

機械翻訳へのユーザの適応と書き換えへの教示効果に関する分析

山下直美^{†1} 坂本知子^{†2} 野村早恵子^{†3}
 石田亨^{†4} 林良彦^{†5}
 小倉健太郎^{†6} 井佐原均^{†7}

機械翻訳を介したコミュニケーションを通じて相互理解を実現するためには、翻訳精度の向上とともに相互作用性の向上が重要である。我々は機械翻訳に対するユーザの適応行動の1つである原文の書き換えに注目した。本論文では、ユーザが母国語だけを用いて原文の書き換え作業を行う方法として折り返し翻訳を検討し、折り返し翻訳を用いてユーザが書き換え作業をする際の作業量を減らす支援方法を考案する。本研究でユーザの折り返し翻訳作業に関する実験を実施、分析した結果、以下の知見を得た。1) 母国語に関する知識が豊富なユーザほど機械翻訳に容易に適応でき書き換え作業量が少なかった。2) ユーザに事前に「良い翻訳結果を得るためのルール集合」の教示を行うと、母国語に関する知識が豊富でないユーザも機械翻訳に容易に適応できるようになり、書き換え作業量が大幅に減った。3) ただし、原文をどのように変更すべきかを明示しない「操作自由型ルール」に対する教示効果は薄く、これらのルール獲得にかかる書き換え作業量は大きく減少しなかった。原文をどのように変更すべきかを明示した「操作指示型ルール」に対する教示効果は高く、これらのルール獲得にかかる書き換え作業量は大きく減少した。4) ルールの教示は、母国語に関する知識が中位のユーザに最も効果的であった。

Analyzing User Adaptation toward Machine Translation Systems

NAOMI YAMASHITA,^{†1} TOMOKO SAKAMOTO,^{†2} SAEKO NOMURA,^{†3}
 TORU ISHIDA,^{†4} YOSHIHIKO HAYASHI,^{†5} KENTARO OGURA^{†6}
 and HITOSHI ISAHARA^{†7}

Translation refinement is often observed when users communicate via machine translation systems. In this study, we analyzed user's translation refinement process through a controlled experiment. In the experiment, users translated sentences using a Japanese-English-Japanese turn-back translation. From the analysis, we discovered the following results: 1) The more knowledge users had about the source language, the better users could refine the original text, 2) Rule instruction was very effective in user's adaptation. Users who were reminded of the rules refined the original text ahead of other users, 3) Instructing operational rules were effective in helping user's adaptation, while conditional rules were not as much effective. 4) Rule instruction was most effective to those who had midium knowledge in their source languages.

†1 NTT コミュニケーション科学基礎研究所
 NTT Communication Science Laboratories

†2 野村総合研究所
 Nomura Research Institute

†3 カリフォルニア大学サンディエゴ校
 University of California San Diego

†4 京都大学情報学研究所社会情報学専攻
 Kyoto University

†5 大阪大学言語文化研究科
 Osaka University

†6 NTT サイバースペース研究所
 NTT Cyberspace Laboratories

†7 通信総合研究所
 National Institute of Information and Communications
 Technology

1. はじめに

近年のインターネットユーザの増大と国際化の潮流により、機械翻訳システムの利用機会は増加の一途をたどっている。また、その利用目的も、単に他言語で書かれた文書の翻訳であったものから多言語間コミュニケーションへと多様化が進行しつつある。国際的な協調作業が年々増加する中、相互理解を実現する機械翻訳システムが切望されている。

これまで機械翻訳の研究の大半は文書の翻訳精度向上を主体としたもので、コミュニケーションの観点からの研究は少ない。話し言葉をうまく翻訳する研究な

どもされているが、依然として正確な翻訳が難しく、翻訳の質の急激な向上は期待できない。同じ翻訳システムを使うにしても、ユーザ側がその使い方を工夫し相互作用性の向上を図ることで相互理解を促進できる可能性がある¹⁾。たとえば、野村らが実施した異文化コラボレーション実験(ICE2002)では、日中韓馬英の5言語の翻訳機能付き掲示板システム上で多言語コミュニケーションが行われたが、ユーザはメッセージを投稿する際、英語の翻訳結果を参照しながら原文を繰り返し書き換えていた^{2),3)}。このような書き換え作業は、機械翻訳に対するユーザの適応の一種であり、この作業によってメッセージの翻訳精度が向上することが示されている⁴⁾。

このように情報を正確に伝えることが重要なタスクを非同期型のコミュニケーションで行う場合、ユーザの書き換え作業は機械翻訳を介した相互理解を構築するうえで有効な手段である。しかし、ユーザの書き換え作業にはいくつかの問題点が存在する。問題点の1つ目は、書き換え作業に多大な作業コストがともなうことである。野村らの実施した異文化コラボレーション実験(ICE2002)ではユーザが1メッセージの投稿あたり平均5回の書き換え作業を行っており、その作業コストが非常に高い様子が観測された。問題点の2つ目は、英語の言語知識を持ったユーザにしか書き換え作業が実施できなかったことである。そのため、母国語だけを用いて書き換え作業が行える仕組みを準備し、さらに書き換え作業にかかる作業コストを軽減する支援が必要である。

そこで、本研究では、母国語だけを用いて書き換え作業を行う方法として、メッセージを対話相手の言語(目的言語)に翻訳後それをまた元の言語に翻訳する「折り返し翻訳」を検討する。ユーザは折り返し翻訳結果を読むことによって、対話相手におよどのような情報が伝わっているかを知ることができる。次に、書き換え作業量を減らす支援方法を考案するにあたり、本論文では特に以下の2点に注目した実験を行う。

- 書き換え作業量に影響を及ぼすユーザ属性にはどのようなものがあるか?
- ユーザに「良い翻訳結果を得るためのルール集合(以後、ルール集合と呼ぶ)」を教示(2.2節参照)することによって書き換え作業量を減らすことができるか?

1点目のユーザ属性に関して、本研究では特に以下の2つを考えた。

- (1) ユーザ属性 A: 専門分野
- (2) ユーザ属性 B: 母国語に関する知識量

実験の分析結果から、相互理解向上に役立つ機械翻訳システムを構築する際に有用な知見を得る。

2. 折り返し翻訳とルール集合

2.1 折り返し翻訳

母国語の知識しか持たないユーザ同士が機械翻訳を介して意思疎通を図ることは容易ではない。特に機械翻訳の翻訳精度が低いとき、話し手が対話相手に情報を伝えたつもりになっていても情報が伝わっていないことや、まったく意図しない情報が相手に伝わるということが考えられる。このように話し手と聞き手の間に情報の不整合があると、ユーザ間の相互理解が阻害され誤解が発生しやすい状態になる。このような状況を未然に防ぐためには、ユーザ自身が相手にどの程度、情報を伝達できているかを知る必要がある。もし対話者にどの程度の情報が伝達できているかの目安となる情報をフィードバックすることができれば、対話者は書き換え作業を行うことによって翻訳精度を高めることが可能になる。

本研究では、そのような手段の1つとして、折り返し翻訳を考える。折り返し翻訳とは、メッセージを目的言語に翻訳し、その翻訳結果をまた元の言語に翻訳することを指す。折り返し翻訳の翻訳結果を見ることによって、対話相手におおよそどのような情報がどの程度の翻訳精度で伝わったかを知ることができる⁵⁾⁻⁷⁾。もちろん、翻訳の非対称性から、目的言語の翻訳精度が良くても、折り返し翻訳の翻訳精度が悪い場合はあるだろう。しかし、目的言語の翻訳精度が良ければ、概して折り返し翻訳精度も良いことが予想できる。少なくとも、折り返し翻訳の翻訳精度が良ければ、目的言語の翻訳精度も良いことは確認できる。

2.2 書き換え作業時に学習するルール集合

ユーザは書き換え作業を繰り返す過程で良い翻訳結果を得るためのルールを学習する。もし良い翻訳結果を得るためのルールをあらかじめユーザに教示することができれば、ユーザの学習過程を短縮し書き換え作業量を減らすことができるはずである。たとえば、ユーザにあらかじめ「口語表現の使用を控える」というルールを教示していれば、ユーザは「彼女は北海道に行くんだって」といった口語表現の文章を入力せず、「彼女は北海道に行くそうです」といった表現を

我々は、工学系の被験者は計算機システムの仕組みに関する知識を持っており、機械翻訳システムの経験が少なくても初期適応が柔軟である可能性が高いと考えた。

表 1 書き換え作業の対象文と獲得すべきルール
Table 1 10 questions and corresponding rules.

問題	模範解答例	ルール
1 過度の森林伐採が、私は環境破壊の原因のように思います。	私は、過度な森林伐採が環境破壊の原因であると思います。	1) 修飾関係は非交差条件を満たす 2) 婉曲表現の回避
2 小人は一万円、大人は二万円が必要です。	小人は一万円が必要で、大人は二万円が必要です。	3) 述語の明記
3 そのテレビが故障しちゃったので、現在新しいのをつかっているんです。	そのテレビが故障したので、現在は新しいテレビを使っています。	4) 曖昧表現の回避 5) 主語の明記 6) 漢字表記 7) 口語表現の回避
4 そのような些細な問題気にする必要はないやん。	そのような些細な問題を気にする必要はありません。	8) 方言の回避 9) 格助詞の明記
5 多くの人々はその動物を嫌います。しかし、私は好きです。	多くの人々はその動物を嫌います。しかし、私はその動物が好きです。	10) 目的語の明記
6 事務機器には、コピー、ワープロ、キャビネットなどの事務機器が含まれます。	コピー、ワープロ、キャビネットなどは、事務機器に含まれます。	11) 不必要な同一語句の削除
7 彼は入院する必要はなし。代わりに、薬を飲み続けるように。しばらく会社を休むべきです。	彼は入院する必要がありません。代わりに、彼は薬を飲み続けるべきです。彼は、しばらく会社を休むべきです。	12) 体言止めの回避 13) 不完全文の回避 14) 主語の明記
8 重要なポイントは言葉の意味ではなく、その言葉が含まれる文脈です。	重要なポイントは言葉の意味ではありません。重要なポイントは、言葉が含まれる文脈です。	15) 長文の分割 16) 受動態の回避
9 彼は土壇場でその仕事を開始しました。結局締め切りに間に合わず、そのプロジェクトは停滞してしまいました。	彼は最後の瞬間まで仕事を開始しませんでした。締め切り前に終わらなかったで、そのプロジェクトは停滞してしまいました。	17) 慣用表現の回避 18) 節と節との論理関係の明確化
10 幸福なことに、私は多くのすばらしい友人と巡りあうことができた。	私は多くのすばらしい友人と出会うことができ、幸福だ。	19) 副詞句表現の回避 20) 動詞の書き換え

用いるだろう。

本研究では、このようにユーザに教示するルール集合、すなわち良い翻訳結果を得るためのルール集合として、機械翻訳システム作成者の提供するガイドライン「原文作成時に有効なルール」⁸⁾を考へる。ここでは、機械翻訳システム作成者が提供するガイドラインから重複部分とシステム特有のルールを取り除き、残されたルールを日本語文法体系^{9),10)}に従って分類して総計 70 のルールに整理したものを「良い翻訳結果を得るためのルール集合(付録 A.1, A.2, A.3 参照)」と定義する。

本実験では、ルール集合の教示が実際にユーザの書き換え作業量の削減に役立つかどうかを調べる。ここで、ルール集合はすべての言語対に共通するものではない。たとえば、「主語を明記する」ルールは日本語から英語への翻訳では効果的であるが日本語から韓国語への翻訳ではさほど効果がない⁴⁾。我々の狙いは、良い翻訳結果を得るためのルールの特定ではなく、ルール教示を行うことが書き換え作業量の削減を行う支援方法として妥当か否かを調べることである。

3. 実験概要

我々は、2003 年 9 月に折り返し翻訳を用いた書き換え作業の実験を実施した。実験参加者は 65 名の日本人学生(男性が 34 名、女性が 31 名)であり、参加者の大半(60 名)は機械翻訳の利用経験を持たない。

実験の参加者は、後述の折り返し翻訳ツールを用いて表 1 で与えられた 10 文の書き換え作業を実施する。表 1 で示すように、各文には獲得すべきルールが 1 つから 4 つ含まれており、ユーザは 10 文の書き換え作業を行う過程で合計 20 個のルールを獲得することが求められる。ここで 20 個のルールは「良い翻訳結果を得るためのルール集合」のうちで ICE2002 によく現れたものを用いた。

良い翻訳結果を得るためのルールの教示効果を調べるために、半数の参加者にはあらかじめルールの教示を行う(3.2 節詳説)。参加者は、各文について妥当な折り返し翻訳結果を得るまで上限を 5 回として書き換え作業を行う。作業終了の判断は各参加者に任せられ、各参加者が入力した原文と折り返し翻訳結果の意味が同じになったと判断した時点で作業を終了することとする。最後に各参加者に日本語力測定テストを実施し、日本語能力を測定する。日本語力測定テストは日本語学研究所による日本語力測定試験¹¹⁾からランダムに問題を 30 問抜粋したものをを用い、30 点を満点として採点した。

3.1 折り返し翻訳ツール

実験参加者は、図 1 の折り返し翻訳ツールを用いて書き換え作業を行った。本ツールは、日本語を英語に翻訳後それをまた日本語に折り返して翻訳するツール

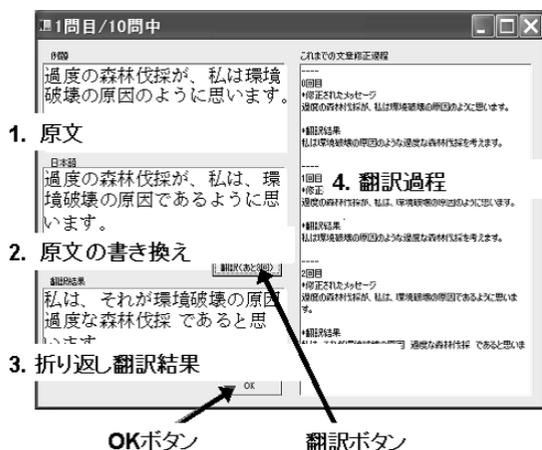


図 1 日英日折り返し翻訳ツール

Fig. 1 Japanese-English-Japanese turn-back translation tool.

である。参加者には英語の翻訳結果は見えず、日本語の折り返し翻訳結果しか見えないように設計されている。また、書き換え作業の過程が記録され、適宜参照できるようになっている。

3.2 良い翻訳結果を得るためのルールの教示

ルールの教示効果を調べるために、我々は書き換え作業前に参加者をランダムに2グループに分類し、片方のグループ(32名)にのみルール集合の教示を行った。ルール集合の教示とは、2.2節で説明した70個のルール集合に例文を付けた資料を配布し目を通す時間を5分与えるという簡単なものである。なお、書き換え作業中はルール教示の配布資料が閲覧できないようにする。

4. 分析手法

我々はまず、実験のログデータから各参加者が書き換えにかかった作業量を計測した。計測に際し、まずは、各参加者が各ルールをいつの時点で獲得できたかを調べ、点数付けを行った。ルール1を1回目の書き換え時に獲得できた場合は5点、5回目の書き換え時に獲得できた場合は1点、最後まで獲得できなかった場合は0点とする。次に、各参加者が全ルールの獲得にかかった合計得点(ルール獲得総合得点と呼ぶ)を算出し、これを参加者が書き換えにかかった作業量とする。ここで得点が高い参加者ほど容易にルールを獲得でき書き換え作業量が少ない。

次に、書き換え作業量(ルール獲得総合得点)に影響

表 2 共分散分析の結果
Table 2 Result of ANCOVA.

ソース	F 値	有意確率
専門分野(文系・理系)	3.263	.078
母国語に関する知識量	11.436	.002
ルール教示(有・無)	17.972	.000

を及ぼすユーザ属性(ユーザ属性 A: 専門分野, ユーザ属性 B: 母国語に関する知識量)を調べる。さらに、共分散分析を用いてユーザ属性の影響を除去したうえでルール集合の教示がルール獲得の作業量に影響を及ぼすか否かを調べる。具体的には、ユーザ属性による因子を共変量、ルール教示の有無を固定変数、ルール獲得総合得点を従属変数として共分散分析を行う。

5. 分析結果

書き換え作業量(ルール獲得総合得点)に影響を及ぼす因子を調べた結果を記す。

5.1 ユーザ属性と書き換え作業量

表 2 は共分散分析の分析結果を表している。表 2 より以下の知見を得た。

(1) 専門分野

本実験では、参加者の専門分野による違い(工学系 31名と文系 34名)はルール獲得総合得点に有意に影響を及ぼさないことが分かった($p=.078$)。

(2) 母国語に関する知識量

本実験で実施した日本語測定テストのスコア(以後、日本語スコアと呼ぶ)がルール獲得総合得点に有意に影響を及ぼしていることが分かった($p=.002$)。また、日本語スコアが高い人ほどルール獲得総合得点が高い(標準化係数 $\beta = .437$)。したがって、母国語(日本語)に関する知識の豊富なユーザほど、機械翻訳システムに解釈可能な文章を容易に作成できることが分かった。

ユーザ属性において、上述の2因子のほかに性別によるルール獲得差についても調べてみたが、男女の性差によるルール獲得総合得点の有意差は確認できなかった($p=.466$)。

5.2 ルール教示と書き換え作業量

表 2 より、ユーザ属性がルール獲得総合得点に及ぼす影響を除去しても、ルール教示はルール獲得総合得点に有意に影響を及ぼしていることが分かる($p=.000$)。なお、ルールの教示を受けた参加者の方が受けなかった参加者と比較してルール獲得総合得点が高い(標準化係数 $\beta = .593$)。

表 3 は、ルールの教示を受けた参加者と受けなかつ

日英翻訳には NTT サイバースペース研究所の翻訳サービス ALT-J/E, 英日翻訳には(株)高電社の翻訳サービス J-Server を用いた。

本分析では有意確率を $p=.05$ とした。

表 3 ルール教示有りの場合とない場合の書き換え作業過程の典型例
Table 3 A typical example of translation refinement process.

書き換え回数	ルール教示あり		ルール教示なし	
	原文	折り返し翻訳結果	原文	折り返し翻訳結果
問 5	多くの人々はその動物を嫌います。しかし、私は好きです。	多くの人々はその動物を嫌います。しかし、私は望みます。	多くの人々はその動物を嫌います。しかし、私は好きです。	多くの人々はその動物を嫌います。しかし、私は望みます。
1 回目	多くの人々はその動物を嫌います。しかし、私はその動物が好きです。	多くの人々はその動物を嫌います。しかし、私はその動物が好きです。	多くの人々はその動物を嫌います。しかし、私は嫌いではありません。	多くの人々はその動物を嫌います。しかし、私はそれを嫌い、そこで、そうではありません。
2 回目			多くの人々はその動物を嫌います。しかし、私は好きなんです。	多くの人々はその動物を嫌います。しかし、私が好きな be。
3 回目			多くの人々はその動物を嫌います。しかし、私はその動物のことが好きです。	多くの人々はその動物を嫌います。しかし、私はその動物が好きです。

た参加者の書き換え作業過程の一例である。この例は、問題（原文）「多くの人々はその動物を嫌います。しかし、私は好きです。」に対して参加者が書き換え作業を行ったものである。ルールの教示を受けた参加者は初めの書き換え時にルール 10（「目的語の明記」）を獲得し精度の高い折り返し翻訳結果を得ることに成功している。これに対して、ルールの教示を受けなかった参加者は原文のどこを修正すればよいか気付くまでに時間がかかり、書き換え作業 3 回目にルールの獲得を果たしている。

ルールをあらかじめ知っておくことの重要性は、実験参加者のアンケートからも確認することができた。書き換え作業に関する感想を自由記述形式で求めたところ、ルールを教示しなかった参加者から「機械翻訳はブラックボックス的で何がどうなっているのかまったく分からない。融通の利かない機械翻訳を使いこなすにはいろんなルールをあらかじめ知っておく必要があると感じた」といった感想が寄せられた。

5.3 母国語の知識量とルール教示効果

図 2 は「母国語に関する知識量」および「ルール教示の有無」がどのように「ルール獲得総合得点」に影響を及ぼしているかを図示したものである。図 2 の横軸は日本語スコア、縦軸はルール獲得総合得点を表している。ただし、参加者の個体差によるばらつきを消すために日本語スコアを 5 点間隔に区切り間隔ごとに参加者のルール獲得点数を平均化した。図 2 より、日本語スコアが低い参加者でもルールの教示を行うことによって、機械翻訳システムに容易に適應できることが分かる。

さらに母国語に関する知識量とルールの教示効果の

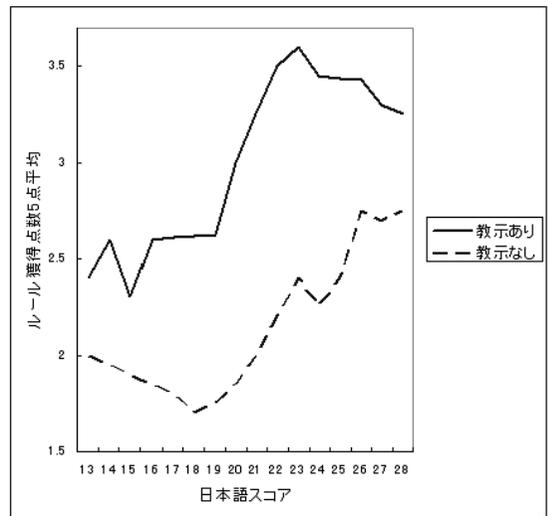


図 2 日本語スコアとルール獲得点数 (5 点平均)
Fig. 2 Relation between Japanese score and rule acquisition score.

関係を調べるために、ユーザを母国語に関する知識量によって上位、中位、下位の 3 グループに分類し、各グループとルール教示効果の関係を調べた (図 3)。ただし、3 グループに分類する際、日本語スコアの閾値は各グループの人数が均等に配分されるように調整した。多重比較法による分析の結果、母国語に関する知識量が中位のユーザに対するルール教示効果が最も高いことが分かった ($p < .01$)。この結果は以下のように解釈できる。すなわち、母国語に関する知識量が低いユーザはルールの教示を行ってもそれを実行できない場合があると考えられる。また、知識量が高いユーザは、もともとルールの獲得能力が高いため、ルール教示効果が比較的薄いと考えられる。

5.4 操作指示型ルールと操作自由型ルール

前節の分析結果より、全体的にルールの教示は書き

本例において、ルールの教示を受けた参加者と受けなかった参加者はともに日本語のスコアが同レベル (中位) に位置していた。

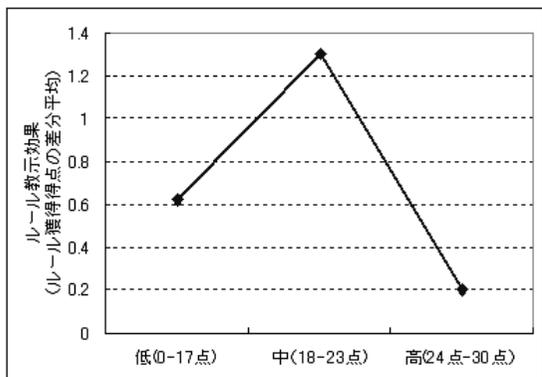


図 3 母国語レベルとルール教示効果

Fig. 3 Relation between Japanese level and effect of rule instruction.

表 4 操作指示型ルールと操作自由型ルール

Table 4 Operational rules and conditional rules.

操作指示型ルール	操作自由型ルール
3) 述語の明記	1) 修飾関係は非交差条件を満たす
5) 主語の明記	2) 婉曲表現の回避
6) 漢字表記	4) 曖昧表現の回避
9) 格助詞の明記	7) 口語表現の回避
10) 目的語の明記	8) 方言の回避
11) 不必要な同一語句の削除	12) 体言止めの回避
14) 主語の明記	13) 不完全文の回避
15) 長文の分割	16) 受動態の回避
18) 節と節との論理関係の明確化	17) 慣用表現の回避
	19) 副詞句表現の回避
	20) 動詞の書き換え

換え作業量を減らすうえで役立つことが分かった。しかし、教示がすべてのルールの獲得に役立つわけではない。

ルール集合は、大きく次の2種類に分けることができる。すなわち、「主語を明記する」といったように、何をどうすべきかを明示した「操作指示型ルール」と、「口語表現の回避」のように、どうしてはいけないことは書かれているがどうすべきかを明示しない「操作自由型ルール」である。表4は本実験で用いた20ルールの分類結果を表している。

我々は、実験参加者の書き換え過程を対象に、各ルールを何度目の書き換え作業時に獲得できたかを調べた。そして、ルールの教示を受けた者と受けなかった者との間で各ルールの獲得までにかかった作業量を比較した。その結果、操作指示型ルールに対する教示効果は高く、ルールの教示を受けたユーザはルールの教示を受けなかったユーザよりもこれらのルールを獲得するまでにかかる書き換え作業量が有意に少ないことが分かった ($p=.000$)。逆に、操作自由型ルールに対する

教示効果は低く、ルール教示を受けたユーザとルール教示を受けなかったユーザの間でこれらのルールの獲得までにかかる書き換え作業量に大きな違いはみられなかった ($p=.276$)。

操作指示型ルールは原文をどのように修正すればよいかイメージしやすいのに対し、操作自由型ルールは原文をどのように修正すればよいかイメージすることが難しいためと考えられる。

6. むすび

機械翻訳を介したコミュニケーションを通じて相互理解を構築するためには、翻訳精度の向上とともに相互作用性の向上が重要である。我々は、機械翻訳に対する知識をほとんど持たないユーザが、非同期型のコミュニケーションを行う際、原文の書き換えを繰り返し行うことに注目した。本論文では、機械翻訳に関する知識をあまり持たないユーザが母国語だけをを用いて書き換え作業を実施する場面を想定し、ユーザが機械翻訳に初期適応する際にかかる書き換え作業量を減らすための方法を考案する実験を実施した。

我々は本研究を通じて、折り返し翻訳が母国語で書き換え作業を行うための手段として有効であることを確認した。これは、大半のユーザが、すべてのルールを獲得して納得のいく折り返し翻訳結果が出力されるまで書き換え作業を継続し、納得のいく折り返し翻訳結果が得られたときに自主的に書き換え作業を終了したためである。母国語で書き換え作業を実施できるということは、母国語の知識しか持たない者同士でも機械翻訳を介してコミュニケーションを実現できることを示唆している。

続いて、本実験より得た知見は以下のとおりである。

- (1) 母国語に関する知識が豊富なユーザほど、多くのルールを容易に獲得できた。
- (2) ユーザに事前に「良い翻訳結果を得るためのルール集合」の教示を行うと、母国語に関する知識が豊富でないユーザでも多くのルールを容易に獲得できるようになり、書き換え作業量が大幅に減った(本実験では平均27.6%減少した)。
- (3) 原文をどのように変更すべきかを明示しない「操作自由型ルール」に対する教示効果は薄く、これらのルール獲得にかかる書き換え作業量は大きく減少しなかった。原文をどのように変更すべきかを明示した「操作指示型ルール」に対する教示効果は高く、これらのルール獲得にかかる書き換え作業量は大きく減少した。
- (4) ルールの教示効果は母国語に関する知識量が中

程度のユーザに対して最も高かった。母国語に関する知識量が高いユーザは教示の必要性がもともと低く、知識量が低いユーザは教示が実際のルール獲得に結びつかない場合があるようだ。

(5) ユーザの専門分野や性別は書き換え作業の作業量に有意に影響を及ぼしていなかった。

以上より、ユーザに「良い翻訳結果を得るためのルール集合」を教示することによってユーザにかかる書き換え作業量を減らすことができることが分かった。このようなルール教示機能を機械翻訳に埋め込むことによって異言語ユーザ間の円滑なコミュニケーションを実現できる可能性がある。

今後の課題は以下のとおりである。まず、本研究では英語に関する知識を保持していたと考えられるユーザに対して日英日の折り返し翻訳の実験を実施したため、今後はユーザが目的言語を知らない場合（たとえば日中日や日韓日の折り返し翻訳など）についても実験を実施する必要がある。

また、ユーザの適応を促進する機能を考案するうえで以下の事柄を調べる必要がある。まず、どのタイミングでどのルールをどのように呈示すると効果的であるかを調査する必要がある。また、ルール獲得の難易度はコミュニケーションの内容や文章の特徴（文長や1文あたりに獲得すべきルール数）にも依存すると考えられる。これらの特徴が書き換え作業量に及ぼす影響も調査し、状況に応じた支援方法を考案する必要がある。なお、言語ごとに教示が有効なルールと有効でないルールを調べ、教示が有効でないルールについては、その原因の究明を行い別の支援方法を考案する必要がある。

以上の課題は、ユーザの書き換えが円滑なコミュニケーションにつながることを前提としているが、ユーザの書き換え自体がかかえる問題についても考えていく必要がある。書き換えを行うことによって、ユーザが当初伝えたいと思っていたニュアンスが抜け落ちてしまうことが予想できる。実際、本実験の参加者の中には、ユーザの書き換えによって原文のニュアンスが落ちてしまうことや使いたい表現が使えなくなってしまったことに対して不満を述べる者がいた。今後、どのようなコミュニケーションの種別や状況において、このようなポイントが問題となるかを明らかにしていく必要がある。

謝辞 本研究の統制実験は、総務省アジアブロードバンドプロジェクトの一環として行われた。また、実験で使用した折り返し翻訳ツールは JST デジタルシティプロジェクトより提供していただいた。本研究を進

めるにあたり、機械翻訳システムに関する有用な議論をいただいた AAMT (Asia-Pacific Association for Machine Translation) 技術動向調査委員会の皆様へ感謝する。また、NTT コミュニケーション科学基礎研究所の松原繁夫氏から本論文の執筆にあたり、有意義なコメントをいただいたことに感謝いたします。

参考文献

- 1) 石田 亨, 林田尚子, 野村早恵子: 異文化コラボレーションに向けて—機械翻訳システムの相互作用性, 電子情報通信学会技術研究報告, AI (2003).
- 2) Nomura, S., Ishida, T., Yamashita, N., Yasuoka, M. and Funakoshi, K.: Open Source Software Development with Your Mother Language: Intercultural Collaboration Experiment 2002, *International Conference on Human-Computer Interaction (HCI-03)*, Vol.4, pp.1163–1167 (2003).
- 3) 野村早恵子, 石田 亨, 船越 要, 安岡美佳, 山下直美: アジアにおける異文化コラボレーション実験 2002: 機械翻訳を介したソフトウェア開発, *情報処理*, Vol.44, No.5, pp.503–511 (2003).
- 4) 小倉健太郎, 林 良彦, 野村早恵子, 石田 亨: 機械翻訳を介したコミュニケーションにおける利用者の機械翻訳システム適応の言語依存性, *自然言語処理*, Vol.12, pp.183–201 (2005).
- 5) 内元清貴, 林田尚子, 石田 亨, 井佐原均: 機械翻訳可能性の自動評価, *言語処理学会第 11 回年次大会発表論文集*, pp.9–12 (2005).
- 6) Yokoyama, S., Kashioka, H., Kumano, A., Matsudaira, M., Shiokizawa, Y., Kodama, S., Ehara, T., Miyazawa, S. and Murata, Y.: An Automatic Evaluation Method for Machine Translation using Two way MT, *MT Summit VIII* (2001).
- 7) Frederking, R.E., Black, A.W., Brown, R.D., Moody, J. and Steinbrecher, E.: Field Testing the Tongues Speech-to-Speech Machine Translation System, *Proc. 3rd International Conference on Language Resources and Evaluation (LREC 2002)* (2002).
- 8) Arendse, B. and Claudia, G.: MTranslatibility, *Machine Translation*, Vol.16, No.3, pp.175–218 (2001).
- 9) 橋本信吉: 国文法体系論, pp.14–21 (1969).
- 10) 徳田政信: 近代文法図説, pp.18–43 (1983).
- 11) 宮腰 賢 (監修): 日本語力測定試験問題集, 日本語学研究所 (2003).

日本語体系による分類		ルール集合
品詞に基づくルール	名詞に基づくルール	<ul style="list-style-type: none"> ・名詞の書き換え <ul style="list-style-type: none"> - 複敬の意味を持つため上手く翻訳されない名詞を適切に書き換える
	動詞に基づくルール	<ul style="list-style-type: none"> ・動詞の書き換え <ul style="list-style-type: none"> - 正しい訳語が出ない動詞を適切に書き換える - 動詞は名詞と解釈されないような形に書き換える <ul style="list-style-type: none"> - 補助動詞を伴うと上手く翻訳されない動詞を適切に書き換える
	形容詞に基づくルール	<ul style="list-style-type: none"> ・形容動詞の書き換え
	助動詞に基づくルール	<ul style="list-style-type: none"> ・助動詞「で」の使用の回避
	副詞に基づくルール	<ul style="list-style-type: none"> ・副詞の削除 <ul style="list-style-type: none"> - 形容詞に書き換えられない副詞を削除する
	接続詞に基づくルール	<ul style="list-style-type: none"> ・文頭の接続詞は文末の述語句に係るようにする
	間投詞に基づくルール	<ul style="list-style-type: none"> ・感動詞の使用は避ける ・呼び掛けの使用は避ける ・助詞を明記する <ul style="list-style-type: none"> - 格助詞を明記する
語の構成に基づくルール	助詞に基づくルール	<ul style="list-style-type: none"> ・主格でない副助詞（係助詞）「は」の使用は避ける、 ・助詞「で」の使用は避ける ・話題提示の助詞「は」は、連用修飾句と同じ扱いにする（ただし3つ以上の述語句のある時は最後の述語句に係るものとする）
	熟語に基づくルール 接頭語・接尾語に基づくルール	<ul style="list-style-type: none"> ・熟語の使用を避ける ・接続表現の書き換え
その他のルール		<ul style="list-style-type: none"> ・システムにとっての未知語の使用を避ける ・正しい訳語が出ない外来語はアルファベット表記する ・無理な造語の使用を避ける ・一般的な語句を用いる ・日本文化特有の事物名の使用を避ける ・漢字を使用する ・同音異義語を持つ語句は漢字で書く <ul style="list-style-type: none"> - 漢字表記可能な部分は漢字で書く ・多義語は一意に意味が特定できる表現に書き換える ・冗長な表現を避ける <ul style="list-style-type: none"> - 不必要な語句は削除する - 削除してもその後の文の意味が変わらない語句は削除する - 不必要な同一語句を削除する - 不必要な同義語は削除する ・鑑りが正しいことを確認する

図 4 良い翻訳結果を得るためのルール集合（単語に基づくルール）
Fig.4 Effective rules when refining original text.

付 録

A.3 良い翻訳結果を得るためのルール集合（表現に基づくルール）(図 6)

- A.1 良い翻訳結果を得るためのルール集合（単語に基づくルール）(図 4) (平成 17 年 7 月 28 日受付)
(平成 18 年 1 月 6 日採録)
- A.2 良い翻訳結果を得るためのルール集合（文章に基づくルール）(図 5)

日本語体系による分類		ルール集合
文の成否に基づくルール	主語に基づくルール	<ul style="list-style-type: none"> ・主語を明記する ・述語を明記する <ul style="list-style-type: none"> - 不完全な文の文末に述語を補完する - 主述関係を明確にするため、文の途中に述語を補完する
	述語に基づくルール	<ul style="list-style-type: none"> ・述語の書き換え <ul style="list-style-type: none"> - 動詞の書き換え - 形容詞・形容動詞の書き換え
	目的語に基づくルール	<ul style="list-style-type: none"> ・目的語を明記する ・目的語を長くしない ・長い修飾語を付けない <ul style="list-style-type: none"> - 長い修飾語は適切な箇所分割する - 連体修飾形を連用修飾形に変える
	修飾語・被修飾語に基づくルール	<ul style="list-style-type: none"> ・修飾語は直近の述語句、体言に係るようにする ・修飾関係は非交差の条件を満たすものとする ・修飾語の書き換え <ul style="list-style-type: none"> - 形容詞を副詞に変える
	独立語に基づくルール	<ul style="list-style-type: none"> ・感動詞の使用は避ける ・呼び掛けの使用は避ける
	その他のルール	<ul style="list-style-type: none"> ・節と節との論理関係（原因、結果、目的、手段、条件など）を明確にする <ul style="list-style-type: none"> - 前節を、連体修飾語を含む副詞節に書き換える - 前節に形式名詞を挿入する ・句読点を適切な箇所に挿入する <ul style="list-style-type: none"> - 漢字が連なり一つの語として解釈されてしまうことを避けるため、語句の切れ目に読点を挿入する - 係り受け関係が明確になるような箇所に読点を挿入する ・長い文を避け、短い文にする <ul style="list-style-type: none"> - 接続助詞を使用した長い文は、接続助詞の前後で分割する - 連用中止を使用した文は、連用中止の前後で分割する ・文スタイルを変更する <ul style="list-style-type: none"> - 箇条書きを避ける - 名詞以外が並列した文章は避ける ・文法的でない構造は避ける ・単純な文法構造にする ・係りと受けを対応させる ・受動態は避ける ・連用中止形を避け、適切な助詞を挿入する ・体言止めを使用しない

図 5 良い翻訳結果を得るためのルール集合（文章に基づくルール）

Fig. 5 Effective rules when refining original text.



山下 直美

1999年京都大学工学部情報工学科卒業。2001年同大学大学院情報学研究科数理工学専攻修士課程修了。同年日本電信電話（株）コミュニケーション科学基礎研究所入所、現在に至る。

2005年より京都大学大学院情報学研究科社会情報学博士課程在学中。



坂本 知子

2004年京都大学大学院情報学研究科修了。現在（株）野村総合研究所勤務。

日本語体系による分類	ルール集合
表現に基づくルール	<ul style="list-style-type: none"> ・曖昧な表現を避ける <ul style="list-style-type: none"> ・具体的な表現に書き換える <ul style="list-style-type: none"> - 形式名詞「もの」「こと」「方」を具体的な表現に書き換える ・準体助詞を具体的な表現に書き換える <ul style="list-style-type: none"> ・語順を変えて係り受けを明確にする ・簡潔な文にする <ul style="list-style-type: none"> ・形式名詞「もの」「こと」「よう」「ところ」を削除する ・副詞句表現の書き換え <ul style="list-style-type: none"> ・複合的で、係り受け解析や適切な訳語の用意が困難な副詞句は、簡潔な表現に書き換える ・文末表現の書き換え <ul style="list-style-type: none"> ・文末複合表現の書き換え ・使役+依頼の複合様相表現の書き換え ・口語表現の使用を避ける ・感情表現の使用を避ける ・慣用表現の使用を避ける ・スラング、方言の使用を避ける ・比喩の使用を避ける ・独特の言い回しを避ける ・複雑な否定表現を避ける ・敬語表現は避ける

図 6 良い翻訳結果を得るためのルール集合（表現に基づくルール）
 Fig.6 Effective rules when refining original text.



野村早恵子

2003年京都大学大学院情報学研究科社会情報学専攻博士課程修了。情報学博士。現在、カリフォルニア大学サンディエゴ校、分散認知とHCI研究室研究員。多言語環境におけるHCI, CMC研究(認知エスノグラフィー)に興味を持つ。



林 良彦(正会員)

1981年早稲田大学理工学部電気工学科卒業。1983年同大学大学院理工学研究科博士前期課程修了。同年日本電信電話公社入社。2004年日本電信電話株式会社退職(退職時, NTTサイバースペース研究所主幹研究員・グループリーダー)。この間, 1994~1995年スタンフォード大学言語情報研究センター滞在研究員。2004年より大阪大学大学院言語文化研究科教授。現在に至る。博士(工学)。自然言語処理, 知的情報アクセスの研究に従事。電子情報通信学会, 人工知能学会, 言語処理学会, ACL, ACM SIG-IR 各会員。



石田 亨(フェロー)

1976年京都大学工学部情報工学科卒業, 1978年同大学大学院修士課程修了。同年日本電信電話公社電気通信研究所入所。ミュンヘン工科大学, パリ第六大学, メリーランド大学客員教授等経験。工学博士。IEEE フェロー。情報処理学会フェロー。現在, 京都大学大学院情報学研究科社会情報学専攻教授, 上海交通大学客員教授。自律エージェントとマルチエージェントシステム, セマンティック Web 技術に取り組む。デジタルシティ, 異文化コラボレーションプロジェクトを推進。



小倉健太郎（正会員）

1978年慶應義塾大学工学部管理工学科卒業．1980年同大学大学院管理工学専攻修士課程修了．同年日本電信電話公社（現NTT）入社．1987～1990年ATR自動翻訳電話研究所へ

出向．現在，NTTサイバースペース研究所主任研究員．機械翻訳の研究に従事．1995年人工知能学会論文賞受賞，2002年電気通信普及財団賞（テレコム・システム技術賞）受賞．電子情報通信学会，人工知能学会，言語処理学会，計量国語学会各会員．



井佐原 均（正会員）

1980年京都大学大学院修士課程修了．博士（工学）．同年通商産業省電子技術総合研究所入所．1995年郵政省通信総合研究所．現在，独立行政法人情報通信研究機構けいはんな

情報通信融合研究センター自然言語グループリーダーおよびタイ自然言語ラボラトリ長．自然言語処理，語彙意味論の研究に従事．言語処理学会，人工知能学会，日本認知科学会，ACL各会員．