

用例に基づく日韓の対話翻訳処理機構

4L-10

山本 和英 古瀬 蔵 飯田 仁

ATR 音声翻訳通信研究所

E-mail: {yamamoto, furuse, iida}@itl.atr.co.jp

1. はじめに

近年の日韓両国間の交流の拡大に伴い、日本語から韓国語（朝鮮語）への言語翻訳の需要は一層増大してきている。日韓言語翻訳はこれまで、後述するような言語的な類似性を背景にした直接翻訳方式が多く採用され、否定文の処理 [Kim93] や助動詞の処理 [Par95] など、直接翻訳方式で処理不可能な述部の翻訳処理を中心として研究が進められてきた。しかし、より自然な翻訳を目指す場合には、語の多義性や敬語の取り扱いなど、両言語間で異なる体系を持つ言語現象に対処しなければならず、言語的類似性に依存した1対1対応を基本とする直接翻訳方式では本質的に限界が存在する。

我々は経験的な知識を利用して翻訳を行う、用例に基づいた変換主導型の翻訳手法を提案し、すでに日英の対話翻訳に関してその有効性が確認されている [Fur94] が、我々は同手法を日韓対話翻訳、および韓日対話翻訳に適用した。以下本稿では、日韓両言語について翻訳する際に問題となる言語現象に触れ、日韓翻訳における変換主導型の翻訳の有効性を考察する。

2. 日韓両言語の相違とその言語翻訳

2.1 日韓両言語の言語的類似性

日本語と韓国語が文法的に類似した言語であることは広く知られている。例えば文中の語順は多くの場合一致するし、日英翻訳で問題となる日本語助詞「は」は多くの場合、韓国語で該当する助詞「는(은) [(n)umn]」に翻訳が可能である。また日韓両言語はどちらも中国から語彙（および漢字）を借用し、自国語の語彙に取り入れてきた経緯があるため、日本語で音読みする漢字語は多くの場合にそれを韓国語読みするだけで翻訳可能である。以上のように、日本語から韓国語に翻訳を行なう際には、韓国語で該当する訳語に置き換えることで、多くの場合に翻訳が可能と考えられている。

2.2 類似性に依存した翻訳の問題点

一方で正確な翻訳処理を困難にする様々な言語現象の相違も存在する。崔らの報告 [Cho96] によると、市販されている主な日韓翻訳システムでの誤りの多くは、形態素解析の誤りや未翻訳を除くと、日本語単語（動詞 / 名詞 / 助詞）の多義性による誤りや対訳語選定の誤りであり、これらは韓国語と日本語の文構造、語彙などの類似性に依存しすぎている翻訳になっていることが原因であると指摘している。

Japanese-to-Korean spoken dialogue translation utilizing example-based approach.

Kazuhide YAMAMOTO, Osamu FURUSE and Hitoshi IIDA.

ATR Interpreting Telecommunications Research Laboratories.

2.3 助詞の訳し分け

例えば例1に示すように、日本語の助詞「に」は韓国語の「에[e]」に対応することが多いとされるが(1-2)のように転成や変化を表す場合は「로[ro]」になり、(1-3)の「乗る」のように、ある特定の動詞と結びつく場合には「를[rwl]」¹となる場合もある。

例1: 助詞「に」に対応する韓国語

(1-1) ソウルに 着く。 (場所 / 到達点)	서울에 도착하다. [sourɛ to:ɣ ^h ak ^h ada]
(1-2) 明日に 延期する。 (転成 / 変化)	내일로 연기하다. [neillo jɔŋgihada]
(1-3) バスに 乗る。 (特定の動詞と結びついて)	버스를 타다. [posurwl t ^h ada]

2.4 対話的述部表現の訳し分け

日本語の一つの特徴として、助動詞や助詞の組み合わせによって微妙なニュアンスを表現する点が挙げられる。このような言語現象は特に対話文に多く現れ、対話文の翻訳の際にはこのような細かい表現の差異を訳し分ける必要がある。たとえば、日本語対話文で頻出する表現「…かと思う」は、該当する韓国語が多数存在する(例2)。

2.5 意味属性による手法の限界

動詞の訳し分けに対して朴らは、連語パターンと呼ぶ動詞の必須格要素とその意味属性の組み合わせによって翻訳する方式を提案している [Par95]。この手法によれば、同一形態素の動詞でも異なる格要素を伴う場合や同一の格要素でも異なる意味属性である場合に訳し分けが可能となるが、格要素が同一で意味属性の設定が困難な場合、例えば「カーテンをかける(달다[talda])」「カバーをかける(덮다[do:p^hta])」「マスクをかける(쓰다[ʔsuda])」の場合には訳し分けが困難となる。

3. 用例に基づく日韓翻訳

3.1 変換主導型機械翻訳の概要

著者らは用例に基づく変換主導型の機械翻訳(TDMT: Transfer-Driven Machine Translation)を提案している [Fur94]。TDMTは変換モジュールがいろいろなレベルの言語表現の経験的知識を入力文に適用して翻訳結果を作り上げるので、用例に基づくさまざまな利点を翻訳処理結果全体に反映させることができる。文献 [Fur94]では、TDMTを日英翻訳に適用して作成したプロトタイプについて報告した。

¹直前の語が子音で終わる場合は「을[wl]」になる。

例2: 「...かと思う」の訳し分け

(2-1) 夕べから熱が出て風邪じゃないか <u>かと思う</u> んですが。 (推測)	이젯밤부터 열이 나서 감기가 아닐까 싶은데요. [ɔdʒe ^t pambut ^h ɔ jori nasɔ kamgiga anil ^p ka sɪp ^h undeyo]
(2-2) フロントにも届けられた方がよろしいか <u>かと思</u> いますが。 (推量、不確実な断定)	프런트에도 신고하시는 게 좋을 것 같은데요. [p ^h urɔnt ^h wedo sɪŋoha ^h sinun ke ʃɔul ko ^t kat ^h undeyo]
(2-3) 少しお待ちくださればすぐ準備できるか <u>かと思</u> います。 (推量)	잠시만 기다려 주시면 곧 준비 될 겁니다. [ʃamʃiman kidarjɔ ʃufɪmjɔn ko ^t ʃunbi twel kɔmnida]

3.2 用例に基づく最尤候補決定処理

TDMT は日本語と英語のように文の語順が大きく異なる言語間の場合の言語構造の変換に有効であるだけでなく、任意の言語間で生じる語句の訳し分けに対しても有効に機能する。例えば例1の用例(1-1)に示すような助詞の訳し分けについては、表1に示すような経験的知識として蓄積された用例の中から、(X, Y) = (ソウル, 着く)とベストマッチする用例を意味距離計算²により求める。この結果、(テグ, 到着する)が最尤候補となり、「に」の翻訳結果として「에[e]」が選択される。

ここで、TDMTでは変換知識の原言語部分を組み合わせることによって入力文をカバーする原言語構造を作成した後、意味距離計算を行なうため、「ソウルに」と「着く」の両語が隣接しない場合、例えば「ソウルに間もなく着く」などのような入力文に対しても対応できる。また、TDMTにおける変換知識は、単語レベルのような小さな言語単位から単文、複文レベルのような大きな単位まで同一の処理として取り扱うことができるため、例2に示したような述部表現に対しても例1と同様の処理によって訳し分けることが可能となる。

表1: 日韓変換知識の一部

変換パターン	対訳	用例
X に Y	X 에 [e] Y X 를 [rul] Y X 로 [ro] Y	((<u>テグ</u> , 到着する), ...) ((<u>電車</u> , 乗る), ...) ((<u>ウォン</u> , 換える), ...)
X かと思う	X 을 것 같다 [[u]l ko ^t ka ^p ta] X 을 것이다 [[u]l ko ^s ida] X 을까 싶다 [[u]l ^p ka sɪ ^p ta] ...	((<u>よろしい</u>), ...) ((<u>できる</u>), ...) ((<u>風邪</u>), ...)

3.3 プロトタイプシステムの実現

我々は日韓翻訳システムのプロトタイプを作成し、ホテルでの予約や観光案内などの旅行会話に関するATR対話コーパス中の約750文の日本語文と対訳韓国語文の分析によって、表1に示すような変換知識や解析知識を構築した。構築したシステムの規模の概要を表2に示す。

²TDMTでの意味距離計算については[Fur94]を参照のこと。

表2: プロトタイプシステムの概要

構築した変換知識数	約550
辞書語彙数	約5400語(基本形)
システム記述言語	Common Lisp
1文の処理時間	約0.3秒
(計測した計算機)	Sun SPARCstation 10)
(計測した文の文長)	約6.4形態素/文)

今回構築したプロトタイプでは概ね良好な翻訳が得られているが、変換知識の量が十分でないために、まだ未知の文に対する頑健性が不十分である。現在は、システムによって出力された翻訳結果を評価すると共に、変換知識の補強を行なっている。

4. まとめと今後の課題

本稿では、対話で出現する文を対象にした日本語から韓国語への翻訳について、計算機での翻訳処理の際に問題となる両言語の言語現象の差異を指摘し、これに対し経験的な知識を利用して翻訳を行う変換主導型の翻訳が有効であることを示した。

今後は、生成したプロトタイプシステムを充実させていくと同時に、その有効性を確認するための定量的な評価を行ない、報告する予定である。

参考文献

- [Cho96] 崔紀鮮, 金泰完: 日韓機械翻訳システムの現状および分析, 年次大会発表論文集, 第2回, pp. 433-443, 言語処理学会(1996).
- [Fur94] 古瀬蔵, 隅田英一郎, 飯田仁: 経験的知識を活用する変換主導型機械翻訳, 情報処理学会論文誌, Vol. 35, No. 3, pp. 414-425 (1994).
- [Kim93] 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).
- [Par95] 朴哲済, 文敬姫, 郭鐘根, 李鐘赫, 李根培: 連語パターンによる日-韓機械翻訳システムの構築とその評価, 研究会資料 NL109-2, 情報処理学会(1995).