

J-038 日本語学習初学者を対象としたやさしい日本語文と機械翻訳文の理解のしやすさ

Comparison of Ease of Understanding between Plain Japanese and Machine Translation for Japanese Beginner

吉野 孝[†]
Takashi Yoshino

宮部真衣[‡]
Mai Miyabe

1. はじめに

やさしい日本語とは、主に災害時などに日本語があまり得意ではない外国人でも理解できる言葉として、通常の日本語を書き換えたものである。やさしい日本語によって情報の提示を行う場合、在日外国人に対してある程度の日本語能力を要求するが、文の内容は正しく、読み易いものであるといえる。しかし、やさしい日本語への書き換えには、わかりやすい単語の利用や言い換えなどの相応のコストがかかる。一方、機械翻訳によって母語での情報提示を行う場合、在日外国人に対して日本語の習得を要求しないが、翻訳精度によっては文意の理解に大きな負荷がかかったり、そもそも理解出来なかったりする場合もある。しかし、やさしい日本語の作成と比較して、日本語から母語への機械翻訳は自動で行うことが可能であり、やさしい日本語への書き換えに比べてコストは小さい。

このように 2 つの支援手法にはそれぞれメリットとデメリットがあるが、在日外国人に対してそれぞれの手法で情報を提示した場合の理解のしやすさについて比較した研究はまだ行われていない。本稿ではやさしい日本語による情報の提示と機械翻訳による母語情報の提示では、在日外国人にとってどちらの方が理解しやすいかを明らかにするために、調査と分析を行った。

2. 関連研究

日本語があまり得意ではない外国人のための言葉として、やさしい日本語が研究されている [1]。この研究は特に災害時や緊急時に外国人でも日本語を理解できるようにすることを目的としたものであり、語彙を制限したり 1 文に主語・動詞のペアを一つにしたりするなど、日本語の文を簡単にするためのルールが定義されている。また、NHK は外国人のためにやさしい日本語を用いたニュースサービスの提供を行っている [2]。このサービスでは、ルビ付きのやさしい日本語で日本の主要なニュースを読むことが出来るほか、音声と動画でもニュースを確認することが可能である。

やさしい日本語を用いた情報提示では、外国人の母語に依存することなく一定に効果を発揮すると考えられるが、ある程度の日本語能力を必要とする。本研究

表 1: 機械翻訳試験文 (原文)・機械翻訳文・やさしい日本語文のペアの例

機械翻訳試験文 (原文)	機械翻訳文	やさしい日本語文
梅雨には天気が変わり易いことに留意することが必要だ。	那的候的他, 完全即使孩子也光。	梅雨は 天気が 変わりやすいので 注意した方が 良い。
書店はもともと利益が低いことに加えて、出版物の需要が鈍化しているため苦しい経営を余儀なくされている。	店因根本利益低事之外又加上, 出版物的需求化被不得已困苦の。	本を 売る店ほ もともと 入ってくる お金 が 少ない。それに 本が 売れないので 店を 続けることが 難しい。

では在日外国人にとって、機械翻訳文による母語情報の提示とやさしい日本語による情報の提示ではどちらの方が文意の理解が容易かを調査および分析する。

3. やさしい日本語文と機械翻訳文との理解のしやすさの比較実験

3.1 評価テキスト

我々は以前にやさしい日本語への書き換えによる機械翻訳結果の影響を検証するために実験を行っている [3]。その際に生成されたやさしい日本語文を比較実験に用いる。具体的には、機械翻訳試験文 [4] から 40 文を用いて、「やさしい日本語」作成のためのガイドライン [5] とリーディングチュー太 [6] を参考に、日本人大学生 3 名が手作業でやさしい日本語へと書き換える作業を行った。作業で用いられた機械翻訳試験文は 40 文で、生成されたやさしい日本語文は合計 120 文である。今回の実験では、機械翻訳試験文 (日本語文) 1 文に対応する機械翻訳文 (中国語文) 1 文とやさしい日本語文 3 文[§]のそれぞれの組み合わせを作成し、合計 120 ペアを実験に使用した。表 1 に実験に使用したペアの例を示す。

3.2 実験方法

Web システムを用いて、やさしい日本語文と機械翻訳文の理解のしやすさについての評価を行った。被験者は 20 代の中国人留学生 6 名で、日本語の学習期間は 1 年未満 (3 ヶ月から 9 ヶ月) であり、実用日本語検定において E 級および F 級[¶]である。本実験では機械翻

[§]ある原文を同じ条件で書き換えを行なっても、人によって書き換え方が違っていたため、3 名が作成したものをそれぞれ用いる。

[¶]外国人の日本語能力を客観的に測定する試験において、初級者レベルである。

[†]和歌山大学 システム工学部

[‡]京都大学 デザイン学ユニット

表 2: やさしい日本語文における被験者の理解のしやすさ

	評価値 (5 段階)					合計
	1	2	3	4	5	
初学者	6 %	30 %	26 %	33 %	4 %	100 %
上級者	0 %	4 %	8 %	54 %	34 %	100 %

評価基準: 1: 強く同意しない, 2: 同意しない, 3: どちらともいえない, 4: 同意する, 5: 強く同意する

訳システムとして高電社の J-Server[7] を用いた。

順序効果を考慮して, 6 名の被験者のうち 3 名はシステム上でのやさしい日本語文と機械翻訳文の表示順序を逆にした。評価テキストは被験者ごとにランダムに表示した。また, 画面の上部には日本語を翻訳者に依頼して中国語に翻訳した文を表示している。これはやさしい日本語および機械翻訳を提示した場合, 日本語初学者は元の文意が全くわからない可能性があるためである。

3.3 実験結果と考察

やさしい日本語文の評価をまとめたものを表 2 に, 機械翻訳文の評価をまとめたものを表 3 に示す。文献 [8] の上級者 (中国人留学生 10 名, 日本語レベル最上級の N1 取得者) の結果を示す。

やさしい日本語文の評価 (表 2) において, 初学者は評価 4 および評価 5 の合計が 37 % であったが, 上級者は 88 % であった。初学者にとっては, やさしい日本語であっても, 理解しにくいと感じていることが多いことがわかった。

機械翻訳文の評価 (表 3) においては, 評価 1 および評価 2 を理解しにくい, 評価 4 および評価 5 を理解しやすいとまとめると初学者と上級者では大きな差は見られなかった。

表 4 に, やさしい日本語と比較して機械翻訳文の方が理解しやすいと判断した文の数とその割合を, 機械翻訳文の翻訳精度評価値毎に示す。機械翻訳文の評価は, 予め翻訳者 3 名に適合性について 5 段階で評価¹⁾してもらったものである。評価者の判断もばらつきがあるため, 中央値を用いている。表 4 から, 中央値 3 および中央値 4 の文の場合, 約 4 割のやさしい日本語と機械翻訳文のペアにおいて, 機械翻訳文の方が理解しやすいと判断していることが分かった。

これらの結果から, 初学者および上級者にとって, やさしい日本語文の方が機械翻訳文より負担無く, 文意を理解できていることがわかったが, 機械翻訳の精度が高いものに関しては, 約 4 割は, やさしい日本語よりも機械翻訳の方が理解しやすいと感じていることが分かった。

¹⁾ 1: 全く違う意味, 2: 雰囲気は残っているがもとの意味と違う, 3: 意味は代替つかめる, 4: 文法などに問題があるが, だいたい同じ意味, 5: 同じ意味

表 3: 機械翻訳文における被験者の理解のしやすさ

被験者	評価値 (5 段階)					合計
	1	2	3	4	5	
初学者	3 %	41 %	39 %	16 %	2 %	100 %
上級者	22 %	27 %	24 %	21 %	6 %	100 %

評価基準: 1: 強く同意しない, 2: 同意しない, 3: どちらともいえない, 4: 同意する, 5: 強く同意する

表 4: やさしい日本語と比較して機械翻訳文の方が理解しやすいと判断した文の数 (機械翻訳文の翻訳精度評価値別)

	中央値 1	中央値 2	中央値 3	中央値 4	合計
文数 (文)	6	39	39	18	102
「機械翻訳文の方が理解しやすい」の平均 (文)	1.7	11.7	16.2	7.0	36.5
割合	28%	30%	41%	39%	36%

4. おわりに

本稿では, 日本語初学者を対象に, やさしい日本語文による情報の提示と機械翻訳文による母語情報の提示ではどちらの方がより文意の理解に役立つのかについて実験を行った。実験の結果, 日本語初学者にとって, 母語で書かれた機械翻訳文よりもやさしい日本語の方が理解しやすいことが分かった。ただし, 機械翻訳の精度の高いものについては, やさしい日本語よりも機械翻訳の方が理解しやすいと感じるものが 4 割程度あることが分かった。今後, 各文の難易度などの影響を考慮し, 詳細に実験結果の検討を行う。

謝辞

本研究の一部は, 独立行政法人科学技術振興機構研究成果最適展開支援事業 (A-STEP) 探索タイプ「検索エンジンと機械翻訳を用いた多言語用語間における文化差検出サービス」の補助を受けた。

参考文献

- [1] 庵功雄, 岩田一成, 筒井千絵, 森篤嗣, 松田真希子: 「やさしい日本語」を用いたユニバーサルコミュニケーション実現のための予備的考察, 一橋大学国際教育センター紀要, 1, pp.31-46 (2010)。
- [2] 田中秀輝, 美野秀弥: やさしい日本語によるニュースの書き換え実験, 情処研報, 自然言語処理, 2010-NL-199 (11), pp.1-8 (2010)。
- [3] 東拓央, 宮部真衣, 吉野孝: やさしい日本語化による翻訳リペア支援効果の検証, 信学技報, 人工知能と知識処理, Vol.111, No.251, pp.19-24 (2011)。
- [4] NTT Natural Language Research Group, <http://www.kecl.ntt.co.jp/icl/mtg/resources/index.php>
- [5] 「やさしい日本語」作成のためのガイドライン, <http://human.cc.hirosaki-u.ac.jp/kokugo/ejgaidorain.html>
- [6] 日本語読解学習支援システム リーディングチュウ太, <http://language.tiu.ac.jp/>
- [7] KODENSHA, <http://www.kodensha.jp/>
- [8] 東 拓央, 宮部真衣, 吉野 孝: 在日外国人を対象としたやさしい日本語文と機械翻訳文間の理解のしやすさの比較, 信学技報, 人工知能と知識処理, Vol.112, No.435, pp.91-96 (2013)。