

目的指向の異言語間コミュニケーションにおける 機械翻訳の有効性の分析

--- 異文化コラボレーション ICE2002 実証実験から ---

小倉健太郎[†] 林良彦[†] 野村早恵子[‡] 石田亨[‡]
日本電信電話株式会社 NTT サイバースペース研究所[†]
京都大学大学院情報学研究科社会情報専攻[‡]

1. はじめに

日本、韓国、中国、マレーシアなど母国語が異なる異文化間での共同ソフトウェア開発などのコラボレーション^[1]に、現在の機械翻訳が有効に機能するかを ICE2002 では実験している。機械翻訳を介したコミュニケーションに現れる言語現象、そこでの翻訳の問題点、利用者が行う機械翻訳システムへの適応を分析した。また、投稿した母国語ごとの参加者のメッセージ理解度に関するアンケート結果とそれに関する考察について述べる。

2. TransBBS コミュニケーションの特徴

実験においては、コラボレーションのためのコミュニケーション手段として、利用者が母国語で書いたメッセージを他の参加者の母国語に翻訳して投稿できる多言語電子会議システム TransBBS と、HTML 形式のソフトウェア開発ドキュメントを多言語に翻訳して閲覧できる文書翻訳システム TransWEB が使用された^[2]。本稿では、TransBBS におけるコミュニケーション（主に、日本語での投稿部分）を分析の対象とした。

TransBBS を介した目的指向コミュニケーションの特徴は以下の通りである。

- ・ 文法的制約が弱く、文脈依存性が高い
- ・ 様相・敬語表現が多い
- ・ 呼びかけ、挨拶が多い
- ・ 依頼、確認、激励表現が多い
- ・ 言い淀み、言い間違い、言い換えが少ない

この他、話し言葉固有の表現（例えば、終助詞「ね」、「もの」、「こと」、「方」の使用）も多く、話し言葉^[3]の特徴を持っている。

しかし、会話などに頻繁に現れる「言い淀み、言い間違い、言い換え」はほとんどない。また、「あの」、「えー」などのフィラーはない。

Analysis of Machine Translation Effectiveness in Goal-directed Communication between Different Languages
--- From Intercultural Collaboration Experiment (ICE2002) ---
Kentaro Ogura, Yoshihiko Hayashi, Saeko Nomura and Toru Ishida

[†] NTT Cyber Space Laboratories

[‡] Department of Social Informatics, Kyoto University

「しちゃう」、「しとく」、「してる」、「じゃ」などの縮訳表現や方言の使用も少ない。

ソフトウェア設計フェーズでは、円滑なコラボレーションを行うために、人間関係を円滑にする呼びかけや挨拶表現や激励表現が多い。それに対して、ソフトウェア開発フェーズでは、質問、依頼、確認の表現が多い。

3. ディスカッション翻訳の問題点

TransBBS を介したコミュニケーションの翻訳をディスカッション翻訳と本稿では呼ぶことにする。「文法的制約が弱く、文脈依存性が高い」ことから生じる翻訳の問題点としては、主語や目的語が省略された場合、適切に補って翻訳することの困難性や助詞の省略を適切に認知できないことなどが挙げられる。

<主語の省略（不適切な主語の補完）>

例 1: とても驚きました。（原文）

-> It was very much surprised. （翻訳結果）

<助詞「を」の省略>

例 2: 何を開発しようとしているか報告してください。

-> What is going to be developed or please report it.

例 2 では、問いかけの終助詞「か」を列挙の副助詞「か」と解釈している。また、例 3 のように、完全な文になっていない表現も、翻訳システムが対処できていない場合がある。

<不完全文>

例 3: ICE2002 へようこそ。

-> It is welcome to ICE2002.

話し言葉は様相や敬語表現などが頻繁に使用され文末表現が複雑になっている。様相が複合している場合や文末複合表現が適切に翻訳できない場合がある。

<「使役+依頼の複合様相処理」>

例 4: みなさんの意見を聞かせてください。

-> Please let listen to everybody's opinion.

<文末複合表現>

例 5: このプロセスは、僕が投稿する日本語の文章を明

確にするのに役立ちます。

-> This process is useful for that * make the sentence of Japanese that I contribute clear.

話し言葉は書き言葉に比べ副詞句表現^[4]の使用頻度が高い。副詞句は名詞や動詞などの翻訳に比べて翻訳が難しい。特に、複合的な副詞句は、係り受け解析や適切な訳語の用意が難しい。

<副詞句表現>

例 6: 私はもう少し後からでもいいと思います。

-> I think that I am good even afterwards to a slight degree.

例 7: 残念なことに、私は中国語を理解できない。

-> I can not understand Chinese to being regrettable.

ディスカッションでよく現れる呼びかけ表現については、あまり複雑でない呼びかけを含む文は、一般の翻訳システムでは比較的適切に翻訳できる場合が多い。本実験で利用したシステムでは、呼びかけを含む文への対処が不十分であり、例 8 のような翻訳失敗が見られた。しかし、例 9 のように呼びかけを含む確認表現のような場合は一般の翻訳システムでうまく翻訳できないことが多い。

<呼びかけ&挨拶> (システム固有の問題)

例 8: 早恵子さん、こんにちは。

-> The ----- of Saeko.

<呼びかけ&確認>

例 9: 皆さん、分かりますか?

-> Is it understood everybody?

「もの」、「こと」、「方」はそのまま“thing”や“direction”と訳すと不適切な翻訳になる場合が多く、特別な構造に変換したり、複合表現の一部として翻訳する必要がある。

「方」には、「ほう」と「かた」の曖昧性もある。例 12 では「ほう」と解釈すべきを「かた」と解釈している。

<「もの」>

例 10: 英語に直接翻訳を行っているものもあります。

-> It is also in the English what it is being translated directly.

<「事が多い」>

例 11: 上手に翻訳される事が多い。

-> There is a lot of what is translated well.

<「方」>

例 12: 第 1トラックでは、初めの方はかなり頑張って書いていました。

-> The first person considerably did their best and wrote in the first track.

ディスカッション翻訳特有の問題ではないが、

外国人名などは未知語になってしまい、適切に翻訳できない。

<未知語>

例 13: 私は、最近、グアックギョンテク監督の映画を見ました。

-> I had seen Supervisor -----'s movie recently.

また、長文が適切に翻訳できない問題や、訳語選択が適切に行われないといった問題は数多くみられた。

4. 利用者の翻訳システムへの適応

TransBBS には、投稿する前に自分が投稿する文がどのように訳されるか確認する機能が付いている。通常、何回かの文の書き換えを行った後投稿するという自己主導型のリペアがなされる。利用者が日本人の場合は、英語への翻訳結果を確認して、その結果により書き換えを行う場合が多い。書き換えには以下の 2 種類がある。

1. 利用者の機械翻訳適応
2. 文章内容の精練

「利用者の機械翻訳適応」は、翻訳が良くない時に、より良い翻訳になるように書き換えるものである。「文章内容の精練」は、必要な情報の追加や、冗長な情報を削除するものである。

市販の機械翻訳を利用する場合、より良い翻訳を得るために、以下のように文を書くことが推奨されている。

- 文を短くする (50文字以内、接続助詞の使用をさける)
- 長い修飾語は付けない
- 節間の接続関係の明確化
- 主語や目的語を省略しない
- 述語を省略しない
- 解釈の曖昧性を避ける (係り受けを明確にする、同音異義語をさける、漢字を使用する)
- 婉曲的な表現を避ける (例: となっている、することになる)
- 抽象性の高い表現は避ける (例: もの、こと)
- 未知語をなくす (辞書登録)

実際に行われた利用者の翻訳システムへの適応で多かったのは、ディスカッション翻訳特有の問題への適応というよりも、頻度の高い一般的な翻訳問題への適応であった。

語句レベルの書き換えにより、適切に訳語選択できない場合に適切な訳語ができるように書き換えている。

例 14: <動詞の書き換え>

あなたの意見を聞かせて下さい。 →
あなたの意見を教えて下さい。 →
あなたの意見を伝えて下さい。

例 15: <名詞の書き換え>

できるだけ TransBBS の上で会議を行って下さい。 →
できるだけ TransBBS の上でミーティングを行って下さい。

副詞（句）の場合は、書き換える場合と削除する場合があった。

例 16: <副詞句表現の書き換え>

日本語で通じてしまうと思うと、思わずそのまま投稿してもよいのではと感じてしまうのです。

→

日本語で通じてしまうと思うと、すぐ投稿してもよいのではないかと感じてしまうのです。

例 17: <副詞の削除>

是非、使ってください。 →
使ってください。

「是非」の訳“by all means”が使用される文脈では強すぎるのでさけたものと思われる。

語句の書き換えの場合、例 18 のように直接翻訳したい英語のアルファベットに書き換えることがある。また、未知語や解析誤りを引き起こす可能性のある名前などの固有名詞は、“Kaname”のように最初から英語のアルファベットで記述する適応が見られた。

例 18: <英語化>

技術オーガナイザの Kaname です。 →
technical organizer の Kaname です。

英語化のためのアルファベット表記への書き換えは諸刃の剣である。アルファベット表記によりその部分は正しく英語に翻訳できる。しかし、未知語になるため、他の語の訳語選択に悪影響を及ぼしたり、正しい構造に変換できない可能性がある。また、英語への翻訳は良くなるが、他の言語への翻訳を悪化させる危険がある。

長文に対しては、文分割を行う適応が見られた。接続助詞の前や連用中止のところで分割する場合が多い。分割した場合は、接続詞などにより接続関係を明確に示すことも行われていた。

例 19: <文分割>

日本語で通じてしまうので、すぐ投稿してもよいのではないかと感じてしまうのですが、他のチームから何を議論しているのかわからないのは問題です。

→

日本語で通じてしまうので、すぐ投稿してもよいのではないかと感じてしまいます。しかし他のチームが何を議論しているのかわからないことは問題です。

主語や目的語などが省略されると適切な翻訳ができない問題に対しては、それらを利用者が

補うという適応が行われている。また、同様な問題として、英語で必要となる所有格ができるように、「の」格を補うという適応も見られた。

例 20: <主語の補完>

日本チームのメンバーです。 →
私は日本チームのメンバーです。

例 21: <「の」格の補完>

意見を聞かせて下さい。 →
あなたの意見を教えて下さい。

「もの」、「こと」の問題に関しては、例 22 のように、抽象的に表現した「もの」をより具体的な「サービス」に書き換える適応があった。しかし、「ものです」、「事が多い」のような特別な構造に変換したり、複合表現の一部として翻訳する必要のあるものからの書き換えによる適応は見られなかった。

例 22: <抽象的な「もの」>

Special Blend は、二つのサービスのうち、翻訳結果のよいものを選択して作られたものです。 →
Special Blend は、二つのサービスのうち、翻訳結果のよいサービスを選択して作られたものです。

呼びかけ表現に対しては、呼びかけ部分を分割することで、呼びかけの翻訳の問題を回避する適応が数多く見られた。呼びかけ表現は、例 23 のように「、」で区切られて表現される。いろいろな場合に、「、」の前後の関係を呼びかけであると適切にシステムが認識することは難しい。また、箇条書きなど特殊な文構造を、普通の文に書き換える適応もあった。

例 23: <呼びかけ部の分割>

早恵子さん、藤代くん、ありがとう。 →
早恵子さん。藤代くん。ありがとう。

例 24: <箇条書きの名詞句を文に変更>

Saeko & Naomi: コミュニケーションコーディネーター。
Kaname: 技術支援者。

→

Saeko & Naomi は、コミュニケーションコーディネーターです。
Kaname は、技術支援者です。

利用者の翻訳システムへの適応は、比較的簡易な書き換えで、かつ効果が明白なものに限られており、市販の機械翻訳で推奨されている文の書き方を実践したものになっている。

TransBBS では、利用者の適応のためのシステムの機能として、翻訳する必要がない語句を指定する機能がある。また、使用する翻訳システムを変更することにより、より翻訳の質の高い翻訳システムを利用者が選んで翻訳する機能がある。利用者の翻訳システムへの適応は、翻訳品質の改善とともにユーザにかなり負担をかけ

ることにもなっている。

5. メッセージ理解度

異文化コラボレーション実験は、日本、韓国、中国、マレーシアの間で行われた。参加者は、母国語で投稿し、他の国からの投稿メッセージは、翻訳された母国語や英語を読むことになる。当然、言語対毎に翻訳の精度の違いがあり、メッセージの理解度に影響がある。どの程度メッセージを理解できたかのコラボレーション参加者へのアンケート結果から、どの程度の翻訳レベルにあれば、コラボレーションを達成できるのかおおまかに知ることができる。

表1、表2、表3、表4はそれぞれ、日本語（日本チームから）、中国語（中国チームから）、韓国語（韓国チームから）、英語&マレーシア語（マレーシアチームから）で投稿されたメッセージが、参加各国のメンバーにどの程度理解されたかを示したものである。

表1. 日本語投稿のメッセージ理解度

チーム	A	O	S	R	N	無
中国	13	50	25	13	0	0
韓国	14	43	29	0	14	0
マレーシア	0	0	75	0	0	25

表2. 中国語投稿のメッセージ理解度

チーム	A	O	S	R	N	無
日本	0	100	0	0	0	0
韓国	0	0	71	29	0	0
マレーシア	0	0	75	25	0	0

表3. 韓国語投稿のメッセージ理解度

チーム	A	O	S	R	N	無
日本	33	67	0	0	0	0
中国	0	0	63	38	0	0
マレーシア	0	0	75	0	0	25

表4. 英語&マレーシア語投稿のメッセージ理解度

チーム	A	O	S	R	N	無
日本	0	67	33	0	0	0
中国	0	25	50	25	0	0
韓国	0	14	29	29	14	14

A: Always, O: Often, S: Sometimes, R: Rarely, N: Never, 無: 無回答
(本データは、実験の第1トラックでのアンケート結果なので、利用者の書き換えによる適応はない。)

表5. 言語ペアごとの翻訳方式

対象言語	日本語	中国語	韓国語	マレーシア語	英語
日本語		直接	直接	直接	直接
中国語	直接		2段	2段	直接
韓国語	直接	2段		2段	2段
マレーシア語	2段	2段	3段		直接
英語	直接	直接	2段	直接	

日本語で投稿されたメッセージに関する理解度は直接翻訳なので概ね高い。マレーシアチームの理解度が低いのは、日マレーシア翻訳がブ

ロタイプシステムなので翻訳精度がまだ低いからであると思われる。

同様に直接翻訳される中国語および韓国語で投稿されたメッセージは日本での理解度が高い。マレーシアからの投稿は、英語で行われた場合も多く、日本での理解度はそれほど低くない。

中国、韓国、マレーシア間の理解度が低いのは、英韓、韓英、韓中、中韓翻訳が別の言語（日本語や英語）への翻訳を介した2段翻訳となっているため、翻訳精度が低いからである。日本語と韓国語は言語としての類似性が高く、日韓/韓日翻訳の精度は高い。これもメッセージ理解度に現れている。

タスクが目的指向のコミュニケーション環境では、相手言語への翻訳の精度が良くない場合においても、タスク・共有知識や英語参照によるメッセージ理解が可能である。また、理解できなかったメッセージ内容に関する質問などの他者主導型リペアを通じた理解もあった。

本実験では、機械翻訳を介したディスカッションをしながら、実際に異文化コラボレーションのためのソフトウェアを開発した。ソフトウェアの完成により、機械翻訳を介したコミュニケーションが異文化コラボレーションに有効であったことを確認できた。

6. おわりに

ある程度の翻訳の質（既存の日英、日韓、日中翻訳システムの翻訳の質）があり、タスクが目的指向で、かつ、利用者の翻訳システム適応が期待できる環境では、機械翻訳がコミュニケーション手段として可能性があることを確認した。また、機械翻訳を介した異文化コラボレーションで行われるディスカッションにおける機械翻訳の問題点を確認した。さらに、そのような環境で行われる機械翻訳システムへの利用者の適応状況を明らかにした。

参考文献

- [1] 野村, 船越, 山下, 安岡, 石田 “機械翻訳を介したオープンソースソフトウェア開発: Intercultural Collaboration Experiment 2002”, 情報処理学会第65回全国大会, 2T6-3, 2003
- [2] 船越, 山本, 藤代, 野村, 石田 “異文化コラボレーション支援システム的设计”, 情報処理学会第65回全国大会, 2003
- [3] 松本, 伝. “話し言葉の形態素解析”, 2001-NL-143-9, 2001
- [4] 小倉, ボンド “日本語副詞の日英対照分析とそれに基づく機械翻訳”, 電子情報通信学会論文誌, Vol. J82-D-II, No. 11, pp. 2048-2057, 1999