

TR-IT-0097

変換主導型翻訳方式による韓国語と日本語間の
対話翻訳に関する研究

A Study on Dialogue Translation Between Korean and Japanese
Using a Transfer-Driven Machine Translation Approach

金 徳奉 古瀬 蔵

Deok-Bong Kim and Osamu Furuse

1995年3月

概要

本報告書で我々は変換主導型翻訳方式を使った日本語と韓国語間の対話機械翻訳システムについて述べる。それから、我々は日本語と韓国語間の言語翻訳の評価実験を行なっていくうえで予想される問題点と、両言語間の言語翻訳に関連する特徴について考察する。具体的に、我々はYokoyama等(1994)が収集して評価した日英翻訳が難しい日本語例文について、日韓翻訳の見地からそれを再検討してみる。

エイ・ティ・アール音声翻訳通信研究所

ATR Interpreting Telecommunications Research Laboratories

©(株)エイ・ティ・アール音声翻訳通信研究所 1995

©1995 by ATR Interpreting Telecommunications Research Laboratories

Contents

1	はじめに	1
2	韓国語と日本語間の言語学的な類似性	3
3	言語学的な類似性が持つ効果	4
3.1	構造的曖昧性 (Structural Ambiguity)	5
3.2	意味的曖昧性 (Semantic Ambiguity)	9
3.3	語彙的曖昧性 (Lexical Ambiguity)	10
3.4	文脈的曖昧性 (Contextual Ambiguity)	11
3.5	生成の問題 (Generation Problem)	13
4	おわりに	13

1 はじめに

日本、ドイツ、米国を中心に多言語の音声自動通訳に関する国際共同研究が行なわれている。音声自動通訳は音声認識と言語翻訳と音声合成の複合的な高度の技術を必要とする。音声認識と音声合成は言語に個別的な特性を持つ。一方、言語翻訳は一つの国の言語を他の国の言語に変えなければならないので、関連国との相互協調関係が必要である。日本のATRと韓国のETRIは昨年、日本語と韓国語間の音声自動通訳のための相互緊密な協調体制を確立した。本研究はこのようなATRとETRIの協調体制下で遂行した韓国語と日本語間の言語翻訳に関するものである。

韓国語と日本語間の対話翻訳は韓国人が日本を、または日本人が韓国を旅行する時、多くの場面で起こりうる対話を対象とする。このための機械翻訳の枠組みは変換主導型機械翻訳(TDMT) [4、5]に従う。その理由はTDMTが既に“国際会議に関する問い合わせ”を対象とする日本語と英語間の対話翻訳への適用に成功し、また次のような特徴を持っているからである。

- TDMTは、翻訳を、原言語の表現をどうやって適切な目的言語の表現に写像すべきであるかを決定する問題と捉える。即ち、TDMTは多くの翻訳システムで採用している“解析-変換-生成”のパイプライン過程を必ずしも遵守せず、与えられた入力から目的言語に写像できる手掛かりを使って変換中心の翻訳を行なう。TDMTは変換を中心にパーシングや生成を簡潔に行なって翻訳を遂行する。このような方法はコンピュータの翻訳時間を大幅に短縮することができる [7、13]。翻訳時間は、実時間の翻訳が要求される対話翻訳において重要な問題である。
- TDMTは、用例に基づいて記述された経験的な変換知識を使って、入力に対する適切な対訳語彙と構造を決定する。TDMTにおいて、変換知識は、意味の同一性を伝達する形態・構文的な単位で、原言語の表現と目的言語の表現間の対応を用例に基づいて表す。このような用例に基づく変換知識の使用は、一般的な規則を使うことに比べて拡張性と頑健性が優れているという長所を持つ [14、4、5]。特にシステムの頑健性は文法的不適格文(文法から逸脱した表現)がよく現れる話し言葉の機械翻訳において必要不可欠である。

TDMTによる韓日の対話翻訳システムは、開発過程における一つの大きな特徴を持っている。それは既に開発した日英の機械翻訳システムを最大限再利用する戦略により短期間(1994年9月から1995年2月まで)で韓日双方向機械翻訳システムを開発したことである。韓日の翻訳システムのために再利用した既存のソフトウェアおよびデータは次のとおりである。

- 日英方向の翻訳システムで使った日本語の形態素解析モジュール
- 英日方向の翻訳システムで使った日本語の生成モジュール [1]
- 日英方向の翻訳システムで使った日本語シソーラス [14]
- 変換知識の記述と処理のためのツール [4、5]
- 多言語の翻訳のための統合インタフェース [9]

それから韓日双方向の対話翻訳のために新しく開発したソフトウェアおよびデータは次のとおりである。

- 韓国語形態素辞書
- 日本語形態素辞書
- 韓国語形態素解析モジュール
- 韓国語形態素生成モジュール
- 日韓方向の変換知識 (14 会話)
- 韓日方向の変換知識 (5 会話)
- 韓国語シソーラス

最終的に、我々は現在韓国語と日本語間の音声翻訳のための双方向言語翻訳のプロトタイプシステムを開発した。今後、我々は実際に海外旅行者の案内ドメインで収集したいろいろな会話文について大量の変換知識を構築し、翻訳に対する評価実験をしなければならない。このような評価実験を通じて韓日の対話翻訳の問題点を捕捉して、それに対する対策を用意することが必要である。本報告書の目的は、実際の評価実験を行なう前に、韓日の対話翻訳の特性と予想される問題点について考察することである。

本報告書は次のように構成されている。2章は韓国語と日本語間の言語学的な類似性について述べる。3章では両言語の言語学的な類似性が両言語間の対話翻訳にどんな影響を与えるかを検討する。具体的に Yokoyama 等 (1994) が検討を行なった日英機械翻訳での問題点と翻訳困難例文 [15、16] について、韓日機械翻訳の立場から再検討してみる。4章では結論を述べる。

2 韓国語と日本語間の言語学的な類似性

多くの文献 [2、3、8、10、12] に述べられているように、韓国語と日本語は言語学的に類似点が多い。その中で、両言語間の機械翻訳をかなり有利にすることができるのは次のような点である。

- 韓国語と日本語は言語構造がかなり類似している。即ち、両言語は語幹に派生接辞か語尾（助詞）が付いて単語を成す膠着語であるだけでなく、語順についてもほとんど同じ特徴を持つ。
- 韓国語と日本語は言葉の省略法が類似している。話し手が互いに主語か目的語が何かを知っていると、多くの場合、主語か目的語を省略する。従って、韓国語または日本語で主語か目的語がない文は、対話の言語外的な状況によってそれが理解されていると言える。

K: 예약을 부탁드립니다. (나는, 당신에게)

J: 予約をお願いします。(私は、あなたに)

K: 근데 하루에 얼마예요? (값이)

J: で、一泊においくらですか。(価格が)

K: 이만 삼천 엔정됩니다. (값이)

J: 二萬三千円ぐらいです。(価格が)

K: 비싸네요. (값이)

J: 高いですねえ。(価格が)

K: 지금 체크인할 수 있습니까? (나는)

J: 今チェックインできますか。(私は)

- 韓国語と日本語は言語文化と言語分布が似ている。即ち、日本語と韓国語は言語を使う二つの国の間の地理的な距離がかなり近いばかりでなく、中国の漢字から来た言葉を共に使っているものが多い。更に、韓国語と日本語は語幹の後に付いて語幹に多様な文法的機能を提供する活用語尾と助詞の言語分布が似ているという特徴を持つ [8]。このような理由等により、韓国語と日本語間には1対1に対応する単語（形態素）が多い。

次に、韓国語と日本語間の言語学的な類似性を実際に示す。下の例で JF340011.DAT は我々が両言語間の機械翻訳を実験するために収集した対話の中の一つである。英語

- J || [あの、すみません] ちょっと伺いたいんですが。
 K || [저, 미안하지만] 길 좀 묻겠는데요.
 J || 세션 문화회관へは、どう行けばよろしいでしょうか。
 K || 세종문화회관은 어떻게 가면 돼요?
- K || [아] 세종문화회관요?
 J || [아] 세션 문화회관ですか。
 K || 우선 이 길을 쭉 가면 세종로라는 큰 길이 나옵니다.
 J || まずこの道をずーっと行きますと、セジョンノという大きな道に出ます。
 K || 거기에서 지하도를 건너서 오른 쪽으로 쭉 가세요.
 J || そこから地下道を通して、上に出ましたら、右の方にずーっと行ってください。
 K || 한 백 미터 정도 걸어가시면 세종문화회관이 있습니다.
 J || 百メートルほど行けば、セジョン文化会館があります。
 K || 금방 알아 보실 겁니다.
 J || すぐお分かりになると思いますよ。
- J || この道をまっすぐ行って、セジョンノの地下道を通して右に行けばいいんですね。
 K || 이 길을 반드시 가서 세종로의 지하도를 건너서 오른 쪽으로 가면 되죠?
- K || 예, 그렇습니다.
 J || はい, そうです。
- J ||分かりました。
 K || 알겠습니다.
 J || どうもありがとうございます。
 K || 감사합니다。

の大文字 J は || の次の文が日本語であることを表し、K は韓国語であることを表す。会話文内で間投詞的な言葉は [] で表す。会話文内のスペースラインは話者の交代を表す。一人の話し手が話した内容は通訳者が文単位として目的言語に翻訳して相手に伝達される。

上のバイリンガル会話のそれぞれの文について形態素単位ごとに分割し、韓日の単語を対応させて表すと、図 1 のように示すことができる。図 1 で単語対応させにくい場合はそのまま表記した。それから通訳者が不必要に付け加えた言葉やなくてもいい言葉は括弧に入れて表記した。また、通訳者が直訳できるにもかかわらず意識した時は⇒の次に可能な直訳を表した。さらに、通訳者が翻訳するとき自然に話すために翻訳しなかったと判断される箇所は〈〉に入れて表した。

3 言語学的な類似性が持つ効果

この章は韓国語と日本語間の言語学的な類似性が両言語間の機械翻訳にどんな影響をもたらすかについて考察する。このために我々は Yokoyama 等 (1994) が発表した日英機械翻訳で翻訳しにくい日本語例文 [15] (以下、日英評価文と略称) と

J||[あの、すいません] ちょっと伺い たい んですが。
K||[저, 미안하지만] 길 좀 물 겠 는데요.
=>[저, 미안하지만] 잠깐 물어보 고 싶 은데요.

J||세종 문화회관 へ は、どう 行けば よろしい でしょうか。
K||세종 문화회관 <으로> 은 어떻게 가 면 되 어요?

K||[아] 세종 문화회관 요?
J||[아] 세종 문화회관 ですか。

K||우선 이 길 을 쪽 가 면 세종로 라는 크 ㄴ 길 이 나오 ㅂ니다.
J||まず この道 を ずーっと 行きます と、セジョンノ という 大きな 道 に 出 ます。

K||거기 에서 지하도 를 건너 서 오른쪽 으로 쪽 가 세요.
J||そこから 地下道 を 通って、(上に出ましたら)、右の 方 に ずーっと 行っ てください。

K||한 백 미터 정도 걸어가 시 면 세종 문화회관 이 있 습니다.
J||百メートル ほど 行け ば、セジョン文化会館 が あり ます。

K||금방 말 아보 실 겁니다.
J||すぐ お 分か り になると 思 いますよ。

J||この道 を まっすぐ 行っ て、セジョン口 の 地下道 を 通っ て 右 に 行け ば いい ン ですね。
K||이 길 을 반드시 가 서 세종로 의 지하도 를 건너 서 오른쪽 으로 가 면 되 죠?

K||예, 그렇 습니다.
J||はい, そう です。

J||分かり ました。
K||알 겠습니다。

J||どうも あり が とう ござ います。
K||<대단히> 감사합 니다。

図1 : 単語対応により表されたバイリンガル会話文

その評価 [16] について日韓機械翻訳の見地から再検討する。3.1 節では構造的曖昧性のために翻訳しにくい例文について検討する。3.2 節は意味的曖昧性のために翻訳しにくい例文について検討し、3.3 節は語彙的曖昧性のために翻訳しにくい例文について見る。3.4 節は日本語の特異な表現と文脈的曖昧性のために翻訳しにくい例文について調べる。最後に 3.5 節は生成に関連した難しさについて検討する。

3.1 構造的曖昧性 (Structural Ambiguity)

言語学的に異なる系統である日本語と英語間の機械翻訳で一番難しい点の一つは構造的曖昧性を解消することである。ここで構造的曖昧性はたいてい日本語と英語が互いに語順の違いによって生じる。例えば、日本語は修飾語が必ず被修飾語の前に現れ

るが、英語はたいていその後に見れる。このために日英機械翻訳では与えられた入力文について係り受け関係を正確に解析しなければならない。しかし、韓国語と日本語間の機械翻訳では両言語の語順がほとんど等しいので係り受け関係についての解析が精密でなくても翻訳することができる。では、次で日英評価文に見れる例文について分析してみる。下で我々は括弧を使って評価文の可能な解析構造を表した。

- (1a) 日本の動物と植物の歴史。
((日本の(動物と植物))の歴史))
((日本の動物)と(植物の歴史))
(日本の((動物と植物)の歴史))

(1a') 일본의 동물과 식물의 역사.

Note 1: (1a) は並列構造の構造的曖昧性を表す。それは三つの異なる構造を持っている。このそれぞれの三つの構造は日英機械翻訳でまったく違う翻訳結果を出力する。即ち、一番目構造が選択されれば、それは“The history of animals and plants in Japan”に翻訳される。二番目が選択されれば“Animals in Japan and the history of plants”に、三番目が選択されれば“The history of animals and plants of Japan”に翻訳される。三つの可能な構造から一つを選択して翻訳する場合にその結果は与えられた文脈によって正しくなることもあるが、まったく違う翻訳になることもありうる。従って、日英機械翻訳では入力文について可能な構造から文脈に適合する構造を選択しなければならない。しかし、(1a)を韓国語に翻訳する場合にはそのような問題が少ない。即ち、三つの可能な構造の中でどれを選択してもそれは同じ翻訳結果(1a')を出力する。これは多くのことを示唆すると思われる。特に、それは韓国語と日本語間の機械翻訳が計算コストが少なくすみ、かつ、翻訳の成功率が日本語と英語間の機械翻訳より高くなることを暗示すると見ることができる。

- (1b) 畜産物価格安定法。
((畜産物 価格) 安定法)
(畜産物 (価格 安定法))

(1b') 축산물가격안정법.

Note 2: (1b) についての翻訳は形態素解析と密接に関係する。一般的に漢字から成る長い複合語について、その構形成態素を正確に分割することは大変難しい。それについての形態素解析が正しくできても、日英機械翻訳ではそれぞれの構形成態素間の係り受けを正確に解析できなければ、正しい翻訳結果を得ることができない。もし(1b)が三つの形態素“畜産物、価格、安定法”から成っているとすれば、それは二つの可能な構造を持つ。それぞれについての翻訳結果は(1a)と同じように全く違う。一方、それに対応する韓国語は(1b')と同じように一つだけ存在する。更に、日本語と韓国語間の翻訳は(1b)のような長い複合語について形態素解析をあまり正確にしなくても

翻訳することができると思われる。

次に現れる例文についても同様の説明が可能であるから特別な点のみ説明を行なう。

- (1c) 大きな飛行機のプロペラ。
((大きな飛行機)のプロペラ)
(大きな(飛行機のプロペラ))
- (1c') 큰 비행기의 프로펠라.
- (1d) アメリカの希望によって、アメリカ国内で生産し、販売するという動き。
(((アメリカの希望によって、(アメリカ国内で生産し、販売する))という)動き)
((((アメリカの希望によって、アメリカ国内で生産し)、販売する)という)動き)
(((アメリカの希望によって、アメリカ国内で生産し)、(販売するという))動き)
- (1d') 아메리카의 희망에 따라, 아메리카 국내에서 생산하고, 판매한다고 하는 움직임.
- (1e) オペランドは必ずしも指定する必要はありません。
((オペランドは必ずしも指定する)必要はありません)
(オペランドは((必ずしも指定する)必要は)ありません)
- (1e') 오퍼랜드는 반드시 지정할 필요는 없습니다.
- (1f) ユーザは置換モードよりも挿入モードの方を高く評価する。
(ユーザは(置換モードよりも挿入モードの方)を高く評価する)
- (1f') 이용자는 치환 모드보다도 삽입 모드 쪽을 높게 평가한다.

Note 3: (1f) で“よりも～の方”のような日本語の比較表現を英語に機械翻訳することは“not only ~ but also”のような英語の非連続熟語を日本語に機械翻訳するときの難しさとほぼ同じ程度の難しさがあると思われる。しかし、韓国語はそれを非連続熟語として扱わなくて、その各構成要素を直訳しても等しい意味を伝達できる。なぜなら、日本語の“よりも”は韓国語の“보다도”に対応し、“の方”は“쪽”に対応するからである。

- (1g) タバコは人体に有害だと言われている。
((タバコは人体に有害だ)と言われている)
(タバコは(人体に有害だ)と言われている)
- (1g') 담배는 인체에 유해하다라고 알려져 있다.

Note 4: (1g) で日本語述語“言われている”によく対応する韓国語は(1g')に表すことのように“알려져 있다”である。ここで日本語の“言われている”は文法的に動詞語幹“言う”、受動接辞“れる”、進行の意味を持つ補助用言“ている”からなっている。しかし、それぞれの構成形態素についての韓国語の対応翻訳は“말하다”、“어지”、“고 있다”になって、その形態素の結合により生成された単語は“말하여지고 있다”になる。つまり、これは(1g')に表した翻訳に比べて文の意味をすぐに

理解することができない。これは韓国語と日本語間に存在する受動接辞の用法が一致しないので起きる。従って、このような受動表現に対する翻訳を自然にするためには構形成態素の個別的な翻訳ではなくて構形成態素の結合による翻訳にならなければならない。これは日本語と韓国語間の機械翻訳で両言語間の変換知識の単位、特に日韓の対訳辞書の見出し語が解析辞書のものを多少変更する必要があることを暗示する。

- (1h) 彼女は鳥に逃げられた。
- (1h') 그녀는 새에게 도망당했다.

- (1i) 東京の人は地震に慣らされている。
- (1i') 동경 사람은 지진에 익숙해져 있다.

Note 5: (1h) と (1i) のような日本語の被害の受け身は英語に表現することが非常に難しいとされている [16]。しかし、韓国語への翻訳はそれほど難しくはない。

- (1j) 組織における実際の作業は、複数人による共同作業という形態をとることが多い【にもかかわらず】、そのような共同作業を支援する環境はあまりなく、個人の生産性向上に重点が置かれているのが現状である。
- (1j') 조직에 있어 실제의 작업은, 복수인에 의한 공동작업이라고 하는 형태를 취하는 것이 많음【에도 불구하고】、그러한 공동작업을 지원하는 환경은 그다지 없고, 개인의 생산성 향상에 중점이 놓여 있는 것이 현상이다.
- (1k) 回路が短絡していると、ヒューズが溶断することは【もちろん】、場合によっては機器を損傷する恐れもあるので、電源を投入する前に回路抵抗を測定することを怠ってはいけません。
- (1k') 회로가 단락하여 있으면, 퓨즈가 용단하는 것은【물론】、경우에 따라서는 기기를 손상할 위험도 있기 때문에, 전원을 투입하기 전에 회로저항을 측정하는 것을 게을리해서는 안된다.
- (1l) この種の推論プロセスの具体的なメカニズムは現時点では明らかにされていない【が】、この推論機構の解明が、自然言語の柔軟な意味理解のモデル化の中心的な問題となる。
- (1l') 이 종의 추론 프로세스의 구체적인 메카니즘은 현시점에서는 명확하게 되어 있지 않【지만】、이 추론 기구의 해명이, 자연언어의 유연한 의미 이해의 모델화의 중심적인 문제로 된다.

Note 6: 実際の機械翻訳システムで長文に対する翻訳は計算的に多くの難しさを持つ。特に、語順が異なる言語間の機械翻訳では原言語の文に対する正確な構造が必要なので、そこから生じる難しさがかなり大きい。長文に対する対処方法としては一般的に (1j) に現れる“にもかかわらず”か (1k) の“もちろん”か (1l) の“が”の

ような特定の手掛かりを使って長文をいくつかの部分に分けて処理する方が良いと言われている。しかし、そのような分割処理は語順が全く異なる言語間の翻訳では分割のために語彙を補充すべきであったり、文の構造を変えなければならないという難しさがある。しかし、韓国語と日本語間の翻訳は分割による構造の変更が最小限にしか起こりえないのでそのような難しさは少ないと言える。

3.2 意味的曖昧性 (Semantic Ambiguity)

前節で我々は日英機械翻訳の難しさの一つである構造的曖昧性を日韓機械翻訳の境地から再検討した。ここでは意味的曖昧性について検討する。

- (2a) 妻は顔がきれいだ。
- (2a-1) 妻の顔はきれいだ。
- (2a') 아내는 얼굴이 이쁘다.
- (2a'-1) 아내의 얼굴은 이쁘다.

Note 7: 普通、日本語の談話助詞‘は’はその前の語が文の主題であることを表す。しかし、(2a)のようにその前の語が文の主題ではなくてその後の言葉が主題である場合、それに対する日英機械翻訳を正確にするためには、その文についての意味解析が必要である。なぜなら、日本語は文を構成する語の間の意味的關係(格關係)をそれぞれに適した助詞を付けて表すが、英語はそのような關係をたいてい語順によって表すからである。従って、日英機械翻訳では(2a)の文を(2a-1)の文と同じ意味として解析して翻訳せざるをえない。一方、韓国語は日本語の談話助詞‘は’にほとんど1対1として対応する助詞‘는(은)’があるので、日韓機械翻訳ではそのような意味解析をする必要がない。ここで韓国語の主題助詞‘는’と‘은’は音韻現象のために區別される異形態である。即ち、‘는’は直前の語が母音で終る場合に使われて、‘은’はそれが子音で終る場合に使われる。

- (2b) セミナーは日程を1日延長する。
- (2b-1) セミナーの日程は1日延長された。
- (2b-2) セミナーの日程は1日追加された。
- (2b-3) セミナーの日程を1日延長する。
- (2b') 세미나는 일정을 1일(하루) 연장한다.
- (2b'-1) 세미나의 일정은 1일(하루) 연장되었다.
- (2b'-2) 세미나의 일정은 1일(하루) 추가되었다.
- (2b'-3) 세미나의 일정을 1일(하루) 연장한다.

Note 8: (2b)は(2a)と似ている現象である。(2a)の‘は’がその前の語とその後語の間の所有關係を表すのに対し、(2b)は部分-全体關係を表すと言える。つまり、(2b)は日英機械翻訳で(2b-1)か(2b-2)か(2b-3)と同じ意味に翻訳さ

れなければならない。しかし、日韓機械翻訳ではそうする必要がない。

(2b) で日本語“1日”に対応する表現を二つ、即ち“하루”と“1일”として表した。それは日本語“1日”をどう発音するかによって変わるからである。それが“いちにち”として発音されると“1일”に対応し、“ついたち”として発音されると“하루”に対応するからである。しかし、(2b)の文脈では“いちにち”がより適切なので“하루”に括弧を付けた。また別な見地から考えると、これは日本語の表記によって生じる語彙的曖昧性であると言える。

(2c) 車が家になる。

(2c') 차가 집이 된다.

(2d) 彼は株で家を建てた。

(2d') 그는 주식으로 집을 지었다.

Note 9: (2c) と (2d) は途中省略が多くて直訳できても意味が通らなかつたり、意識も難しい例文である [16]。しかし、これに対する韓国語への翻訳はあまり難しくない。実際に、この例文に対する韓国語への翻訳は日本語の格助詞に対応する韓国語の格助詞を適切に決定するための解析だけが要求される。即ち、語彙的曖昧性を解消するための解析のみ要求される。

3.3 語彙的曖昧性 (Lexical Ambiguity)

語彙的曖昧性是一个の単語が一つ以上の意味として解析されるとき起きる。このような語彙的曖昧性は普通二つに分類される。一つは与えられた単語が二つ以上の文法範疇(品詞)を持っているとき起きる品詞の曖昧性で、もう一つは与えられた単語が二つ以上の意味を持っているとき現れる同音異議語の問題である。語彙的曖昧性についてのこのような二つの分類は主に原言語を分析する見地から行なわれる。しかし、我々はこの節で語彙的曖昧性を翻訳の見地から次のように見る:それは原言語の一つの単語について全く異なる意味を持つ二つ以上の目的言語の単語があるときに起きるのである。次の例のように日英評価文の語彙的曖昧性は多くのものが日本語と英語間の翻訳だけではなく日本語と韓国語間の翻訳でも同様に現れる。

(3a) 勉強【のために】、目が悪くなった。

(3a') 공부【때문에】, 눈이 나빠졌다.

(3b) 勉強【のために】、本を買った。

(3b') 공부【하기 위하여】, 책을 샀다

(3c) 彼女【のために】、プレゼントする。

(3c') 그녀【를 위하여】, 선물한다.

Note 10: (3a) と (3b) は日英評価文にある例文で、(3c) は日本語 “ のために ” に対応する韓国語がもう一つあることを示した例である。

- (3d) 満員のバスに乗っ【て】会社に行く。
 満員のバスに乗っ【て】遅刻した。
 満員のバスに乗っ【て】恥をかいた。
- (3d') 만원 버스를 타【고】 회사에 간다.
 만원 버스를 타【서】 지각했다.
 만원 버스를 타【서】 부끄러움을 받았다.
- (3e) まだ【着られた】服を捨てた。
 天皇陛下が【着られた】服を拝見した。
- (3e') 아직【입을 수 있는】 옷을 버렸다.
 천황 폐하가【입으셨던】 옷을 삼가해 보았다.

しかし、日本語と韓国語間の翻訳で語彙的曖昧性は両言語の言語分布が似ているから日英翻訳でよりかなり少ないことが調査されている [6]。表1は TDMT の日英翻訳システムと日韓翻訳システムにおいて、日本語格(係)助詞についての変換知識がそれぞれいくつあったかを調べた結果である。

表1：日英翻訳と日韓翻訳における助詞の訳し分け

日本語助詞	日英での対訳の数	日韓での対訳の数
X は Y	25	1
X が Y	9	1
X の Y	32	2
X を Y	12	1
X に Y	18	5
X で Y	18	5

3.4 文脈的曖昧性 (Contextual Ambiguity)

敬語、慣用句、省略等による文脈的曖昧性は、機械翻訳システムにとって現実的には解消するのが困難な部分である。しかし、このような文脈的曖昧性は言語文化の差異から来るものが多い。これは日本語と英語間の機械翻訳で文脈的曖昧性のために翻訳しにくい例文と分類した日英評価文の大部分が韓国語へ直訳できることを見ても分かる。

- (4a) この病気は注射しなければ直らない。
- (4a') 이 병은 주사하지 않으면 낫지 않는다.
- (4b) その時、彼は酒を飲んでいて。
- (4b') 그 때, 그는 술을 마시고 있었다.

- (4c) 佐藤は佐藤なりの仕事しかできない。
- (4c') 사토는 사토 나름의 일 밖에 할 수 없다.
- (4d) 今夜は私の番なので、それが悲しくて泣いているのです。
- (4d') 오늘 저녁은 내 당번이기 때문에, 그것이 슬퍼서 울고 있는 것이다.
- (4e) 危険な液体には色がつけてある。
- (4e') 위험한 액체에는 색이 붙어 있다.
- (4f) 息子には、現金を考えている。
- (4f') 아들에게는, 현금을 생각하고 있다.
- (4g) 首を切る。
- (4g') 머리를 자른다.

この以外にも対話文によく現れる日本語“はい”と“いいえ”は英語で単純に“はい”を“yes”に、“いいえ”を“no”に対応させることができない。なぜなら、先行する対話が否定疑問文である場合には質問者の意図によって“はい”が“no”に、“いいえ”が“yes”に翻訳されるからである [11]。理解を助けるために(4h)と(4i)にKunoの日本語文法書 [11]にある例文を引用しておく。つまり、これらを正確に翻訳するためには先行する対話を理解しなければならない。しかし、日本語と韓国語は肯定と否定に対する表現が等しいので(4h')と(4i')に表せるようにそれほど難しくない。日本語“はい”に対応する韓国語は必ず“예”で、“いいえ”に対応するのは必ず“아니오”である。

- (4h) 昨日、学校に行きませんでしたか。
はい、行きませんでした。 "No, I didn't."
いいえ、行きました。 "Yes, I did."
- (4h') 어제 학교에 안 갔습니까?
예, 안 갔습니다.
아니오, 갔습니다.
- (4i) 昨日、学校に行ったんじゃないですか。
はい、行きました。 "Yes, I did."
いいえ、行きませんでした。 "No, I didn't."
- (4i') 어제 학교에 간 것은 아닙니까?
예, 갔습니다.
아니오, 안 갔습니다.

しかし、日本語と韓国語間の対話翻訳においても、非常に解決が難しい文脈的曖昧性がある。それは“どうも、どうぞ”や“おはようございます、こんにちは、ごんばんは”などのあいさつ表現である。ここで“どうも、どうぞ”はそれに対応する韓国語の言葉がまったくないので日韓方向の翻訳を困難にする。そして、“おはようございます(午前)、こんにちは(午後)、ごんばんは(夜)”は、対応する韓国語の言葉がたった一つの同じものであるから、日韓方向の翻訳では問題がないが、韓日方向

の翻訳で問題がある。けれども、韓国語のあいさつの言葉“안녕하세요”について日本語でいつも“こんにちは”に翻訳してもあまり支障がなく、あいさつの言葉は比較的問題が少ないと言える。最も困難なのは“どうも、どうぞ”のような日本語表現に対する韓国語への翻訳である。

3.5 生成の問題 (Generation Problem)

日本語と英語間の機械翻訳で生成に関連する大きな問題は冠詞生成と時制の一致である。これらの問題は、日本語が英語のような冠詞を全く使わないこと、時制の一致に相当する現象があまりないことによって生じる。一方、日本語と韓国語間の機械翻訳にはこのような問題がない。なぜなら、韓国語は日本語と同じように英語のような冠詞がなく、また時制の一致もあまり起こらないからである。

韓国語と日本語間の対話翻訳で生成に関連する問題は、両言語が言語的に類似な点が多いので、たいていの場合、形態素の結合とそれによる音韻処理とに関連づけられる。構文処理が必要とされる部分は少ない。今まで検討されている構文処理が必要な現象に、次のような尊敬表現の一致がある。

(5a) このカードに【お】名前と【ご】住所を【お】書きください。

(5a') 이 카드에 【성함】과 주소를 써 주 【세】요(시+어요).

Note 11: 韓国語と日本語は尊敬の意味を持つ形態素の文法的特性が異なっている。韓国語は尊敬の意味を持つ形態素(例:시)が語尾の性質を持っているので用言語幹の後にだけ接続できるが、日本語(例:お、ご)は接頭辞の性質を持っていて用言の前に出てくる。また、日本語は用言が支配する名詞の尊敬表現をその名詞の直前に尊敬の接頭辞を置いて表すが、韓国語はその名詞自体に対応する尊敬の意味を持つ名詞があればそれを使い、なければそのまま表す。例えば、(4j)の日本語“名前”に対応する韓国語は“이름”であるが、その尊敬表現は“성함”である。しかし、日本語“住所”に対応する韓国語“주소”は尊敬表現が特にないのでそのまま表現される。

4 おわりに

本報告書で我々は変換主導型機械翻訳による日本語と韓国語間の対話翻訳システムの開発について簡潔に述べた。それから、我々は日本語と韓国語間の機械翻訳の実験、評価をしていくうえで予想される問題点と両言語間の機械翻訳に関連する特徴について考察した。具体的に我々は Yokoyama 等 (1994) が収集して評価した日英機械翻訳が難しい日本語例文について日韓機械翻訳の見地からそれを再検討してみた。

本研究を通じて我々は機械翻訳での問題点は目的言語に依存する面がかなり多いこ

とが分かった。日本語と英語間の機械翻訳に起きる多くの問題が日本語と韓国語間の機械翻訳では問題ではないことを確認した。実際に日英機械翻訳での多くの問題は語順と関連する構造的曖昧性と言語文化と関連する文脈的曖昧性から生じている。しかし、このような曖昧性は日韓機械翻訳では、両言語が語順がほとんど同じで言語文化も似ているので、深刻な問題にならないことを示した。結果的に、韓国語と日本語間の対話翻訳でのほとんどの問題は語彙的曖昧性にあると思われる。

結論的に、新しく開発した日本語と韓国語間の対話翻訳システムは既存の日英翻訳システムよりかなり高い翻訳正解率を示すことが予想される。今後、海外旅行者の案内ドメインで収集したいろいろな会話文について翻訳知識を構築し、翻訳に対する評価実験を行なっていく予定である。また、音声認識システムや音声合成システムとの統合を試み音声翻訳システムを構築することも今後の課題である。

参考文献

- [1] Akamine, S., O. Furuse, and H. Iida. 1994. Integration of example-based transfer and rule-based generation. Proceedings of ANLP-94.
- [2] An, D. U., G. C. Kim, and J. H. Lee. 1994. Corpus-Based Modality Generation for Korean Predicates. Literary and Linguistic Computing (to appear).
- [3] Choi, K. S., N. Nomura, K. Muraki, J. H. Lee, C. H. Kim, B. H. Choe, D. U. An, and G. C. Kim. 1988. A Consideration for Development of Korean Translation System Under Development Environment for Japanese Translation System. NLP 68-4, pp.1-8 (in Japanese).
- [4] Furuse, O. and H. Iida. 1992. Cooperation between Transfer and Analysis in Example-Based Framework. Proceedings of COLING-92, pp.645-651.
- [5] Furuse, O., E. Sumita, and H. Iida. 1994. Transfer-Driven Machine Translation Utilizing Empirical Knowledge. Transaction of Information Processing Society of Japan, Vol.35, No.3, pp.414-425 (in Japanese).
- [6] Furuse, O., S. Akamine, J. Kawai, D. B. Kim, and H. Iida. 1995. Dialogue Translation Mechanism Using Empirical Language-Knowledge - A system for Japanese-English and Japanese-Korean Dialogue Translation -. Proceedings of JNLP-95 (to appear; in Japanese).

- [7] Higuchi, T., K. Handa, N. Takahashi, H. Iida, E. Sumita, K. Oi, and H. Kitano. 1994. The IXM2 Parallel Associative Processor for AI. IEEE COMPUTER, pp. 53-63.
- [8] Hong, S. M. 1988. 韓・日語比較文法論 - 特殊助詞と副助詞 - . (韓国) 慶北大学出版部 (in Japanese).
- [9] Kawai, J. 1994. TDMT 翻訳知識作成ツールの利用方法. ATR TR-IT-0088 (in Japanese).
- [10] Kim, T. S. and S. Ura. 1992. Generation of Korean Based on Connection Forms of the Semantics in Japanese-Korean Machine Translation. Transaction of Information Processing Society of Japan, Vol.33, No.12, pp.1578-1588 (in Japanese).
- [11] Kuno, S. 1973. 日本文法研究. (日本) 大修館書店 (in Japanese).
- [12] Lee, S. H. and S. Ozawa. 1992. Inflecting Process of Verbs by Phonological Form for Machine Translation between Korean and Japanese. Transaction of Information Processing Society of Japan, Vol.33, No.12, pp.1565-1577 (in Japanese).
- [13] Sumita, E., K. Oi, O. Furuse, H. Iida, T. Higuchi, N. Takahashi, and H. Kitano. 1993. Example-Based Machine Translation on Massively Parallel Processors. Proceedings of IJCAI-93, pp. 1283-1288.
- [14] Sumita, E. and H. Iida. 1992. Example-Based Transfer of Japanese Adnominal Particles into English. IEICE TRANS. INF. & SYST., Vol.E75-D, No.4, pp.585-594.
- [15] Yokoyama, S., A. Kumano, Y. Mochizuki, M. Uemura, M. Kawagoe, Y. Tokioka, S. Ando, and H. Tanaka. 1994. Collection and Classification of Sentences Difficult to Machine-Translate. NLP 101-5, pp.33-40 (in Japanese).
- [16] Yokoyama, S., S. Tokunaga, M. Kameda, Y. Uchio, T. Ashizaki, T. Noda, H. Aizawa, H. Naka, H. Tanaka. 1994. Machine Translation and Evaluation of Japanese Sentences Difficult to Translate. NLP 101-6, pp.41-48 (in Japanese).