、翻訳テーブルの形態を拡張させた拡張翻訳テーブルを提案する。

4. 拡張翻訳テーブル方式

日本語と韓国語の類似性を用いた日韓機械翻訳の基本的な概念は、単語別の1対1対応関係からの置き換えによる訳語の生成である。1対1の対応では処理できない活用語や多義語などについては、単語ごとに前後に接続できる単語との接続関係を記述し、その接

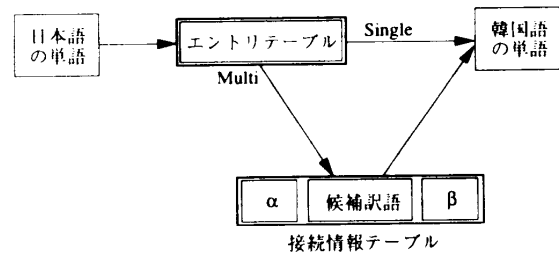


図3 拡張翻訳テーブルを用いた品詞別翻訳処理

Fig. 3 Translation process by using extended translation tables.

続関係を用いて適切な訳語を選ぶ特別な処理を行うことで解決できる。活用語別に前後接続単語との接続関係を記述したテーブルを拡張翻訳テーブルといい、これから拡張翻訳テーブルを用いた翻訳方式を拡張翻訳テーブル方式と呼ぶことにする。

4.1 拡張翻訳テーブル方式の翻訳処理

拡張翻訳テーブルを用いた翻訳処理の流れを図3に示す。従来の翻訳テーブル方式とは異なって品詞に依存しないので、大変シンプルな翻訳処理になっている。ここで候補訳語を登録するときには前接情報 α 、後接情報 β も一緒に記述する。

4.2 拡張翻訳テーブルでの接続要素

前後に接続された単語の情報 α, β を表すために単語の属性、すなわち、品詞、体言の意味素性、用言の活用形、助詞および助動詞の一連番号などを用いる。品詞の分類は従来の翻訳テーブル方式とほぼ同じであるが、名詞には「サ変名詞」を、そして動詞には「可能動詞、進行動詞、補助動詞」などを追加分類した。そして、体言の意味素性はIPAL辞書の分類方法を用いており⁵⁾、用言の活用形は活用形態により15種類に分類した。また、翻訳テーブルでは助動詞の文法的な意味を用いたが、拡張翻訳テーブルではその代わりに助詞、助動詞、記号の一連番号を接続情報として利用する。

- 品詞
 名詞 (する名詞), 動詞 (上1, 下1, カ行変格, サ行変格, 5段)(可能, 進行, 補助), 形容詞, 形容動詞, 助動詞, 助詞 (格, 接続, 副, 終), 副詞, 連体詞, 接続詞, 感動詞, 記号など
- 体言の意味素性
 具体名詞 (動物, 人間, 組織および機関, 植物, 生物の部分, 自然物, 生産物および道具), 現象名詞, 抽象名詞 (動作および作用, 精神, 言語作品, 性質, 関係, 空間および方角, 時間, 数量), その他
- 用言の活用形

(a) エントリテーブル							(b) 接続情報テーブル											
原形	品詞	エントリ	単語の情報		代表 訳語	マルチ	エン トリ	品詞	活用形	No	前接情報			後接情報			候補 訳語	
			活用形	No							品詞	活用形	No	品詞	活用形	No		
行く	5動	行か	未然1		が	No												
行く	5動	行こ	未然2		が	No												
行く	5動	行き	連用1		가서	Yes	行き	動詞	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:	가		
											体言	LOC				행		
													助詞			가기		
行く	5動	行っ	連用2		갔	Yes	行っ	動詞	連用2				助詞		22:て	가		
行く	5動	行く	連体		가는	Yes	行く	動詞	終止				記号		1:	간다		
													助動		15:そう	간다		
													助詞			갈때		
													助詞		37:ほど	갈		
行く	5動	行け	假定		가	Yes	行け	動詞	命令1							가		

図4 動詞「行く」の拡張翻訳テーブル
Fig. 4 Extended translation tables of the verb "IKU".

表5 接続形態の分類
Table 5 Classification of connection types.

区分	品詞 第1要素	意味素性または活用形 第2要素	一連番号 第3要素
Type1	[名詞]	[体言の意味素性]	[NULL]
Type2	[用言]	[用言の活用形]	[NULL]
Type3	[助動詞]	[助動詞の活用形]	[助動詞の 一連番号]
Type4	[助詞]	[NULL]	[助詞の 一連番号]
Type5	[記号]	[NULL]	[記号の 一連番号]
Type6	[その他]	[NULL]	[NULL]

未然1, 未然2, 未然3, 連用1, 連用2, 連用3, 連用4, 連用5, 連用6, 終止, 連体, 假定, 命令1, 命令2, 語幹

- 助詞, 助動詞, 記号の一連番号

助詞 58個, 助動詞 19個, 記号 30個

前接情報 α および後接情報 β は上記の要素を3個まで結合して表示する。その種類は6つのタイプに分けられており、前接情報 α および後接情報 β は同じ形態で表現する。表5にその種類を表示する。

4.3 拡張翻訳テーブルの構成

拡張翻訳テーブルは2種類の形式のテーブルで表現する。1つは、すべての品詞のエントリ、属性、代表的な訳語などが1対1の対応関係で登録されているエントリテーブルであり、もう1つは、訳語が複数である場合、前後の接続関係により、適切な訳語を選択することができる接続情報テーブルである。体言のほとんど、助詞の一部などの1個の訳語にしか翻訳できな

い単語はエントリテーブルを用いた置き換え処理により簡単に翻訳できるが、用言、助動詞、一部の助詞など、2個以上の訳語に翻訳できる単語であれば、代表的な訳語はエントリテーブルに、それ以外の候補訳語は接続情報テーブルに用意する。拡張翻訳テーブル方式では、品詞に依存せず、単語ごとに単語の前接情報を3個の接続要素として記述し、同様に後接できる単語の情報を3個の接続要素で記述する。これを接続情報テーブルの1個の接続関係にして、前接にも後接にも両方向同時の接続にも同じ形態の接続要素を利用することができるように作成し、接続関係の記述に一貫性を維持した。図4に動詞「行く」に対する拡張翻訳テーブルの例を示す。図4の(a)エントリテーブルには「行く」の6種類の活用語がすべて登録されている。その中で、「行か」、「行こ」は1個の訳語だけに翻訳可能であるが、「行き」、「行っ」、「行く」などは前後の接続関係により訳語が複数存在するので、図4の(b)のような接続情報テーブルに接続関係と候補訳語を用意する。拡張翻訳テーブルに基づく日韓機械翻訳システムは以下のような様々な利点が期待できる。

- 既存の翻訳テーブル方式では記述できなかった前後の接続単語との細かい接続関係が付属語を別々に処理させることにより可能になる。したがって、細かい接続関係を拡張翻訳テーブルに記述することができ、翻訳率が向上する。
- 動詞、形容詞、形容動詞などの翻訳テーブルは接

表6 例文から接続形態の抽出

Table 6 Extraction of connection types from an example.

区分	単語	品詞	意味素性 または活用形	一連番号
Type1	学校	[名詞]	[場所]	[NULL]
Type4	に	[助詞]	[NULL]	[004]
Type2	行き	[5 段動詞]	[連用 1]	[NULL]
Type3	ませ	[助動詞]	[未然 1]	[015]
Type3	ん	[助動詞]	[終止]	[005]
Type3	でし	[助動詞]	[連用 1]	[012]
Type3	た	[助動詞]	[終止]	[017]
Type5	.	[記号]	[NULL]	[001]

統関係と訳語の数が決められていたが、拡張翻訳テーブルでは、接続関係を自由に記述することができ、接続関係の数を活用語ごとに調節できる。このことで、活用語ごとに独立した接続関係の記述ができ、候補訳語の追加、削除を始め、辞書のメンテナンスなどが簡単になる。

- 活用語別の代表訳語を用意して、同じ訳語の重複登録を避けることができる。
- 拡張翻訳テーブルは同じ形式のエントリテーブルや接続情報テーブルにすべての品詞が納められるように設計して、辞書の統一性を図り、一貫性のあるシステムの構築が可能である。すなわち、翻訳システムの構築、メンテナンスなどが容易である。

4.4 拡張翻訳テーブル方式による翻訳処理

たとえば、日本語の文「学校に行きませんでした」を翻訳させるためには、形態素解析の情報が必要で、拡張翻訳テーブル方式では前後の接続情報を利用するため、表6のような情報を利用する。ここで、候補訳語の中で適切な訳語を選択するため、接続関係の評価を行う。3個の接続要素は、各々その接続の強度が違うので、重みを付与して接続要素の優先順位を求める。第1接続要素(品詞)は比較的緩やかな制限であるので重みは1点とする。第2接続要素(活用型)は、第1接続要素より具体的な制限であり、重みに2点を与える。第3接続要素(一連番号)は、一連番号であり、単語そのものを示しているので重みを3点とする。そして、詳しい接続関係の評価方法をアルゴリズム1に表記した。日本語の単語「学校」は1:1の対応により「학교」に翻訳できる。しかし、助詞「に」の場合は「에, 에서, (으)러, 데」などの複数の訳語が存在するので、前後に接続されている単語の情報(「学校」および「行き」)を用いて、適切な訳語を選択する必要がある。図5は助詞「に」の拡張翻訳テーブルを示す。

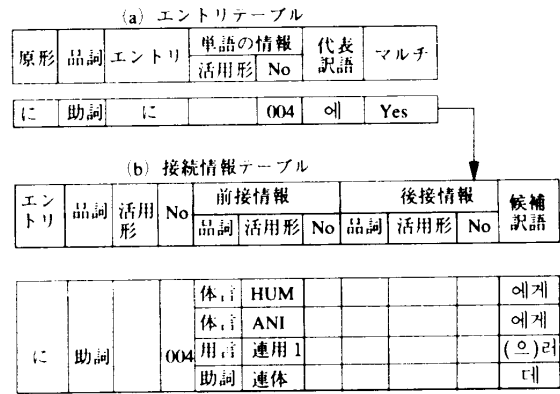


図5 助詞「に」の拡張翻訳テーブル
Fig. 5 Extended translation tables of the particle “NI”.

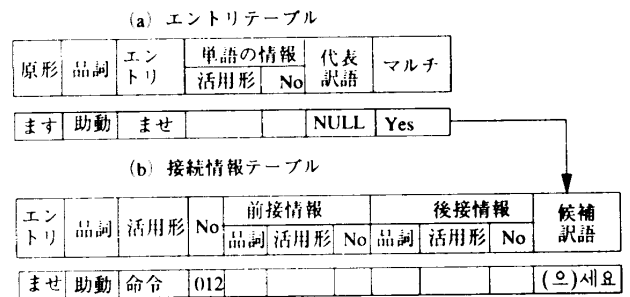


図6 助動詞「ませ」の拡張翻訳テーブル
Fig. 6 Extended translation tables of the auxiliary verb “MASE”.

「に」の拡張翻訳テーブルには、前後の接続関係による複数の訳語が登録されており、例文には「学校[名詞][場所][NULL]」を前接情報 α とし、「行き[5 段動詞][連用 1][NULL]」を後接情報 β とする条件を満足する訳語を接続情報テーブルから探してみるが、その条件に満足する接続関係が存在しないので代表的な訳語であるエントリテーブルの「에」に訳す。同じ方法で動詞の連用形「行き」を翻訳していく。「に[助詞][NULL][004]」を前接情報 α とし、「ませ[助動詞][未然 1][012]」を後接情報 β とする訳語を接続情報テーブルから探してみる。図4を参照すると、最初の接続関係は第2接続要素(活用型)が一致していないので評価値に-1点を与える。2番目と3番目も同じく-1点が付与される。4番目の接続関係は後接3要素がすべて一致しているので、第1接続要素から1点、第2接続要素から2点、第3接続要素から3点が与えられ、評価値は6点になる。すなわち、各々の評価値は-1点、-1点、-1点、6点、-1点、1点となり、4番目の「가」に訳される。次の「ませ[助動詞][未然 1][012]」の翻訳は、前接情報 α 「行き[5 段動詞][連用 1][NULL]」と後接情報 β 「ん[助動詞][終止][005]」を用いるので、図6の「ませ」の拡張翻訳

(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		
ん	助動詞	ん	連体	7	지 않는	YES

(b) 接続情報テーブル

エントリ	品詞	活用形	No	前接情報			後接情報			候補訳語
				品詞	活用形	No	品詞	活用形	No	
ん	助動詞	終止	7							지 않
							助詞		37:ほど	지 않을
							記号		1:Period	지 않는다
							助詞		12:と	지 않는다
		助動詞	未然 1	12:ます	記号		1:Period	지 않습니다		
	連体	7				助詞		15:ので	지 않	

図7 助動詞「ん」の拡張翻訳テーブル
Fig. 7 Extended translation tables of the auxiliary verb "N".

テーブルから接続関係を評価した結果 -1 点になり、代表的な訳語「NULL」に翻訳される。

そして、「ん [助動詞][終止][005]」は、前接情報 α 「ませ [助動詞][未然 1][012]」と後接情報 β 「でし [助動詞][連用 1][019]」との接続関係を用いて、図7の拡張翻訳テーブルから、「지 않」に翻訳できる。同じ方法で残りの単語をすべて翻訳させると「학교에 가지 않습니다」になる。アルゴリズム1に前接单語の接続情報と後接单語の接続情報を用いた訳語の選択アルゴリズムを示す。

アルゴリズム1: 適切な訳語を選択するアルゴリズム
Algorithm 1: A Selecting Algorithm for Suitable Target Words

```

/* エントリテーブルから日本語の単語（見出し語）を探す */
翻訳対象単語をエントリ辞書から探す
if (存在するか)
    選択訳語にエントリ辞書に用意されている韓国語をセットする
else
    { 未登録単語の処理をする
      return(0)}
if (訳語が1個であるか)
    return (選択訳語)
/* 訳語が複数存在する場合の処理をする */
ループの初期状態で総合評価値に0を割り当てる
S[1]..S[6]にNULLをセットする
前接单語  $\alpha$  の接続要素を S[1],S[2],S[3] に割り当てる
後接单語  $\beta$  の接続要素を S[4],S[5],S[6] に割り当てる
while (接続情報辞書のエントリと翻訳対象単語が同じである間、繰り返す) {

```

```

候補単語の評価値に0を割り当てる
C[1]..C[6]にNULLをセットする
辞書の接続情報を C[1]..C[6]に割り当てる
/* 適切な接続要素を辞書から選択する */
for (i=1 から, 6まで, 1ずつ増加) {
    if (S[i] または C[i] が NULL であるか)
        次の比較のため, ループの最後まで行く
    else
        if (S[i] と C[i] が同じであるか)
            /* 品詞の一致:1, 活用型の一致:2, */
            /* 番号の一致:3をいれ, 加重点数を与える */
            { 評価値にiを3に割った残りを足すその残りが0であれば3を足す }
        else
            評価値に -1 をセットし, ループから抜ける
    }
/* 適切な訳語の選択処理 */
if (評価値が総合評価値より大きい)
    { 総合評価値に評価値を代入する
      選択訳語に接続情報辞書の候補訳語を代入する }
接続情報辞書のポイントを1増やす
}
return (選択訳語)

```

大部分の場合、日本語の前接单語や後接单語との接続関係だけで活用語を翻訳することが可能である。し

(a) エントリテーブル							(b) 接続情報テーブル										
原形	品詞	エントリ	単語の情報		代表訳語	マルチ	エントリ	品詞	活用形	No	前接情報			後接情報			候補訳語
			活用形	No							品詞	活用形	No	品詞	活用形	No	
行く	5動	行って	連用4		가서	Yes	行って	動詞	連用4				助詞		29:も	가	
													助詞		7:から	가서	
													動進			가고	
													補動			가	
行く	5動	行った	連用5		간	Yes	行った	動詞	連用5				助詞		13:が	갔	
													助詞		15:ので	갔기	
													助詞		16:から	갔기	
													助動		11:そうだ	갔	

図8 複合語「用言+て(で), た(だ)」の拡張翻訳テーブル

Fig. 8 Extended translation tables of compound word (Declinable Word + TE (DE), TA (DA)).

かし、接続助詞「て(で)」の場合は、用言の後ろに接続され、同じ後接関係を与えてしまい適切な訳語を選択することができない。この現状は過去、完了、状態などを表す助動詞「た(だ)」、丁寧の助動詞「ます」にも現れる。本システムでは、「て(で), た(だ), ます」と連結される用言を特殊形として、連用4形、連用5形および連用6形とし、複合語として処理する。図8は「て(で), た(だ)」と一つになった動詞「行く」の連用2形である「行っ」の複合語処理を表している。その結果、動詞「行く」は、連用2形「行っ」の代わりに「行って」および「行った」が連用4形、連用5形として用意され、後接した単語との関係を表すことができる。また、「行く」の連用1形「行き」には「ます」との複合語「行きます」を連用6形として追加、分類する。

5. 翻訳実験の評価および検討

従来の翻訳テーブル方式を評価する際に2つの測度を利用した。1つは生成された韓国語の正しさの理解度であり、もう1つは原文の意味がどのくらい伝わったのかを観測する忠実度である¹³⁾。表7, 表8は翻訳結果のテーブルであり、理解度を1, 2, 3, 4に分けて横軸に記入し、忠実度をA, B, Cに分けて縦軸に表示した。

理解度は

- 1) 韓国語の文章として正しく、十分理解できる。
- 2) 韓国語の文章としてややおかしいが、理解できる。
- 3) 韓国語の文章としてかなりおかしく、かなり推論しないと意味が分からない。
- 4) 韓国語の文章として正しくなく、理解不可能である。

表7 従来の翻訳テーブル方式による翻訳結果

Table 7 Results of translation using the translation table.

	1	2	3	4	合計
A	183	8			191
B	53	150	52	4	259
C			18	6	24
合計	236	158	70	10	474

表8 拡張翻訳テーブル方式による翻訳結果

Table 8 Results of translation using the extended translation table.

	1	2	3	4	合計
A	216	3			219
B	97	98	35	3	233
C			18	4	22
合計	313	102	53	7	474

忠実度は

- A) 原文の意味がそのまま訳されている。
- B) 原文の意味にやや足りない単語で訳されている。
- C) 原文の意味に合わない訳文である。

と定義して実験のデータを求めた。従来の翻訳テーブルを用いた韓国語の生成処理実験¹³⁾で得られた結果と拡張翻訳テーブル方式を用いた翻訳処理の結果を表7および表8に示す。実験に使った文章は科学記述文献、新聞記事、機械のマニュアルなど、文法的な誤りが少ない文である。また、既存の翻訳テーブル方式との比較のため、今回の実験にも翻訳テーブル方式を評価するときに使った474文¹³⁾をそのまま用いた。ここで、理解度1, 2と忠実度A, Bを翻訳の合格としたら、両方式の翻訳率は各々83.1% (394)と87.5% (415)であり、拡張翻訳テーブル方式が約4.4%程度の高い翻訳率を表した。しかし、韓国語の理解度の面で見ると、1, 2段階の数字の差に注目する必要がある。

表9 翻訳結果の例

Table 9 Example of translation results.

No	段階	例文	翻訳された訳文	正しい訳文
1	1A	行って見る あの本ならば、今も買える 全部食べたように見える 行こうと思った	가 본다 그책이라면, 지금도 살 수 있다 전부 먹은 것처럼 보인다 가려고 생각했다	가본다 그책이라면 지금도 살 수 있다 전부 먹은 것처럼 보인다 가려고 생각했다
2	1B	被告人から手紙が来た	피고사람에게서 편지가 왔다	피고인에게서 편지가 왔다
3	2A	行く必要がある	가는 필요가 있다	갈 필요가 있다
4	2B	科学というのは	과학과 말하는 것은	과학이라고하는 것은
5	3B	改善されつつある	개선되어 계속 있다	계속 개선되고 있다
6	3C	十分考えられること	충분생각되는 일	충분생각할 수 있는 일
7	4B	根も葉もない	뿌리도 입도 없다	밑도 끝도 없다
8	4C	3億円を引いた金額	3 억원을 뺀 금액	3 억원을 뺀 금액

る。理解度の1が49.8% (236) から66.0% (313) まで上がり、訳文の理解はできるとしても、かなりの後編集が必要であったものが拡張翻訳テーブル方式に依ることによって約16%以上効率が良くなった。表9に各段階別の代表的な翻訳結果の例を示す。翻訳結果に対しての説明は次のとおりである。

- (1) 活用語、助詞、助動詞などが正しく翻訳されたケースである
- (2) 体言の多義性による翻訳失敗
- (3) 未来形が現在形になったテンスの翻訳失敗
- (4) 後接する用言のパターン分類ができていない場合の助詞の翻訳失敗
- (5) 語順の合わない翻訳失敗
- (6) 助動詞の多義性による翻訳失敗
- (7) 慣用語の翻訳失敗
- (8) 用言の多義性による翻訳失敗

(2), (6), (8)の多義性による翻訳失敗を解決するためには、今後、多義性解消のために提案された特別な処理^{11), 16)}を加える必要があると思われる。また、用言に対しては品詞、活用形だけを用いており、用言のパターン分類をしていない。より良い翻訳のためには、今後、用言のパターン分類を考慮する必要がある。そして、テンスの翻訳失敗(3)、両国語の順番が合わないことからの翻訳失敗(5)、慣用表現の翻訳失敗(7)などは、本システムでは処理しにくい失敗であり、後編集などを行って、訳語の質を上げる必要がある。

6. むすび

最近、日韓機械翻訳に対する研究が活発に行われている。そのほとんどの方式は類似性を活かした1対1対応による直接翻訳方式であり、著者らは、翻訳テーブルを用いた日韓機械翻訳システムを提案して、用言の活用処理部分での有効性を確認した。しかし、翻訳テーブル方式は翻訳辞書に大きく依存するシステムに

なり、翻訳辞書の効率を良くするための研究が必要となった。その要求に合わせて、本論文では、翻訳テーブルを改良した拡張翻訳テーブルを用いる翻訳処理方式を提案した。拡張翻訳テーブルの特徴としては、まず、助詞、助動詞の1個1個を接続の要素として記述することによって、翻訳テーブルでは表現できなかった細かい接続関係が拡張翻訳テーブルには記述できた。また、活用単語別に接続関係が自由に記述でき、訳語の追加、削除を始め、辞書のメンテナンスなどが簡単になった。そして、品詞別に独立して定義されていた翻訳テーブルの項目を統一化して、翻訳処理の一貫性を目指した。

拡張翻訳テーブルを用いた翻訳処理が有効であるかを確認するため、従来の翻訳テーブル方式で使った474個の例文を用いて翻訳処理を行った結果、韓国語の理解度の面で1レベル(韓国語として十分理解できる)に属する文章の合格率が49.8% (236) から66.0% (313) に大きく改善されて自然な訳語の生成ができること、そして既存の翻訳システムより4.4%向上した87.5%の翻訳率が確認でき、従来の翻訳テーブル方式での短所の相当の部分が解決できたことが分かった。しかしながら、特殊な慣用表現、「受身、可能、自発、尊敬」などの多義性が生じる助動詞「れる、られる」などは、本システムでの問題点として残っており、今後、さらに改良を続けたい。

参考文献

- 1) 野村浩郷：言語処理と機械翻訳，講談社(1991)。
- 2) 田中穂積：自然言語解析の基礎，産業図書(1989)。
- 3) 野村浩郷，田中穂積：機械翻訳，bit別冊，公共出版(1988)。
- 4) 長尾 真：機械翻訳サミット，オーム社(1989)。
- 5) 村田賢一，ほか：計算機用日本語動詞辞書IPAL (Basic Verbs) — 解説編，情報処理振興事業協会技術センター(1987)。

- 6) 田近洵一：くわしい国文法，文英堂 (1987).
- 7) 黄 燦鎬ほか：韓日語対照分析，明志出版社，韓国 (1993).
- 8) Cho. K.B.：ハイライト高校文法，志学社，韓国 (1992).
- 9) Choi. K.S.ほか：日本語翻訳システム環境下での韓国語翻訳システム開発のための一考察，情報処理学会自然言語処理研究報，Vol.86, No.4 (1988).
- 10) 李 義東，中嶋正之，安居院猛：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).
- 11) 李 義東，中嶋正之，安居院猛：格形式を介した語と語の関係を用いた意味解析による韓日単文機械翻訳システム，電子情報通信学会論文誌，Vol.J73-D-II, No.3, pp.418-426 (1990).
- 12) Kim, T.S. and Ura, S.: A Japanese-Korean Machine Translation Based on Conjugated Words Analysis, *ICEIC*, pp.199-203 (1991).
- 13) 金 泰錫，浦 昭二：日韓機械翻訳における意味接続関係を考慮した翻訳テーブル，慶応義塾大学，理工学研究科，No.92002 (1992).
- 14) 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，第43回情報処理学会全国大会論文集，Vol.3, pp.201-202 (1991).
- 15) 李 秀絃，小沢慎治：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).
- 16) 金 政仁，大駒誠一：日韓機械翻訳における動詞の多訳性処理，第45回情報処理学会全国大会論文集，Vol.3, pp.97-98 (1992).
- 17) 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方法，情報処理学会論文誌，Vol.33, No.12, pp.1578-1588 (1992).
- 18) 金 泰錫，浦 昭二：日韓機械翻訳における否定文の処理，情報処理学会論文誌，Vol.34, No.5, pp.892-904 (1993).
- 19) 黄 道三，長尾 眞：類似性に基づいた日韓対訳テキストの文対応，電子情報通信学会自然言語処理研究会，Vol.99-12, pp.87-93 (1994).
- 20) 金 政仁，金 泰錫，大駒誠一：日韓機械翻訳における話し言葉の翻訳処理，第47回情報処理

- 学会全国大会論文集，Vol.3, pp.179-180 (1993).
- 21) Kim, J.I., Kim, T.S. and Okoma, S.: A Processing of Polysemy and Multi-Translatable Verbs on Japanese-Korean Machine Translation, *Proc. ICCTA '94*, pp.144-148 (1994).
- 22) 金 政仁，金 泰錫，大駒誠一：拡張翻訳テーブルを用いた日韓機械翻訳，第51回情報処理学会全国大会論文集，Vol.3, pp.89-90 (1995).

(平成7年12月14日受付)

(平成8年6月6日採録)

金 政仁 (正会員)



1961年生。1986年韓国啓明大学校理工学部統計学科卒業，同年国際コンピュータエンジニアリング(株)入社。1989年未来情報システム(株)を経て，1990年来日，1993年慶応義塾大学大学院理工学研究科管理工学専攻前期博士課程修了，1996年同後期博士課程単位取得退学。現在，浦項工科大学校情報通信研究所専任研究員。自然言語処理，機械翻訳などの研究に従事。言語処理学会，電子情報通信学会，大韓電子工学会各会員。

大駒 誠一 (正会員)



1936年生。1959年慶応義塾大学工学部計測工学科卒業。同年小野田セメント(株)入社。1964年慶応義塾大学工学部管理工学科助手。現在同大理工学部管理工学科教授。工学博士。プログラミング，コンパイラ，アルゴリズム，機械翻訳などに関する研究に従事。著書「アセンブリプログラミング入門」培風館，「改訂FORTRAN77」サイエンス社，「数値計算のためのCプログラミング技法」HBJ出版局，「COBOL基礎と応用」サイエンス社など。日本ソフトウェア科学会，計量国語学会，ACM，アジア太平洋機械翻訳協会各会員。