

折返し翻訳を用いた翻訳リペアの効果

宮部 真衣^{†a)} 吉野 孝^{††,††b)} 重信 智宏^{†††c)}

Effects of Undertaking Translation Repair Using Back Translation

Mai MIYABE^{†a)}, Takashi YOSHINO^{††,††b)}, and Tomohiro SHIGENOBU^{†††c)}

あらまし 翻訳リペアは、不適切な翻訳箇所が少ない文章を作成する方法として、機械翻訳を用いたコミュニケーションにおいて重要な役割と果たすと考えられる。しかし、適切に翻訳されていない文章を見てユーザが原文の書換えを行った場合、その翻訳リペアの結果がどの程度の精度をもつかについての評価は行われていない。そこで本論文では、折返し翻訳を用いた、日本語の入力による英語・中国語・韓国語間の翻訳リペアの実験を行った。本実験では、15文字以上32文字以下の文の翻訳リペアを行い、その翻訳リペア結果の精度評価及び作業のコストについての評価及び考察により、以下の知見を得た(1)本実験では約6回の翻訳リペアを行うことにより、実験で用いた65%の文の3言語の平均翻訳精度を正しく意味伝達が可能なレベルまで改善できた(2)正しく意味伝達が可能な文へ修正するための修正コストは、修正困難な語句・表現の存在や、修正前の折返し翻訳精度に依存する。翻訳精度が低い場合、ユーザによる修正箇所の特定や修正方法の発見は容易ではなく、翻訳精度に応じたコストを要する。

キーワード 折返し翻訳, 翻訳リペア, 機械翻訳

1. ま え が き

現在、世界規模のインターネットの普及に伴い、多言語によるコミュニケーションの機会が増加している。多言語間でコミュニケーションを行うには、様々な言語による対応が求められる。しかし、一般に多言語を十分に習得することは非常に難しいため、母国語以外の言語によりコミュニケーションを行うことは困難であり[1]~[3]、相互理解ができない可能性が高い[4],[5]。そのため、母国語を用いてコミュニケーションを行うために、機械翻訳などの技術が利用されている[6],[7]。

近年、翻訳技術は急速に進展しているが、翻訳精度には限界があり、完璧な翻訳を行うことは困難である。

機械翻訳を用いる場合、短文では比較的高精度な翻訳が期待できるが、長文では不適切な翻訳箇所が増加する。しかし、コミュニケーションにおいて、不適切な翻訳箇所を含む文章は話者間の相互理解を困難にし、円滑なコミュニケーションの妨げとなる。チャットのようなコミュニケーションは、短い文章でも成立するため、機械翻訳を用いた相互理解は可能であると考えられる。一方、ある程度の長さの文章をコミュニケーションで用いる場合、翻訳結果に不適切な翻訳箇所を含む可能性が高まり、相互理解が困難になる。したがって、円滑にコミュニケーションを行うためには、不適切な翻訳箇所の少ない文章を作成しなければならない。文章中の不適切な翻訳箇所を減少させるために、入力文章を書き換えていくことを「翻訳リペア」と呼ぶ。

翻訳リペアは、不適切な翻訳箇所の少ない文章を作成する方法として、機械翻訳を用いたコミュニケーションにおいて重要な役割を果たすと考えられる[8]。しかし、翻訳リペアを行うには相応のコストを要する[9]。

一般ユーザにとっては、母国語ではない他国語の翻訳結果を見て翻訳ミスを見つけ出すことは困難である。そのため、被験者が翻訳結果を理解できるように、折返し翻訳が利用されている[10]。折返し翻訳により各言語の翻訳結果をもとの言語にし、その翻訳結果を見

[†] 和歌山大学大学院システム工学研究科, 和歌山市
Graduate School of Systems Engineering, Wakayama University, 930 Sakaedani, Wakayama-shi, 640-8510 Japan

^{††} 和歌山大学システム工学部, 和歌山市
Faculty of Systems Engineering, Wakayama University, 930 Sakaedani, Wakayama-shi, 640-8510 Japan

^{†††} 独立行政法人情報通信研究機構言語グリッドプロジェクト, 京都府
Language Grid Project, National Institute of Information and Communications Technology, 3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto-fu, 619-0289 Japan

a) E-mail: s085051@sys.wakayama-u.ac.jp

b) E-mail: yoshino@sys.wakayama-u.ac.jp

c) E-mail: shigenobu@nict.go.jp

ることで、他言語を十分に習得していないユーザの翻訳リペア作業が容易になる。

これまでに、翻訳ミスを含む翻訳文を見て、正しい翻訳文になるように原文を書き換えた場合、その翻訳リペアの結果がどの程度の翻訳精度をもつかは評価されておらず、翻訳リペアのコストに見合うだけの翻訳精度が得られるかについては検証されていない。

そこで、本論文では翻訳リペアの実験及び翻訳リペア結果の評価を行い、以下の点について検証を行う。

(1) 翻訳リペアにより翻訳結果は意味の伝達が可能な文になっているか。

(2) 1回の修正により、大幅な翻訳精度の改善は可能なのか。

これらの検証を行うことにより、コミュニケーションにおいて意味を伝えるにあたり、問題のないレベルの翻訳リペアを行うことができるか、また翻訳リペアの修正コストが円滑なコミュニケーションを妨げる可能性があるかを明らかにする。

2. 関連研究

機械翻訳を用いたコミュニケーションでは、相互の発言内容理解は翻訳精度に依存する。低い翻訳精度の場合、相互理解ができず、思い違いが発生する [11]。翻訳リペアは、このような思い違いの回避においても重要であると考えられる。

翻訳リペアの関連研究として、これまでにリペア（翻訳に適した文章への修正）支援機能の性能予測 [10] が行われており、折返し翻訳機能の検証実験によって、折返し翻訳結果の提示はリペアに有効であることが示されている。

異文化協調作業におけるコミュニケーションにおいて行われた自己リペア [8] では、母国語が英語に翻訳され、その英語を見ながら母国語の修正を行っており、母国語の修正により、文章の翻訳精度が向上している。しかし、母国語以外を十分に習得していない場合、翻訳結果を見ながら修正を行うのは困難である。母国語だけをういてリペア作業を行う方法として、折返し翻訳の利用が検討され [12]、折返し翻訳がリペア手段として有効であると示されている。また、折返し翻訳のためのリペア支援機能の評価実験 [13] により、修正にかかるコスト（時間、回数）評価が行われている。

これらの研究では、折返し翻訳を用いた翻訳リペア作業が行われているが、修正により得られた文の評価は行っておらず、翻訳リペアによる翻訳結果の改善効

果の検証は行われていない。そのため、適切で円滑なコミュニケーションを支援するために、折返し翻訳を見ながら翻訳リペアを行った場合、どの程度翻訳結果を改善可能なのか検証する必要がある。小倉らの研究 [14] により、翻訳リペアは言語に依存し、特定言語におけるリペアが他の言語において有効であるとは限らないことが分かっている。そこで今回の実験では、多言語での利用を想定し、英語・中国語・韓国語の翻訳リペアを同時に行い、リペアが完了した言語から翻訳結果を確定していく。3言語すべてのリペアが完了した時点での修正時間と修正回数を修正コストとして、評価を行った。今回の実験では、翻訳精度の高い韓国語翻訳、中程度の英語翻訳、精度の低い中国語翻訳を同時に行い [14]、修正コストの平均化を図った。

3. 翻訳リペア実験

3.1 実験手順

折返し翻訳の結果を見ながら翻訳リペアを行う場合、翻訳リペアにより得られる翻訳結果の改善効果、3言語の翻訳リペアに要するコスト及びユーザによる原文の書換え方法を検証するために、折返し翻訳を用いた翻訳リペアの実験を実施した。

実験の被験者は、和歌山大学システム工学部及び大学院の学生5名である。被験者のうち、2名は折返し翻訳結果を見ながらの翻訳リペア作業の経験がある。他の3名は翻訳リペア作業の未経験者であり、実験の前に折返し翻訳及びツールの使い方について説明した。

翻訳リペア作業には、英語・中国語・韓国語の折返し翻訳が可能なツールを使用した。使用したツールの画面例を図1に示す。翻訳システムは高電社のJ-Server [15] を言語グリッド [16] を介して使用した。本ツールでは、入力エリアに文を入力すると、指定された対象言語の翻訳結果が折返し翻訳表示エリアに表示される。折返し翻訳の確定チェックボックスをチェックすることにより、翻訳結果を確定し、入力内容の変更に伴う影響を受けないようにすることができる。

予備実験により、15文字未満の文は高精度の翻訳結果が得られる可能性が特に高い傾向があることが分かった。一方で、機械翻訳システムでは、50文字以上の文の解析は非常に困難であり、80文字以上の文については解析がほとんど失敗すると指摘されている [17], [18]。現在、機械翻訳に必要な機能の改善が進んでいるが、十分な精度は得られていない [19]。また、本実験でリペア対象文として用いる機械翻訳機能試験

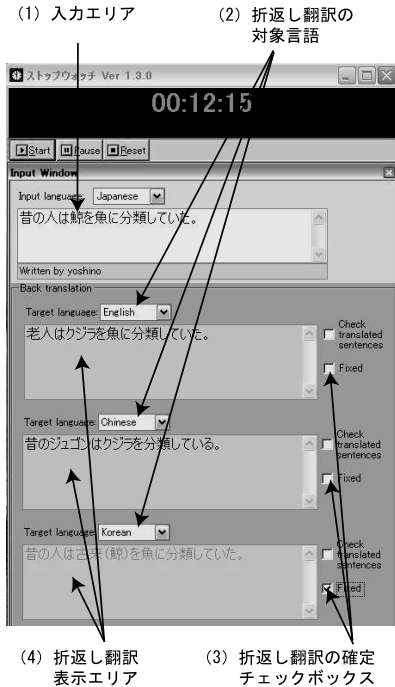


図 1 翻訳リペア実験ツールの画面例

Fig. 1 Sample screen of an experiment tool for translation repair.

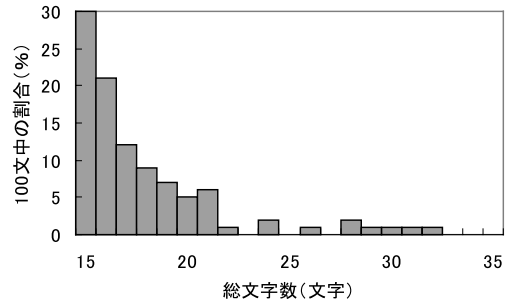


図 2 実験に利用した文の総文字数の分布

Fig. 2 Distribution of the total number of characters in the experiment sentences.

表 1 3言語の合計修正時間と合計修正回数

Table 1 Total repair times and the number of translation repairs in the three languages in the experiment.

被験者	3言語の合計修正時間		3言語の合計修正回数	
	平均 (分:秒)	標準偏差 (分:秒)	平均 (回)	標準偏差 (回)
A (経験あり)	02:43	02:18	8.3	7.7
B (経験あり)	01:42	01:17	4.1	3.3
C (経験なし)	01:30	01:07	4.6	4.2
D (経験なし)	02:57	01:56	7.9	6.0
E (経験なし)	02:07	01:39	6.8	5.9
平均	02:12		6.3	

文 [20] の文字数については、40 文字以下の文が全体の 98.5% を占めている．そこで本実験では、簡単な内容の文ではなく、ある程度の内容を表現可能である、15 文字以上 40 文字以下の日本語文をリペア適用範囲とし、翻訳リペア作業を行う．実験で利用した日本語文は、機械翻訳機能試験文 [20] のうち、15 文字以上 40 文字以下である文を最初から 100 文選択したものである．実験で利用した 100 文の文字数の分布を図 2 に示す．今回利用した文の最小文字数は 15 文字，最大文字数は 32 文字である．

実験の手順を以下に示す．

(1) 日本語文を入力エリアにコピー&ペーストする．

(2) 折返し翻訳の結果を見ながら、原文と同じ意味になるように入力日本語文を修正する．

(3) 折返し翻訳の結果が同じ意味になったと判断したら、確定機能を利用して各言語の修正結果を確定する．

(4) 3 言語すべて確定したら、1~3 の手順を用意した 100 文すべてについて行う．

コピー&ペースト直後の折返し翻訳結果が原文と同

じ意味になっていると被験者が判断した場合は、リペア作業を行わず、そのまま確定可能とした．また、5 分程度リペア作業を行っても同義の文へと修正ができない場合、被験者の判断により修正を放棄することも可能とした．実験には 3~6 時間を要し、実験中は適宜休憩をとるように指示した．

実験では、文字の入力過程や修正過程などのリペア作業の様子の記録を、画面キャプチャを利用して行った．

3.2 実験結果

(1) 翻訳リペアのコスト

3 言語の合計修正時間及び合計修正回数を表 1 に示す．また、実験中被験者がリペアを行わなかった文数を表 2 に、被験者が修正を放棄した各言語の文数を表 3 にそれぞれ示す．今回の実験は、被験者数が 5 名であり限定的な結果になっている．コストの検証を行うには不十分であり、多数の被験者により検証を行う必要がある．

表 1 における被験者の修正時間及び修正回数は、100 文のうち修正放棄文を除いた文の平均である．修正放棄文に関しては、被験者の判断に応じてコストが必要

以上に増加する可能性があるため、平均コスト算出の対象としていない。また、リペアを行わなかった文については、翻訳結果取得までのコスト及び文意一致判定のコストを考慮し、平均コスト算出の対象とした。

今回の実験では、一つの言語を修正した後、続けて残りの言語の修正を行っており、修正時間と修正回数は3言語の合計とした。また、被験者 A, B は翻訳リペア作業の経験者であり、被験者 C, D, E は未経験者である。

今回の実験では、一つの文について3言語を修正するのに要する平均時間は2分12秒、3言語を修正した合計回数は6.3回を要した。本実験における被験者は5名であり、今回計測したコストもまた限定的である。修正回数の少ない被験者は、翻訳リペアの経験が

表2 リペア不要の文数

Table 2 The number of unnecessary repair sentences in the experiment.

被験者	英語 (文)	中国語 (文)	韓国語 (文)
A (経験あり)	22	20	45
B (経験あり)	18	21	48
C (経験なし)	27	23	58
D (経験なし)	17	20	51
E (経験なし)	27	23	58
平均	22.2	21.4	52.0
標準偏差	4.8	1.5	5.9

表3 放棄の文数

Table 3 The number of sentences given up in the experiment.

被験者	英語 (回)	中国語 (回)	韓国語 (回)	平均 (回)
A (経験あり)	5	10	2	5.7
B (経験あり)	2	6	1	3.0
C (経験なし)	5	4	0	3.0
D (経験なし)	11	14	1	8.7
E (経験なし)	16	22	5	14.3
平均	7.8	11.2	1.8	6.9
標準偏差	5.6	7.2	1.9	4.8

表4 放棄した人数が3人以上の文

Table 4 Given-up sentences list by over three subjects in the experiment.

日本語文	言語	原因
浅間山が噴火口から煙を出している。	中	「浅間山」「煙」の翻訳が不適当
その壺は口が大きく開いている。	中	「壺」の翻訳が不適当
私はとなりどうしの球の距離をストロボの発光時間間隔で割る。	英, 中	全体的に語順がおかしくなる
この七億円は所得隠しにあたる。	中	「所得隠し」の翻訳が不適当
棚が高く子供には手が届かない。	中	「棚」の翻訳が不適当
太郎は腰を曲げてゴミを拾った。	英, 中	「腰を曲げて」の翻訳が不適当
象は鼻が長い、豚は鼻が短い。	英	「鼻」の翻訳が不適当
子供はおもちゃをもとの位置に戻した。	中	「元の位置に戻した」の翻訳が不適当

ある被験者 B と未経験者である被験者 C であり、両者の平均修正回数は5回以下である。また、修正時間が長い被験者は被験者 A と被験者 D であり、2分半以上かかっている。実験では、修正完了の判断を被験者自身で行っているため、翻訳結果の精度にかかわらず被験者が翻訳結果の意味が原文と同じであると早期に判断すれば、修正時間は短くなり、また修正回数は少なくなると考えられる。

(2) 修正の放棄

表3より、修正を放棄した文数は、韓国語が平均1.8文であるのに対し、英語と中国語はそれぞれ7.8文、11.2文と多くなっている。このことから、韓国語と比較して、英語と中国語では修正が困難である場合がより多く生じたと考えられる。

3人以上が修正を放棄した文は、100文中8文であった。該当文の内容、修正を放棄した言語及び放棄の原因を表4に示す。被験者全員が放棄した文は、「私はとなりどうしの球の距離をストロボの発光時間間隔で割る」(表4)という文の中国語翻訳である。この文は、中国語の折返し翻訳では「私とフラッシュのが時間間隔に発光するでなる達ボールの距離を等分する」となり、名詞の翻訳が不適当であり、また全体的に語順がおかしくなったため、修正が困難であったと考えられる。

表3に示される修正を放棄した文のほとんどが、翻訳リペアが困難な名詞や表現を含んでおり、それらの翻訳が不適当であることが放棄の一因であると考えられる。具体的には、「腰を曲げて」という表現は英語の折返し翻訳では「下の後ろを曲げて」となり(表4)、被験者3名は表現の言い換えができずに修正を放棄している。他の被験者2名は翻訳の確定をしており、1人は「腰を曲げて」という表現自体を削除した後確定をしている。もう1人の被験者は、「腰を曲げて」という表現を「前屈して」のように同義の表現へと書き換

表 5 実験中に被験者によって書き換えられた例
Table 5 Examples of sentences rewritten by subjects in the experiment.

書換え方法	文の種類	書換え内容
(A) 単語及び表現の言換え	原文	今年のクリスマスは日曜日にあたる。
	書換え文	今年のクリスマスは日曜日だ。
(B) 表現の補完	原文	あの人は言う事が変わっている。
	書換え文	あの人は発言内容が普通の人と比べておかしい。
(C) 文の簡略化	原文	我々は試合まであと 5 日を残している。
	書換え文	我々は試合まで 5 日を残している。
(D) 文の分割	原文	母親は寝ていた子供を起こした。
	書換え文	子供は寝ていた。母親は子供を目覚めさせた。
(E) 語順の変更	原文	この会場には観客が 5 万人入る。
	書換え文	観客がこの建物には 5 万人入る。
(F) 名詞の区別	原文	衛星放送を見るにあたっては、特別のアンテナとチューナーが要る。
	書換え文	“衛星放送”を見るために、特別なアンテナとチューナーが要る。

えることで対応した。前者の対応（該当表現の削除）による翻訳文は、「腰を曲げて」という表現がなくなったことで「タローはゴミを拾った」となった。この文では原文の意味が失われており、後述の 4. で述べる翻訳精度の評価値が低くなった。一方、後者の対応による翻訳文は、「タローは前にかがみ、ゴミを拾った」となり、原文と同等の意味とみなされ、翻訳精度の評価値が高かった。このように、うまく翻訳される同義の表現を見つけ出すことができれば、名詞・表現の不適切な翻訳の精度を改善することができる。

(3) 原文の書換え方法

実験中、被験者は以下の方法によって原文の書換えを行っていた。各方法の具体例を表 5 に示す (A)~(E) の方法については、山下らの研究 [11] における良い翻訳結果を得るためのルール集合と対応がとれている。

(A) 単語及び表現の言換え

被験者が行う主な修正作業は、表現の変更である。実験では、動詞・名詞の同義な表現・単語への変更（「会場」を「ホール」や「建物」にする等）や、助詞の変更（「～に」を「～へ」にする等）が行われていた。

(B) 表現の補完

被験者は、「あの人は言う事が変わっている」という文において、「普通の人よりも」のように、補助的な表現を追加することにより修正を行っていた。

(C) 文の簡略化

「我々は試合まであと 5 日を残している」という文では、「あと」という単語を削除し、より簡単な文へ修正していた。

(D) 文の分割

中国語翻訳では、文法的におかしな翻訳結果になる

ことも多い。一部の被験者は、文をより単純な二つの文へと分割することにより、翻訳結果の改善を行っていた。

具体的にいえば、「母親は寝ていた子供を起こした」という文は「母は寝る子供を始めた」と翻訳される。この文を「子供は寝ていた。母親は子供を目覚めさせた」という文に修正することにより「子供は眠っている。母は子供に目覚めさせた」という翻訳結果を導出していた。

(E) 語順の変更

もとの日本語文の語順でうまく翻訳できない場合、語順の変更により修正をしていた。「この会場には観客が 5 万人入る」という文では、「観客が」の位置を変更して「観客がこの建物には 5 万人入る」という文にする修正の様子が見られた。

(F) 名詞の区別

地名や名前などの固有名詞の翻訳がおかしいとき、被験者は名詞を“”や「」でくくっていた。実験では、地名（盛岡など）を括弧でくくることにより、翻訳結果を正しく修正できていた。しかし、一部の単語は括弧でくくっても修正できていなかった。「浅間山」は中国語では「浅い間に山」と翻訳されるが、括弧でくくっても同じ結果であり、修正が困難であった。このような場合、被験者は、括弧でくくったことで名詞と判断できるとし、修正を確定した人と、修正を放棄した人に分かれた。

4. 翻訳文の主観評価実験

4.1 実験手順

翻訳リペア実験により得られた翻訳リペア後の翻訳結果の精度を検証するために、翻訳リペア結果の主観

表 6 各被験者間の有意確率

Table 6 Significance probability in each subject.

比較被験者	英語	中国語	韓国語
A-B	0.870	0.824	0.029
A-C	0.821	0.000*	0.000*
A-D	0.000*	0.687	0.275
A-E	0.982	0.635	0.015
B-C	0.685	0.001*	0.001*
B-D	0.001*	0.787	0.295
B-E	0.989	0.750	0.113
C-D	0.000*	0.003*	0.000*
C-E	0.432	0.003*	0.024
D-E	0.007	0.536	0.337

*: 有意差あり $p < 0.005$

評価実験を行った。評価実験の被験者は、和歌山大学システム工学部及び大学院の学生 3 名であり、翻訳リペア実験の被験者とは別の学生が評価した。

実験内容は、以下の 2 組の文について、翻訳文がもとの日本語文と同じ意味になっているかどうか比較することである。

(1) 翻訳リペア実験で利用した日本語試験文と折返し翻訳直後の文(未修正)

(2) 翻訳リペア実験で利用した日本語試験文と被験者による修正後の折返し翻訳文(3言語分)

評価は Walker らの適合性評価(5段階評価)[21]により行った。評価値は、5: All(同じ意味), 4: Most(文法などに多少問題があるが、大体同じ意味), 3: Much(意味はなんとなくつかめる), 2: Little(雰囲気は残っているが、もとの意味は分からない), 1: None(全く違う意味)である。

評価値が 3 未満の場合、折返し翻訳文からは原文の意味が読み取れないことを意味する。評価値が 3 の場合は、なんとなく意味が分かるレベルの折返し翻訳であることを意味し、評価値が 4 以上の場合、かなり精度が高く、コミュニケーションなどで問題なく利用できるレベルの折返し翻訳であることを意味する。

評価は、並べられた 2 文(もとの日本語文及び折返し翻訳文)を見て直観で評価し、1 組の文に対して 30 秒以内で評価するものとした。

また、中間言語(母国語に折返し翻訳する前の翻訳言語)については、一般に折返し翻訳の精度が良ければ中間言語の翻訳精度も良いとされている[12]。そのため、今回の実験では中間言語の翻訳精度評価については議論しない。

4.2 評価実験結果

ウィルコクソンの符号付順位検定により各被験者間

表 7 3 言語の平均評価値

Table 7 Average evaluated value in three languages in the experiment.

言語	未修正	修正後平均	有意確率	放棄率(%)
英語	2.61	4.06	0.000	7.8
中国語	2.65	3.96	0.000	11.2
韓国語	3.78	4.60	0.000	1.8
平均	3.01	4.21		

の評価の多重比較を行った結果を表 6 に示す(ボンフェローニの不等式より、有意水準 0.5%において有意差があったものに*印を付加した)。英語では被験者 D、中国語及び韓国語では被験者 C において他の被験者との有意差が見られる。このような被験者の違いは、修正困難な文における放棄の有無や、文意一致判定基準によるものと考えられ、一部の修正困難な文を除いて、各被験者間の評価に大きな差はないと考えられる。そのため本論文では、これら 5 名の評価値により議論を進める。

未修正折返し翻訳文と修正後折返し翻訳文(被験者全員の平均評価)の各言語における平均評価値及びウィルコクソンの符号付順位検定による修正前後間の有意確率を表 7 に示す。表 7 における未修正、修正後平均の値は、3 人の評価者による 5 段階評価結果の平均値である。

実験の結果から、未修正折返し翻訳文の評価は、3 言語の平均で 3.01、修正後の折返し翻訳文は 4.21 となっており、文意の理解に対する翻訳リペアの効果が見られた。言語ごとに見ると、韓国語の評価は未修正折返し翻訳文でも比較的高く、韓国語の機械翻訳では、原文の意味がある程度はつかめる翻訳が可能であると考えられる。一方、英語と中国語の評価はいずれも 3 以下であり、入力した文の意味をつかむことが困難な、不適切な翻訳箇所を含む文になることを示している。5 人の被験者による修正後の折返し翻訳文の平均評価値は、英語 4.06、中国語 3.96、韓国語 4.60 となっており、またどの言語に関しても、修正前後で有意差がある。このことから、修正によって各言語の評価値は増加したといえる。

100 文それぞれの未修正折返し翻訳文及び修正後折返し翻訳文に対する評価値のグラフを、言語別に図 3、図 4、図 5 に示す。

どの言語についても、100 文のほぼすべてにおいて評価値の上昇が見られる。特に英語と中国語では、未修正では意味が分からなかった文(評価値が 3 未満)

が修正後には意味がつかめる文であると評価されている。言語によって差は見られるが、翻訳リペアを行うことにより、翻訳結果は確実に改善されているとみなすことができる。

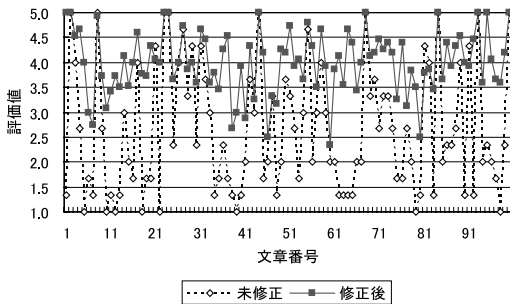


図3 修正前後における各文の評価値の比較（英語）

Fig. 3 Comparing the evaluations before and after translation repair in the experiment (using English-language machine translation).

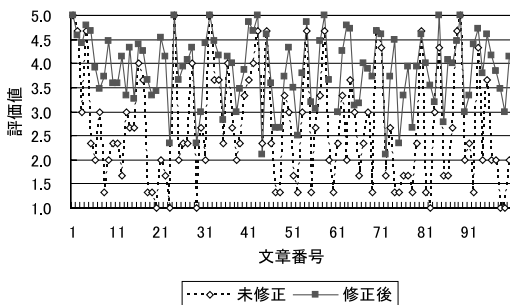


図4 修正前後における各文の評価値の比較（中国語）

Fig. 4 Comparing the evaluations before and after translation repair in the experiment (using Chinese-language machine translation).

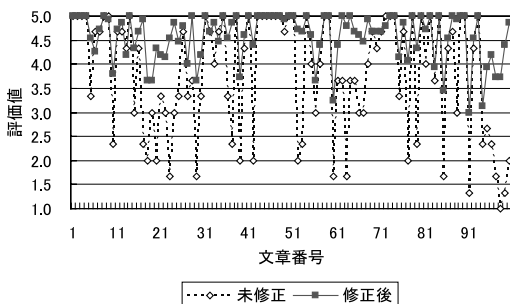


図5 修正前後における各文の評価値の比較（韓国語）

Fig. 5 Comparing the evaluations before and after translation repair in the experiment (using Korean-language machine translation).

5. 考 察

5.1 翻訳リペアによる改善効果

精度の高い折返し翻訳文数の変化を表 8 に示す。表 8 における 3 言語平均は、英語、中国語、韓国語の折返し翻訳文の評価の平均値である。

表 8 において、各言語の修正後折返し翻訳については、被験者 5 人の平均の評価値を用いた。中国語のみ 5 人全員が修正を放棄した文が 1 文あるため、合計が 99 文となっている。

3 言語の平均を見ると、評価値が 3 以上の文については、修正により 46 文増加し、100 文中 99 文が何となく意味がつかめるレベルまで修正できている。評価値が 4 以上（コミュニケーションに問題なく利用できるレベル）の文については、未修正の場合 100 文中 14 文であったのに対して、修正後は 65 文まで増加した。

また各言語では、未修正折返し翻訳結果を見ると、英語と中国語は、評価値が 3 以上の文が 100 文中 38 文である。一方、修正後は英語、中国語のどちらも評価値 3 以上の文が 9 割前後になっている。また、韓国語に関しては、未修正でも評価値 3 以上の文が 76 文と多いが、修正後にはすべての文の精度を評価値 3 以上に改善できている。

被験者別の平均評価値を表 9 に示す。

表 8 修正前後の高精度翻訳文数の変化

Table 8 The number of highly accurate translations before and after translation repair.

言語	翻訳結果の状態	評価値 3 以上の文数 (文)	評価値 4 以上の文数 (文)
英語	未修正	38	22
	修正後	94	54
中国語	未修正	38	20
	修正後	88	48
韓国語	未修正	76	55
	修正後	100	86
3 言語平均	未修正	53	14
	修正後	99	65

表 9 被験者別平均評価値

Table 9 Average evaluations of each subject in the experiment.

被験者	英語	中国語	韓国語	平均
A (経験あり)	3.99	4.07	4.73	4.27
B (経験あり)	3.97	3.99	4.63	4.19
C (経験なし)	3.94	3.61	4.39	3.98
D (経験なし)	4.33	4.05	4.66	4.35
E (経験なし)	4.07	4.09	4.60	4.26
平均	4.06	3.96	4.60	4.21

表 10 未修正文の評価値区間における 4 以上に修正できた文の平均修正回数及び評価値増加分の平均

Table 10 Average number of repair times and average increasing value of four-point or more improved sentences relative to the adjacent sections of un-repaired sentences.

未修正折返し翻訳文の評価値区間	該当文数(文)	平均修正回数(回)	評価値増加分の平均	修正 1 回当たりの増加量
1 以上 2 未満	8	7.93	2.32	0.29
2 以上 3 未満	32	7.30	1.45	0.20
3 以上 4 未満	39	6.33	0.90	0.14
4 以上 5 未満	14	3.17	0.31	0.10

どの言語についても、被験者ごとの評価値は同程度である。評価値に関しては、経験者と未経験者の違いはそれほど見られないが、被験者 C は他の被験者と比較すると若干評価値が低くなっている。被験者 C の修正時間及び修正回数(表 1)を見ると、修正時間・修正回数のどちらも他被験者と比較して少ないことが分かる。

修正完了の判断は被験者が行っているため、評価値は被験者の修正完了の判断基準に依存すると考えられる。特に、修正完了の判断、つまり原文と折返し翻訳結果の文意一致判定を厳しく行っていた場合、修正コストが増加し、評価値は上昇する。逆に、文意の一致判定が緩やかであれば、修正コストは減少するが、評価値は低下すると考えられる。すなわち、被験者 C の文意一致判定の基準は他被験者よりも緩やかであると考えられ、修正初期段階において更に改善可能な折返し翻訳結果のまま確定した文があり、その文の評価値が被験者 C の平均評価値を下げていると考えられる。

今回の実験では、どの被験者についてもリペアにより翻訳結果を改善できており、実験で用いた 100 文の内、65%の文については 3 言語の平均翻訳精度を正しく意味伝達が可能なレベルまで改善できた。

5.2 翻訳リペアによる修正コスト

今回の翻訳リペア実験では、修正コストとして修正時間と修正回数を計測した。計測の結果、被験者ごとに 1 回の修正に要する時間は異なっていた。修正時間は、被験者個人の能力や状況、文意一致判定の基準に依存するため、翻訳リペアによる評価値の変化の指標として修正時間を用いるのは適切でないと考えられる。そこで、折返し翻訳文の評価値変化の指標として修正回数について考察する。

未修正折返し翻訳の評価値を四つの区間に分けたときの、正しく修正された折返し翻訳文(3 言語の評価値の平均が 4 以上)の平均修正回数、評価値増加分の

表 11 隣接した評価値区間の有意確率

Table 11 Significance probability between adjacent-section evaluations.

未修正折返し翻訳文の評価値区間	平均修正回数の有意確率	評価値増加分の平均の有意確率	修正 1 回当たりの増加量の有意確率
1 以上 2 未満 2 以上 3 未満	0.565	0.000	0.156
2 以上 3 未満 3 以上 4 未満	0.326	0.000	0.065
3 以上 4 未満 4 以上 5 未満	0.002	0.000	0.992

平均及び修正 1 回当たりの評価値増加量を表 10 に示す。また、マン・ホイットニーの検定による、表 10 における隣接評価値区間の有意確率を表 11 に示す。

評価値が 4 以上であるものは原文と同義の意味をもつ文であると評価された折返し翻訳文である。表 10 では、修正後折返し翻訳文の 3 言語の評価値の平均が 4 以上である文について、未修正折返し翻訳文の評価値が「1 以上 2 未満」「2 以上 3 未満」「3 以上 4 未満」「4 以上 5 未満」の評価値区間に分類し、各文の評価値の増加量を求める。該当文数は各評価値区間における平均評価値が 4 以上である文の数である。また、評価値区間ごとに、求めた増加量と修正回数の平均値を算出し、修正 1 回当たりの評価値増加量を求める。また、求めた修正 1 回当たりの増加量に違いがあるかどうか比較を行う。

表 10 より、未修正のときに意味の伝達が不可能であると評価された文(評価値区間が 1 以上 2 未満)でも、意味の伝達が可能なレベルまで精度を上げることが可能であると分かる。

各評価値区間における平均修正回数を比較すると、未修正折返し翻訳文の評価値が低いほど修正回数が多く、評価値が上昇するにつれて徐々に減少している。各評価値区間における修正 1 回当たりの評価値の増加量

もまた、未修正折返し翻訳文の評価値が低いほど多い。しかし、隣接する評価値区間（「1 以上 2 未満」と「2 以上 3 未満」、「2 以上 3 未満」と「3 以上 4 未満」、「3 以上 4 未満」と「4 以上 5 未満」）の増加量の差はそれぞれ 0.09, 0.06, 0.04 であり、また統計的に有意差はない。したがって、未修正折返し翻訳の評価値における修正 1 回当たりの増加量は隣接した評価値区間ではほぼ同等であり、1 回の修正で大幅に精度が上がるとは考えにくい。

これらの要因として、評価値が低い未修正折返し翻訳文の翻訳リペア過程の観察により下記の傾向が見られた。

- 修正が必要な箇所を複数含んでいる。
- 1, 2 回の修正では直せないような修正が困難な箇所を含んでいる。
- 折返し翻訳の結果から、原因であると考えられる語句を修正しても、修正した内容が他の翻訳部分に影響を及ぼしている。

このように、評価値が低い場合、複数の修正が必要な場合が多くあり、1 回の修正により、大幅な翻訳精度の改善は困難であることが分かった。修正コストの増加は円滑なコミュニケーションを妨げるため、未修正折返し翻訳文の精度を考慮したりペア回数を減らすための支援の仕組みが必要である。

また、入力文中に翻訳リペアが非常に困難な名詞や表現が含まれていた場合に、何とかして修正しようとしたことによる修正回数の増加が被験者の翻訳リペア過程において多数観察された。すなわち、修正困難な語句や表現の有無は、修正コストに大きく影響すると考えられる。このような修正困難な語句や表現がある場合には、言語グリッド [16] のコミュニティ辞書を機械翻訳と関係させることにより、対処が可能であると考えられる [22]。

5.3 翻訳リペアの手法

翻訳リペア実験において、被験者が最も多く行っていった書換え方法は単語や表現の言換えであった。適切な言換えを行うことにより、翻訳精度は向上する。適切な言換えを短時間で考え出すことができれば、より少ない修正コストで翻訳精度を上げることができると考えられる。しかし、一般に適切な言換えを考え出すことは容易ではない。そこで、言換えに関する技術 [23] を利用することにより、より容易な翻訳リペアを行うための支援が可能であると考えられる。

6. む す び

機械翻訳を用いて円滑に多言語コミュニケーションを行うためには、不適切な翻訳箇所の少ない文章を作成する必要があり、翻訳リペアが重要な役割を果たす。本論文では、翻訳リペアのコストに見合うだけの翻訳精度が得られるかについて検証するために折返し翻訳を用いた翻訳リペア実験を行った。本実験では、15 文字以上 32 文字以下の文を翻訳リペアの対象として用いた。また翻訳リペア実験により得られた結果を用いて折返し翻訳文評価実験を行い、精度改善や修正コスト増加に関する考察を行った。

実験により得られた知見は以下のとおりである。

(1) 本実験では、一部の修正が困難な単語を含む文を除いて、約 6 回の翻訳リペアにより、英語・中国語・韓国語の 3 言語の翻訳結果を正しく意味伝達が可能な文へと改善可能であった。3 言語の平均では、コミュニケーションに問題なく利用可能な翻訳精度へと 100 文中 65 文を修正できた。

(2) 未修正折返し翻訳文の精度による修正 1 回当たりの精度の増加量への影響はほとんどなく、未修正折返し翻訳文の精度が低いほど、翻訳リペアの修正コスト（修正回数）は大きくなっている。そのため、1 回の修正で大幅に精度を改善することは難しいと考えられる。また、修正困難な語句・表現を含んでいる場合修正コストが増加し、修正の放棄につながっていた。これらの修正コスト増加要素を考慮した翻訳リペア支援の仕組みが必要である。

今後の課題は、今回の実験結果を踏まえ、効果的な修正方法について検討し、コミュニケーションツールに適用することにより、ユーザがより少ない修正コストで容易に翻訳リペアが行えるかどうかを検証する。

謝辞 本論文に対して、編集委員、査読者の方々から有益な御指摘を頂いた。ここに深く感謝の意を表す。なお本研究の一部は、日本学術振興会科学研究費基盤研究 (B ¥ 19300036) の補助を受けた。

文 献

- [1] Y. Takano and A. Noda, "A temporary decline of thinking ability during foreign language processing," *J. Cross-Cultural Psychology*, vol.24, pp.445-462, 1993.
- [2] M. Aiken, C. Hwang, J. Paolillo, and L. Lu, "A group decision support system for the Asian Pacific rim," *J. International Information Management*, vol.3, pp.1-13, 1994.

- [3] K.J. Kim and C.J. Bonk, "Cross-cultural comparisons of online collaboration," J. Computer Mediated Communication, vol.8, no.1, (online) available from (<http://jcmc.indiana.edu/vol8/issue1/kimandbonk.html>), 2002.
- [4] M. Aiken, "Multilingual communication in electronic meetings," ACM SIGGROUP, Bulletin, vol.23, no.1, pp.18-19, 2002.
- [5] L.L. Tung and M.A. Quaddus, "Cultural differences explaining the differences in results in GSS: Implications for the next decade," Decision Support Systems, vol.33, no.2, pp.177-199, 2002.
- [6] 石田 亨, 内元清貴, 山下直美, 吉野 孝, "機械翻訳を用いた異文化コラボレーション," 情報処理, vol.47, no.3, pp.269-275, March 2006.
- [7] 藤井薫和, 重信智宏, 吉野 孝, "機械翻訳を用いた異文化間チャットコミュニケーションにおけるアノテーションの評価," 情処学論, vol.48, no.1, pp.63-71, Jan. 2007.
- [8] 安岡美佳, 中小路久美代, 大平雅雄, 石田 亨, 野村早恵子, "異文化協調作業における共有理解構築の機会としてのコミュニケーションエラー現象の利用," 情処学研報, 2003-HI-103, pp.47-54, May 2003.
- [9] 坂本知子, 野村早恵子, 石田 亨, 井佐原均, 小倉健太郎, 林 良彦, 石川 開, 小谷克則, 島津美和子, 介弘達哉, 畠中伸敏, 富士 秀, 船越 要, "機械翻訳システムに対する利用者適応の分析—異文化コラボレーションを目指して," 情処学研報, 2003-ICS-135, pp.125-130, March 2004.
- [10] 林田尚子, 石田 亨, "翻訳エージェントによる自己主導型リペア支援の性能予測," 信学論 (D-I), vol.J88-D-I, no.9, pp.1459-1466, Sept. 2005.
- [11] N. Yamashita and T. Ishida, "Automatic prediction of misconceptions in multilingual computer-mediated communication," Proc. 11th International Conference on Intelligent User Interfaces, pp.62-69, 2006.
- [12] 山下直美, 坂本知子, 野村早恵子, 石田 亨, 林 良彦, 小倉健太郎, 井佐原均, "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," 情処学論, vol.47, no.4, pp.1276-1286, April 2006.
- [13] 吉野 孝, 松原繁夫, 喜多千草, 石田 亨, "多言語コミュニケーションツールの異文化間対面協調作業への適用," 第20回人工知能学会全国大会, 3E1-2, 2006.
- [14] K. Ogura, Y. Hayashi, S. Nomura, and T. Ishida, "User adaptation in MT-mediated communication," Proc. 1st International Joint Conference on Natural Language Processing (IJCNLP04), pp.596-601, March 2004.
- [15] <http://www.kodensha.jp/>
- [16] T. Ishida, "Language grid: An infrastructure for intercultural collaboration," IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06), pp.96-100, 2006.
- [17] 金 淵培, 江原暉将, "日英機械翻訳のための日本語長文自動短文分割と主語の補完," 情処学論, vol.35, no.6, pp.1018-1028, June 1994.
- [18] 黒橋禎夫, 長尾 眞, "長い日本語文における並列構造の推定" 情処学論, vol.33, no.8, pp.1022-1031, Aug. 1992.
- [19] 乾 孝司, 乾健太郎, "複数のパーザを利用した統計的部分係り受け解析," 情処学論, vol.42, no.12, pp.1234-1246, Dec. 2001.
- [20] <http://www.kecl.ntt.co.jp/icl/mtg/resources/index.php>
- [21] K. Walker, M. Bamba, D. Miller, X. Ma, C. Cieri, and G. Doddington, Multiple-Translation Arabic (MTA) Part 1, Linguistic Data Consortium (LDC) catalog number LDC2003T18 and ISBN 1-58563-276-7.
- [22] 重信智宏, 藤原義功, 村上陽平, 吉野 孝, 石田 亨, "ドメイン指向翻訳を備えた言語グリッドコミュニケーションツール," 情報処理学会第69回全国大会講演論文集(分冊4), pp.67-68, 2007.
- [23] 藤田 篤, 降幡建太郎, 乾健太郎, 松本裕治, "語彙概念構造に基づく言い換え生成—機能動詞構文の言い換えを例題に," 情処学論, vol.47, no.6, pp.1963-1975, June 2006.
(平成19年3月16日受付, 6月21日再受付)

宮部 真衣



2006 和歌山大・システム工・デザイン情報中退。現在、同大大学院システム工学研究科博士前期課程在学中。多言語コミュニケーション支援に関する研究に従事。

吉野 孝 (正員)



1992 鹿児島大・工・電子卒。1994 同大大学院工学研究科電気工学専攻修士課程了。2004 和歌山大学システム工学部デザイン情報学助教授(2007より准教授)、現在に至る。博士(情報科学)東北大。コラボレーション支援の研究に従事。

重信 智宏 (正員)



2003 和歌山大学大学院システム工学研究科博士前期課程了。2006 同大学院システム工学研究科博士後期課程了。博士(工学)。現在(独)情報通信研究機構研究員。異文化コラボレーションに関する研究に従事。