

日韓機械翻訳における話し言葉の翻訳処理

5P-3

金 政仁、金 泰錫、大駒誠一
慶応義塾大学理工学部管理工学科

1 はじめに

機械翻訳処理の実験評価の対象には、科学技術文献が用いられることが多い。これは、小説の文章とか話し言葉に比べて科学技術文献には曖昧さが少ないためである。著者らも、日本語と韓国語の間の違いが多い助詞と述部の表現を自然な訳語に翻訳するために、翻訳テーブルを提案し、科学技術文献を対象として、その有効性を示してきた(2-6)。本報告では、音声認識に関する研究開発が近年、急速に進歩していることをふまえ¹⁾、著者らが提案してきた翻訳テーブルを話し言葉に適用して、その有効性を確かめた結果を示す。これは、今後、一般の日韓自動通訳システムに拡張していく予定である。

2 日本語と韓国語の話し言葉の類似点

話し言葉は、書き言葉と比べて文法に合わない場合が多い。日本語の話し言葉では、語順が無視されたり、話者と聴者が話題の対象としている主語が省略されたりする場合が多い。これらの話し言葉の特徴は、韓国語にもあてはまる。日本語と韓国語の間の類似性は、書き言葉だけではなく、話し言葉にも次のような類似点が存在する。

- (1) 語順がほとんど同じである(話し言葉の場合は、語順が異なっても意味が通じる)。
- (2) 話者と聴者が話題の対象を知っている場合は、その主語をしばしば省略する。
- (3) 単数、複数の区別が曖昧であり、冠詞がない。
- (4) 真偽の疑問文に対する答えは、その命題の真偽に関係する。
- (5) 敬語が発達している。

3 翻訳処理の対象

Processing of spoken-language in Japanese-Korean Machine Translation .

Kim Jung-In, Kim Tae-Suk, Okoma Seiichi
KEIO University

本論文ではホテル予約⁷⁾で使われる話し言葉を採用した。これは、ホテルの予約に使われる文形式が比較的少ないことと、限られた単語でその実現が可能であるためである。また、音声の解析処理が終わった段階、つまり、発声された言葉の文法情報(品詞名、活用形)と意味情報(体言と助動詞の意味判別)が明かに解析された結果をもとに翻訳処理を行なうものとする。但し、意味を持っていない不用語(例えば、「えー」、「あー」など)については音声解析の段階で削除されることを前提とした。

4 話し言葉の翻訳処理

4.1 翻訳テーブルの作成

著者らのこれまでの研究では科学技術文献を対象としていたために不要であった、終助詞(「な」、「ぞ」、「さ」、「わ」など)の終助詞の翻訳テーブルを追加する。終助詞の翻訳テーブルの作成方法は各助詞ごとに接続可能な品詞や文法情報、または意味情報を考慮した訳語を登録する。表1は、終助詞「な」の翻訳テーブルを示す。

表1. 終助詞「な」の翻訳テーブル

No.	前接単語	前接単語の意味および文法情報	対応する韓国語	適用例
1	動詞	終止形	마 (ma)	友達にも話すな.
2	助動詞	終止形	마 (ma)	一緒に行かせるな.

4.2 ホテル予約の話し言葉

ホテルを予約するときに使われる話し言葉では、約27個の文形式がよく使われる。(KDDの資料参考)本論文では、この27個の文形式に当てはまる話し言葉を対象とする。表2は、ホテル予約で使われる文形式の例⁷⁾を示す。

表2. ホテル予約に使われる文形式と記号の例

(文形式)	
文形式 1 :	もしもし予約をしV1.
文形式 2 :	予約をV2.
文形式 3 :	DATEからDATEまでBE.
文形式 4 :	DATEからDUR間BE.
文形式 5 :	YES.
文形式 6 :	NO.
.	
文形式 2 6 :	R1BE.
文形式 2 7 :	R2BE.

(記号の定義)			
V1 :	たいです/たいですが	V2 :	お願いします
DATE :	[MONTH]DAY	DUR :	DAY/WEEK
BE :	です/ですが	YES :	はい
NO :	いいえ	R1 :	Single/Double/ Twin
R2 :	普通/高級/特別	NAME :	A/B/a/b/...
SO :	そうです	MONTH :	1月/.../12月
WEEK :	1週間/.../n週間	DAY :	1日/.../31日
WDAY :	月曜日/.../日曜日		

4. 3 翻訳処理の手順

翻訳処理では、まず、音声処理と日本語の形態素解析によって得られた各単語とその意味をもとに各単語の翻訳テーブルを辞書から読み込む。次に、単語の意味接続関係によって単語翻訳テーブルから適切な韓国語を選択する。例えば、「ダブルですが」は、表1の文形式26番であり、「ダブル」と「です」と「が」の3つの翻訳テーブルから韓国語を生成する。翻訳テーブルから適切な翻訳テーブルを選択する方法については文献4)を参照されたい。表3は、ホテル予約の話し言葉の翻訳例を示す。

表3. 翻訳処理の例

No.	翻訳結果
1	もしもし、予約をしたいですが。
	여보세요, 예약을 하고 싶은데요. (yoboseyo, yeyakulhagosupundeoyo.)
2	8月10日から3日間です。
	8월10일부터 삼일간입니다. (8wol10ilbutosamilkanimnida.)
3	そうです。
	그렇습니다. (gurutsumnida.)

No.	翻訳結果
4	ダブルですが。
	다블입니다. (doubleimnida.)
5	TANAKAです。
	다나카입니다. (tanakaimnida.)
6	ありがとうございました。
	감사합니다. (kamsahamnida.)

5 評価および考察

本論文では、科学技術文献を対象として提案した翻訳テーブルを用いる日韓機械翻訳システムを両国語間の自動通訳システムに拡張するために、話し言葉の例としてホテル予約によく使われる文形式と約300語を用いて翻訳処理を試みた。その結果、両国語間の類似点を活かしながら、複雑な活用語尾や助詞助動詞を簡単に処理するために提案した翻訳テーブルが話し言葉にも十分適用可能であることがわかった。今後、音声認識技術の進歩と伴ない、日韓自動通訳システムの実現に向け、さらに改良、拡張を図りたい。

謝辞：この研究に対して、貴重な助言と資料を頂いた(株)KDD研究所の山本誠一氏、井ノ上直己氏に感謝します。

参考文献

- 1) 黒岩真岩, 他: 内線電話への接続をタスクとする電話対話収集システム, 信学技報VOL93, No31 SP93, pp.43-48 (1993.5).
- 2) 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, 情報処理学会第43回全国大会(3), pp. 201-202 (1991).
- 3) 金泰錫, 浦昭二: 意味接続関係に基づいた翻訳テーブルを用いた日韓機械翻訳システム, 情報処理学会第43回全国大会(3), pp.203-204 (1991).
- 4) 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, TECHNICAL REPORT No:92002, Dep. of Administration Eng., Faculty of Science and Technology, KEIO UNIV. (1992.4)
- 5) 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 VOL33, No.12 pp.1578-1588(1992)
- 6) 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 VOL33, No.12 pp.1578-1588(1992)
- 7) ETTRI(韓国) : HOTEL RESERVATION TASK文例集 (内部資料)