

翻訳不適箇所の指摘による翻訳リペア効率の改善効果の検証

宮部 真衣^{†1} 吉野 孝^{†2}

機械翻訳を用いたコミュニケーションにおいて、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重要な役割を果たす。一方、翻訳リペア作業はユーザに対して負担が大きく、コミュニケーションへ翻訳リペアを適用する場合、リペア効率の改善が不可欠である。リペア効率の改善手法の1つとして、翻訳不適箇所を指摘する手法が提案されている。従来研究では、精度向上の成功率や適切でない修正をした割合によって、翻訳不適箇所指摘の効果が評価されている。しかし、実際のコミュニケーションへの利用の観点から、修正時間などのコストに関する効率化が不可欠であり、翻訳不適箇所指摘の有効性を議論するためには、コストの軽減効果について検証する必要がある。そこで本論文では、翻訳不適箇所指摘を用いた翻訳リペア実験を行った。本実験では20文字以上30文字以下の文の翻訳リペアを行い、翻訳不適箇所指摘のリペア効率改善効果に関する考察により、以下の知見を得た。(1)本実験では、翻訳不適箇所の指摘により、修正箇所の判定に関する学習期間の短縮効果が見られた。特に、低精度な折り返し翻訳文のリペアにおいて、提案手法による学習期間短縮の高い効果が見られた。しかし、この効果は大きなものではなく、リペア支援を行うためには翻訳不適箇所指摘以外の手法の検討が必要である。(2)被験者の不正確判定率が適切な翻訳リペアの実施に影響している可能性が高く、不正確判定を減少させるための正確な文意一致判定の仕組みが必要であると考えられる。(3)本実験では、コスト軽減の観点から、翻訳不適箇所指摘は大きなコスト軽減に貢献できず、翻訳リペア支援の効果は不十分であることを示した。

Verification of Improvement Effects of Translation Repair Efficiency Using Providing Inaccurately Translated Segments

MAI MIYABE^{†1} and TAKASHI YOSHINO^{†2}

Translation repair using back translation plays an important role in multi-lingual communication using machine translation. The cost of the translation repair is considerable high for the user. The support of efficient translation

repair is required. Providing inaccurately translated-segments is presented as the method of improving translation repair efficiency. This method has been evaluated in the following views: the success rate of improving the translation accuracy and the rate of inadequate repair. However, the reduction of the translation repair costs, such as the repair time, is essential to apply the translation repair in actual communication. In the discussion of the effectiveness of providing inaccurately translated-segments, it is necessary to verify the improvement effects of translation repair cost using this method. We performed the experiment of translation repair to evaluate the effectiveness of providing inaccurately translated segments. In the experiment, we used the sentences that contained 20-30 letters. The results of the experiment are as follows. (1) For users who are inexperienced in translation repair work, the start time of the repair work is improved by providing translation inadequacy part. However, this improvement is not effective in actual communication. To support translation repair in actual communication, we need to develop more efficient methods. (2) The inaccurate-judgment rate can affect the improvement of the efficiency of translation repair. In order to reduce the inaccurate judgments, it is necessary to develop a method that measures the similarity between an input sentence and the back-translated sentence. (3) Providing inaccurately translated-segments cannot contribute to the reduction of the translation repair costs. Therefore, providing inaccurately translated-segments would be not effective to support the translation repair.

1. はじめに

世界規模のインターネットの普及により、インターネットにおける使用言語の多様化が進んでいる。それにともない、ネットワークを介した多言語コミュニケーションの機会が増加している。多言語コミュニケーションを行うためには、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容易ではない。母国語以外の言語を用いたコミュニケーションは困難であり、相互理解ができない可能性が高い^{(1),(2)}。母国語を用いた他言語話者間のコミュニケーションを実現するために、機械翻訳などの技術を利用した取り組みが現在行われている⁽³⁾⁻⁽⁵⁾。

近年、機械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。機械翻訳においては、長文や複雑な構造の文になると不適切な翻訳箇

^{†1} 和歌山大学大学院システム工学研究科

Graduate School of Systems Engineering, Wakayama University

^{†2} 和歌山大学システム工学部

Faculty of Systems Engineering, Wakayama University

所が急速に増加する。円滑な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。翻訳文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。

我々はこれまでに、折り返し翻訳を用いた翻訳リペアにより得られる翻訳結果の精度検証実験を行い、翻訳リペアによって翻訳精度が改善できることを確認した⁶⁾。一方、小倉らによって翻訳リペア作業はユーザに対して負担をかけると指摘されている⁷⁾。コミュニケーションへ翻訳リペアを適用するためには、リペアの効率を改善する必要がある。

リペア効率の改善手法の1つとして、翻訳不適箇所を指摘する手法が提案されている⁸⁾。翻訳不適箇所の指摘は、入力文中の修正すべき箇所をユーザに提示することにより、ユーザの翻訳不適箇所の推定作業を補助するものである。翻訳不適箇所の指摘はリペア作業自体を支援することはできないが、リペアを行うべき箇所の明示により、リペアの効率化が可能であると示されている。しかし、従来研究では、精度向上の成功率や適切でない修正をした割合により、翻訳不適箇所の指摘による効果が評価されており、リペア効率の改善に関する要因分析はなされていない。そのため、翻訳不適箇所の指摘は、どのような点において、どの程度効率改善に貢献できるのかは明らかにされていない。翻訳リペア作業の支援においては、修正時間などのコストに関する効率化が不可欠である。翻訳不適箇所指摘の有効性を議論するためには、コストの軽減効果について検証する必要がある。

そこで、本論文では翻訳不適箇所の指摘を用いた翻訳リペア実験を行い、以下の項目について検証を行う。

- (1) 翻訳不適箇所の指摘を行うことにより、どのような点においてどの程度のリペア効率改善が可能か。
- (2) リペア効率の改善において考慮すべき要因が存在するか。

これらの検証により、翻訳不適箇所指摘のリペア効率改善における有効性および翻訳リペアを適切に行うために必要な支援を明らかにする。

以下、2章において本研究と関連する先行研究について述べる。3章では翻訳不適箇所の指摘を用いた翻訳リペア実験および実験結果を示し、4章で実験結果に関する考察を行う。5章で翻訳不適箇所の指摘の効果について述べる。最後に6章でまとめと今後の課題について述べる。

2. 関連研究

機械翻訳を用いたコミュニケーションにおいて、相互の発言内容理解は翻訳精度に依存する。翻訳精度が低い場合、十分な相互理解ができず、思い違いが発生する⁹⁾。このような思い違いを回避するためには、機械翻訳によって得られる文章の精度を改善することが重要であると考えられる。機械翻訳によって得られる文章を理解できる文章へと改善するアプローチとして、これまでに機械翻訳のための文の簡易化や翻訳リペアの検討が行われている。

機械翻訳性能を向上させる手法として、フィンチらは英語の入力文から不必要な語を取り除く手法（自動簡易化）を提案している¹⁰⁾。フィンチらは、入力文と自動簡易化文を機械翻訳システムに与え、その比較により訳質が向上することを示している。また、自動簡易化処理を評価するため、人手による文の簡易化との比較を行い、同等の性能が得られることを示している。この研究においては、英語文の簡易化を実現しているものの、日本語文への適応は難しく、さらに別の処理が必要であると述べられている。また、文の簡易化を行うことにより、機械翻訳精度は改善しているが、元の文に含まれていた詳細な情報が欠落する可能性があると考えられる。

翻訳リペアに関する先行研究については、母国語だけを用いてリペア作業を行う方法として、折り返し翻訳の利用の検討が行われている¹¹⁾。折り返し翻訳を用いた文章書き換え作業の実験により、折り返し翻訳が母国語によるリペア手段として有効であると示されている。そこで、本論文では折り返し翻訳を利用した翻訳リペアにより検証実験を行い、その実験結果により議論を進める。

3. 翻訳不適箇所指摘を用いた翻訳リペア実験

翻訳リペアを効率的に行うためには、ユーザが適切なリペアを行うことができるように支援する必要がある。翻訳リペア作業においては、入力文と折り返し翻訳結果を比較し、修正すべき箇所を的確に判断できなければ、適切なリペアを行うことができない。適切なリペアが行われなかった場合、リペア効率が低下する。具体的には、修正着手の遅れ、修正時間の増加が発生すると考えられる。

そこで、リペア効率改善手法の1つである翻訳不適箇所指摘を利用した場合の翻訳リペア効率の改善効果を検証する。本論文では、「リペア着手時間」「修正時間」を翻訳リペア効率の指標とする。翻訳不適箇所指摘を用いた翻訳リペア実験を行い、翻訳リペア効率の改善度合いについて、以下の仮説について検証を行う。

[仮説 1]: 翻訳不適箇所の指摘により, 修正着手に要する時間(リペア着手時間)が短縮される.

[仮説 2]: 翻訳不適箇所の指摘により, 修正時間が短縮される.

3.1 実験内容

翻訳不適箇所の指摘による翻訳リペア効率の改善効果を検証するために, 翻訳不適箇所指摘を用いた翻訳リペア実験を実施した. 実験の被験者は, 和歌山大学システム工学部および大学院の学生 30 名である.

翻訳不適箇所指摘を用いた翻訳リペア実験を行う前に, 日本語検定公式模擬・練習問題集¹²⁾の中級テストを実施した. テスト後, 実験を行う際に折り返し翻訳および実験ツールの使い方について説明した.

本実験では, 以下の 2 項目の作業を行った.

- (1) 翻訳不適箇所の指摘がある状態での翻訳リペア作業
- (2) 翻訳不適箇所の指摘がない状態での翻訳リペア作業

各項目について 100 文ずつ, 計 200 文の翻訳リペア作業を行った. また, 1 回目に指摘なし, 2 回目に指摘ありでリペアを行った被験者と, 逆の順序でリペアを行った被験者は各 15 名である.

本実験で使用したツールの画面例を図 1 に示す. 入力エリア(図 1(4))に入力した文の折り返し翻訳結果が, 折り返し翻訳表示エリア(図 1(2))に表示される. 翻訳システムは, 言語グリッド¹³⁾を介して高電社の J-Server¹⁴⁾を使用した. 本実験では, 機械翻訳システ

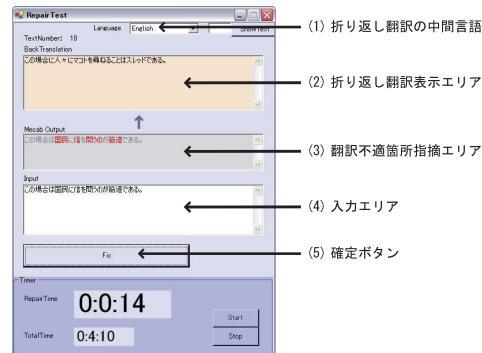


図 1 実験ツールの画面例

Fig. 1 Screenshot of an experimental tool for translation repair.

ムによる性能への依存性を減少させるために, 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の 3 種の機械翻訳システムを利用した. なお, 今回の実験では, 被験者は対象言語の翻訳結果を見ることはできない. 翻訳対象言語に対する習熟度は被験者に依りて異なり, また, 3 言語(英語, 中国語, 韓国語)すべてを被験者が理解できるとは限らない. そのため, 全被験者の実験条件を統一するため, 対象言語の翻訳結果は提示せず, 折り返し翻訳のみを提示することとした. また, 形態素解析により推定した翻訳不適箇所は, 翻訳不適箇所指摘エリア(図 1(3))に表示される. 入力エリアの内容が書き換えられると, 自動的に折り返し翻訳が行われ, 折り返し翻訳表示エリアおよび翻訳不適箇所指摘エリアの内容も更新される. 指摘なしの実験では, 翻訳不適箇所指摘エリアは非表示とした. 翻訳不適箇所の指摘については, 形態素解析を利用した単語の比較により, 翻訳不適箇所を推定する仕組みを用いた. なお, 今回は推定の対象を名詞と動詞とした. 翻訳不適箇所指摘の流れを図 2 に示す. 以下の手順により翻訳不適箇所の指摘を行った.

- (1) 機械翻訳により, 入力文(「彼女は手紙を送ることにより分かれを告げた。」)の折り

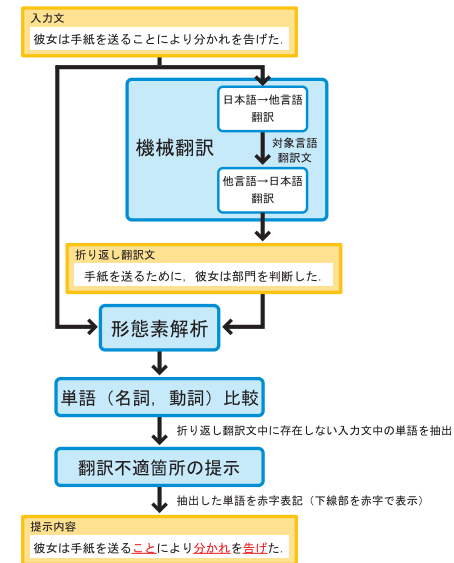


図 2 翻訳不適箇所指摘の流れ

Fig. 2 Procedure of providing translation inadequacy parts.

返し翻訳文(「手紙を送るために、彼女は部門を判断した。’)を取得する。

- (2) 形態素解析器 MeCab¹⁵⁾ を利用し、入力文および折り返し翻訳文の形態素解析を行う。
- (3) 入力文中の単語(名詞、動詞)のうち、折り返し翻訳文中に存在しない単語を抽出する。
- (4) 抽出した単語を翻訳不適箇所と判断し、赤字で強調表示する(「こと」「分かれ」「告げ」を赤字で表示)

ユーザは、入力文中の強調表示された箇所と折り返し翻訳文を確認し、正しく翻訳されていなければ該当箇所の修正を行う。

先行研究により、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ている¹⁶⁾。また、これまでに行った実験では、20文字以下の文についても高精度であり、修正が不要な文章が多かったため、本実験では20文字を実験対象の下限とする。一方で、機械翻訳システムでは、50文字以上の日本語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文²⁰⁾の文字数については、30文字以下の文が全体の94.7%を占めている。また、これまでの実験において32文字以下の文については、リペア作業の適用可能性を確認できているため、本実験における実験対象の上限を30文字とする。以上のことから、本実験では20文字以上30文字以下の日本語文について翻訳リペア作業を行う。実験では、20文字以上30文字以下であり、リペア作業が必要である文を200文選択したものを利用した。実験で利用した200文のうち、10文のテキストを表1に示す。また、200文の文字数の分布を図3に示す。

実験の手順を以下に示す。

- 1) 折り返し翻訳結果を見ながら、原文と同じ意味になるように入力文を修正する。
- 2) 折り返し翻訳結果が同じ意味だと判断したら、次の文の修正を行う。
- 3) 1)~2)の手順を100文行う。
- 4) 翻訳不適箇所指摘の有無を切り替え、別の100文について1)~3)の手順を行う。

一番初めの折り返し翻訳結果(未修正の翻訳結果)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。

表1 実験に用いたテキストの一部

Table 1 Examples of sentences used in the experiment.

(1)	首相は経済をめぐる諸問題について語った。
(2)	ここで私は体積の変化は考えないものとする。
(3)	これは汎用機用とパソコン用を合わせたものである。
(4)	この写真は兄弟が二人仲良く並んだものです。
(5)	母親というものは自分の子供を弁護するものだ。
(6)	上司は部下に対して外国に出張することを命令する。
(7)	彼女は手紙を送ることにより分かれを告げた。
(8)	データが表示されることはないものとする。
(9)	駝鳥が飛ばないことを除いては、鳥は飛ぶ。
(10)	その案は費用を沢山要することから、採用を見送られた。

これらのテキストは、機械翻訳機能試験文²⁰⁾から20文字以上30文字以下である文を200文選択したものの一部である。

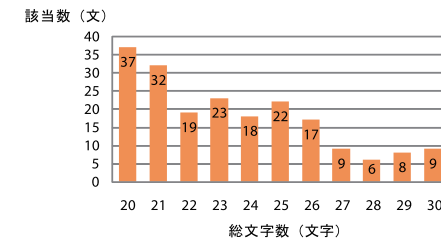


図3 実験に利用した文の総文字数の分布

Fig. 3 Distribution of the total number of characters in the experiment sentences.

3.2 実験結果

指摘の有無の切替え前のリペア(1~100文)を1回目、切替え後のリペア(101~200文)を2回目とする。今回の実験では、被験者30名中4名の正確なデータが取得できなかったため、4名を除いた26名の実験結果により検証を行う。1回目に指摘なし、2回目に指摘ありでリペアを行った被験者は12名(以下グループ1)、1回目に指摘あり、2回目に指摘なしでリペアを行った被験者は14名(以下グループ2)である。

各グループにおける各指標の平均、標準偏差および指摘の有無における差の有意確率を表2に示す。

- (1) リペア着手時間

表2より、1回目については、指摘ありのリペア着手時間が平均3秒早く、有意水準

表 2 リペア着手時間と修正時間

Table 2 Delay in the start of the translation repair and repair time in repair work.

		1 回目		2 回目	
		グループ 1 指摘なし (分:秒)	グループ 2 指摘あり (分:秒)	グループ 1 指摘あり (分:秒)	グループ 2 指摘なし (分:秒)
リペア着手時間	平均	00:23	00:20	00:20	00:21
	標準偏差	00:20	00:29	00:19	00:24
	有意確率	0.000*		0.584	
修正時間	平均	01:48	01:43	01:27	01:23
	標準偏差	01:36	01:22	01:14	01:08
	有意確率	0.988		0.462	

1%で有意差が見られる。2 回目においては、指摘ありのリペア着手時間が平均 1 秒早いですが、指摘の有無において有意差は見られない。

(2) 修正時間

表 2 より、各実施段階において、指摘の有無における修正時間には有意差が見られない。すなわち、指摘の有無は修正時間の短縮に影響していないと考えられる。

3.3 アンケート結果

翻訳不適箇所指摘を用いた翻訳リペアに関するアンケート調査を後日行った。アンケート調査は、メールにより被験者へアンケートを送信し、回答もメールによって返信してもらう形式で実施した。

アンケートの質問項目は以下の 2 項目である。

質問 1：翻訳不適箇所の指摘がある状態とない状態では、どちらが修正しやすかったか？

質問 2：修正を行う際に、翻訳不適箇所の指摘をよく参照した。

質問 1 については、2 つの選択肢（(1) 指摘がある状態、(2) 指摘がない状態）からどちらか一方を選択し、回答する形式とした。質問 2 については、5 段階評価（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）からあてはまるものを選択し、回答する形式とした。また、どちらの質問についても選択した理由を記述してもらった。

被験者 30 名のうち、17 名からアンケートの回答を得た。質問 1 については、17 名中 15 名（88%）が指摘のある状態が修正しやすいと回答した。表 3 に回答の選択理由を示す。なお、類似した選択理由が多かったため、表 3 には代表的な理由のみを提示した。表 3 より、指摘ありを選択した被験者は指摘によるメリットを、指摘なしを選択した被験者はデメリット

表 3 質問 1 の回答理由

Table 3 Reasons for answer of question 1.

質問 1：翻訳不適箇所の指摘がある状態とない状態では、どちらが修正しやすかったか？

選択項目	記述内容
指摘あり	<ul style="list-style-type: none"> ・何も指摘されない状態で修正するのは手探りになってしまい、指摘があるときよりも難しかったため。 ・修正すべき箇所の見当を付けやすかったため。 ・うまくいかないことはあるものの、指摘された方が楽に作業ができたため。
指摘なし	<ul style="list-style-type: none"> ・指摘箇所の修正により、翻訳精度がより低下してしまうことが多かったため。 ・指摘箇所ばかり意識させられてしまい、リペア作業が進まないことが多かったため。

表 4 質問 2 の回答理由

Table 4 Reasons for answer of question 2.

質問 2：修正を行う際に、翻訳不適箇所の指摘をよく参照した。

選択評価	記述内容
5	<ul style="list-style-type: none"> ・まず最初に指摘箇所が修正すべき部分であると検討をつけてリペア作業を行っていた。 ・わかりやすく的確に指摘されていたため。
4	<ul style="list-style-type: none"> ・どこから直せばいいのかわからないため、まずは指摘された部分から修正しようと思ったため。 ・まずは指摘を見て修正していた。ただし、慣れてくると指摘されていなくても問題なく修正できるようになった。
3	<ul style="list-style-type: none"> ・指摘を参照した場面もあったが、基本的にはあまり参照しなかったため。 ・修正箇所を悩んだ場合は参照したが、簡単な文の場合は指摘の有無にかかわらず修正できた。
2	<ul style="list-style-type: none"> ・どうしても直らない場合のみ参照したが、あまり参照しなかった。

表中の評価値は 5 段階評価（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）による。

トをあげている。メリットとしては、ある程度修正すべき箇所の目安になり、容易に作業ができたという回答が多く得られた。デメリットとしては、翻訳不適箇所指摘では修正の妨げになったという回答が得られた。

また、質問 2 については、平均評価が 3.9（標準偏差 0.9）となっており、比較的参照が行われていたと考えられる。表 4 に回答理由を示す。表 4 より、あまり参照していなかった被験者（選択評価値が 2 または 3）は、困ったときに参照したが、それほど参照していないと回答した。しかし、参照していた被験者（選択評価値が 4 または 5）は、まず指摘された箇所を参照したと回答しており、すぐに修正箇所が判別できない場合には、翻訳不適箇所の指摘は必要とされると考えられる。

以上の結果から、翻訳不適箇所の指摘により、被験者は比較的容易に作業ができていたと考えられる。

4. 考察

4.1 仮説に関する考察

実験を行うにあたり、2つの仮説を立てた。それぞれの仮説に関して、考察を行う。

[仮説1] 翻訳不適箇所の指摘を行うことにより、修正着手に要する時間（リペア着手時間）が短縮される。

表2より、1回目の実施については、翻訳不適箇所の指摘により、リペアへの着手が平均3秒早くなっている。また、2回目の実施については、指摘の有無において有意差は見られず、指摘の有無にかかわらず、同程度の時間で着手していると考えられる。したがって、翻訳不適箇所の指摘によって、リペアに慣れていない段階においても、リペアに慣れたと考えられる2回目のリペア着手時間と同等の時間で着手できていたと考えられる。すなわち、翻訳不適箇所の指摘を行うことにより、修正箇所の判定に関する学習期間が短縮される効果があると考えられる。

[仮説2] 翻訳不適箇所の指摘を行うことにより、修正時間が短縮される。

表2より、修正時間に関しては、翻訳不適箇所指摘の有無による有意差は見られなかった。したがって、指摘の有無は修正時間の短縮に影響しておらず、翻訳不適箇所の指摘では修正時間を短縮させることは難しいと考えられる。

修正時間は、行った修正により翻訳精度が改善しなかった場合に増加する。翻訳精度が改善されない修正には、修正する必要のない箇所の修正と、適切に翻訳されない修正があると考えられ、翻訳不適箇所の指摘は、前者の修正の減少が可能であると考えられるが、後者の修正の減少は難しい。リペアにおいては、適切に翻訳されない修正が多数行われており、修正時間を短縮させることができなかつたと考えられる。

4.2 被験者の不正確判定率

未修正の折り返し翻訳文の精度の主観評価を Walker らの適合性評価²¹⁾を用いて行った。Walker らの適合性評価は5段階評価であり、2名以上の評価者により行われる。評価基準は、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 (意味は何となくつかめる), 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 (まったく違う意味), である。本実験では、3名の評価者により評価を行った。また、評価結果が4以上であった文を「精度が高く修正が不要である文」とする。本実験で使用したテキスト200文のうち、精度が高く修正が不要であった文の数は英語が3文、中国語が1文、韓国語が6文であり、それ以外の文については何らかの修正を行う必要がある。

表5 日本語検定および不正確判定率の平均値

Table 5 Average of Japanese language proficiency test results and average of the inaccurate-judgment rates.

	日本語検定		不正確判定率	
	平均(点)	標準偏差(点)	平均	標準偏差
グループ1	58.2	7.0	0.078	0.062
グループ2	62.5	8.3	0.064	0.057
全体	60.5	7.9	0.070	0.058

表6 日本語検定の結果と各指標の相関

Table 6 Correlation between Japanese language proficiency test result and each measure of repair efficiency.

	相関係数	有意確率
リペア着手時間(1回目)	-0.348	0.082
リペア着手時間(2回目)	-0.248	0.222
修正時間(1回目)	-0.001	0.996
修正時間(2回目)	-0.001	0.997

修正不要文を除いたテキストにおいて、被験者が何も修正せずに修正回数0で確定していた場合、被験者の文意一致判定が緩く、正確な判定がされていないと考えられる。

そこで、本来修正すべき文を、被験者が修正不要であると判断した割合を不正確判定率とし、以下の計算式により定義する。

$$\text{不正確判定率} = \frac{(\text{修正回数0の文数}) - (\text{修正不要文の該当数})}{(\text{総テキスト数}) - (\text{修正不要文数}) - (\text{放棄文数})}$$

表5に実験前に実施した日本語検定結果および不正確判定率の平均値を示す。不正確判定率の平均は0.070であり、100文のリペアにおいて、平均7文は正確な判定がなされていないことを意味する。また、最大値は0.231、最小値は0であった。

日本語検定の結果と不正確判定率の相関係数は-0.255、有意確率は0.209であり、日本語検定の結果と不正確判定率は無相関である。すなわち、不正確判定率は、日本語の能力に依存しないと考えられる。

また、日本語検定の結果と各指標の相関を表6、実験結果から求めた不正確判定率と各指標の相関を表7に示す(5%水準で有意であったものには、*印を付加した)。表6より、日本語検定の結果はどの指標とも相関がない。一方、表7から、不正確判定率については、修正時間についての相関係数は、1%水準で有意であり、各実施段階の修正時間は不正確判定率と負の相関関係にある。

表 7 不正確判定率と各指標の相関

Table 7 Correlation between the inaccurate-judgment rate and each measure of repair efficiency.

	相関係数	有意確率
リペア着手時間(1回目)	-0.271	0.181
リペア着手時間(2回目)	-0.072	0.725
修正時間(1回目)	-0.580	0.002*
修正時間(2回目)	-0.656	0.000*

表 8 修正された文の精度と日本語検定結果、不正確判定率の相関

Table 8 Correlation coefficients of the evaluated value.

		相関係数	有意確率
日本語検定の結果	修正された文の精度(1回目)	0.050	0.808
	修正された文の精度(2回目)	0.100	0.628
不正確判定率	修正された文の精度(1回目)	-0.510	0.008*
	修正された文の精度(2回目)	-0.595	0.001*

また、実験において修正された文の精度^{*1}と日本語検定結果、不正確判定率との相関を表 8 に示す。表 8 より、修正された文の精度と不正確判定率には負の相関関係が見られる。

以上のことから、不正確判定率が高いほど短時間の修正となり、精度が改善されなくなると考えられる。すなわち、ユーザが短時間でリペアを終了しても、精度が改善していない場合もあり、リペア効率の改善とともに、ユーザの不正確な判断による不十分なりペア状態での修正終了を防ぐ仕組みを検討する必要がある。

5. 翻訳不適箇所指摘の効果と今後の課題

本章では、従来研究における知見と本論文で得られた結果について議論する。また、今回の実験結果から得られた今後の課題について述べる。

5.1 従来研究における知見と本論文の実験結果

これまでに、林田らは理想的な翻訳不適箇所の指摘を行った状態におけるリペア実験を行っている⁸⁾。林田らの実験では、折り返し翻訳を用いず、英語の翻訳結果を見ながらリペアを行っている。従来研究では、翻訳不適箇所の指摘だけでは、翻訳に適した文章の作成な

どのリペア作業自体の支援は困難であるものの、翻訳不適箇所の指摘により、見込みのない変更をする可能性が減少しており、効率の良いリペア作業のために有効であることを示している。しかし、この知見は精度向上の成功率および適切でない修正をした割合により議論された内容であり、リペア作業の支援において重要である、コスト軽減に関する観点からの議論はなされていない。

本論文では、リペアコストの面から翻訳不適箇所指摘の効果について検証を行った。3.3 節で示したアンケート結果から、アンケート回答者の 88%は翻訳不適箇所の指摘のある方がリペアを行いやすいと回答しているが、実験結果から、翻訳不適箇所の指摘によるリペアコストの改善はできていないことが分かった。4.1 節で述べたように、今回の実験ではリペア着手時間の短縮効果が見られた。しかし、翻訳不適箇所の指摘によるリペア着手時間の有意な差は見られたものの、その短縮効果は大きなものではない。また、実際のコミュニケーションへの適用において問題となる修正時間については、翻訳不適箇所の指摘による短縮効果は見られなかった。翻訳リペアをコミュニケーションへ適用するためには、修正時間の短縮が不可欠である。すなわち、翻訳不適箇所の指摘を用いただけでは、実際のコミュニケーションへの翻訳リペアの適用は難しいと考えられる。

これらの実験結果より、本論文の知見として、コスト軽減の観点において、翻訳不適箇所指摘は大きなコスト軽減に貢献できず、翻訳リペア支援の効果は不十分であることを示した。

5.2 今後の課題

翻訳リペアを実際に適用するためには、以下の問題への対応が必要である。

(1) リペア作業の支援

従来研究でも述べられているように、翻訳不適箇所の指摘だけでは、翻訳に適した文章の作成などのリペア作業自体の支援は困難である⁸⁾。今回の実験結果においても、翻訳不適箇所の指摘による修正時間の短縮効果はなく、翻訳不適箇所指摘以外の支援が必要であると考えられる。

翻訳結果の精度検証実験において、被験者が最も多く行っていた書き換え方法は単語や表現の言い換えであった。適切な言い換えを行うことにより、翻訳精度は向上する。適切な言い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、一般に適切な言い換えを考え出すことは容易ではない。そこで、類語の提示による言い換えの支援が考えられる。インターネット上の類語辞典などを利用し、うまく翻訳されない単語の類語を提示することで、ユーザのリペア作業を支援できる可能性がある。また、言い換えに関する技

*1 未修正の折り返し翻訳結果の精度評価と同様に、Walker らの適合性評価²¹⁾によって、被験者により修正された文の精度評価を行った。

術²²⁾を利用することによって、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。

(2) 不正確判定の防止

本論文では、不正確判定率を定義した。翻訳リペアは高精度な文章を作成するための手法であり、翻訳リペアを効果的に適用するためには、ユーザの不正確な判断による不十分なりペア状態での修正終了を防ぐ必要がある。

これまでに、翻訳システムの評価のために、BLEU²³⁾ や NIST²⁴⁾ など様々な翻訳自動評価手法の提案や評価がなされている。これらの評価手法は基本的には対象言語への翻訳結果とバイリンガルが作成した対象言語の参照訳とを比較し、類似度を計算するものであるが、折り返し翻訳結果と原文の比較による手法も提案されている²⁵⁾。本研究では折り返し翻訳を利用しており、後者の手法を利用して類似度計算を行うことが可能である。折り返し翻訳結果と入力文の類似度計算を行うことにより、十分な類似度を持たない折り返し翻訳結果については、さらなるリペアを求めることが可能になると考えられる。

6. おわりに

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と考えられる。一方、翻訳リペア作業はユーザに対して負担をかけるため、リペア効率の改善が必要である。

本論文では、リペア効率改善手法の1つである「翻訳不適箇所の指摘」を用いた場合の効率化が可能な点および効率の改善度合いを検証するために、翻訳不適箇所指摘を用いた翻訳リペア実験を行った。本実験では20文字以上30文字以下の文を対象として用いた。実験結果から翻訳不適箇所指摘による効率改善効果に関する考察を行った。

本研究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、翻訳不適箇所の指摘を作業に慣れていない段階において行うことにより、リペアに着手する時間が早くなっており、修正箇所の判定に関する学習期間が短縮される効果があると考えられる。特に低精度な折り返し翻訳文のリペアにおいて、翻訳不適箇所の指摘によるリペア着手時間の短縮効果が高いと考えられる。しかし、この効果は大きなものではなく、リペア支援を行うためには翻訳不適箇所指摘以外の手法の検討が必要である。
- (2) 被験者の不正確判定率が翻訳リペア効率改善に影響している可能性が高く、不正確判

定を減少させるための正確な文意一致判定の仕組みが必要であると考えられる。

- (3) 従来研究においては、精度向上の成功率および適切でない修正をした割合の観点から、翻訳不適箇所の指摘だけでは、翻訳に適した文章の作成などのリペア作業自体の支援は困難であるものの、翻訳不適箇所の指摘により、見込みのない変更をする可能性が減少しており、効率の良いリペア作業のために有効であることが示されていた。しかし、本実験では、コスト軽減の観点から、翻訳不適箇所指摘は大きなコスト軽減に貢献できず、翻訳リペア支援の効果は不十分であることを示した。

今後は、リペア作業自体の支援手法についての検討が必要である。また、不十分なりペア状態での修正終了を防ぐために、類似度計算などを利用した、不正確判定防止の仕組みを検討する必要がある。

謝辞 本論文に対して、査読者の方々から有益なご指摘をいただいた。ここに深く感謝の意を表す。本研究のデータ分析において、貴重な助言や指摘をいただいたNTTコミュニケーション科学基礎研究所の山下直美博士に深く感謝する。本研究の実験において、機械翻訳を利用するために言語グリッドを提供していただいた言語グリッドプロジェクトに深く感謝する。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。

参考文献

- 1) Aiken, M.: Multilingual Communication in Electronic Meetings, *ACM SIG-GROUP, Bulletin*, Vol.23, No.1, pp.18-19 (2002).
- 2) Tung, L.L. and Quaddus, M.A.: 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).
- 3) Fujii, K., Yoshino, T., Shigenobu, T. and Munemori, J.: Development of an Intercultural Collaboration System with Semantic Information Share Function, *Proc. 9th International Conference on Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies (KES 2005)*, LNAI 3681, pp.425-430, Springer Verlag (2005).
- 4) Sakai, S., Gotou, M., Tanaka, M., Inaba, R., Murakami, Y., Yoshino, T., Hayashi, Y., Kitamura, Y., Mori, Y., Takasaki, T., Naya, Y., Shigeno, A., Matsubara, S. and Ishida, T.: Language Grid Association: Action Research on Supporting the Multicultural Society, *International Conference on Informatics Education and Research for Knowledge-Circulating Society (ICKS-08)* (2008).
- 5) Inaba, R.: Usability of Multilingual Communication Tools, *Proc. Lecture Notes in*

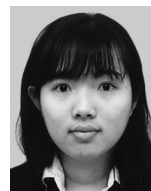
Computer Science, Vol.4560, pp.91-97 (2007).

- 6) 宮部真衣, 吉野 孝, 重信智宏: 折返し翻訳を用いた翻訳リペアの効果, 電子情報通信学会論文誌, Vol.J90-D-I, No.12, pp.3142-3150 (2007).
- 7) 小倉健太郎, 林 良彦, 野村早恵子, 石田 亨: 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, 第 65 回情報処理学会全国大会論文集 (2003).
- 8) 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, 電子情報通信学会論文誌, Vol.J88-D-I, No.9, pp.1459-1466 (2005).
- 9) Yamashita, N. and Ishida, T.: Automatic prediction of misconceptions in multilingual computer-mediated communication, *Proc. 11th International Conference on Intelligent User Interfaces*, pp.62-69 (2006).
- 10) フィンチ・アンドリュー, 下畑光夫, 隅田英一郎: 機械翻訳のための文簡易化, 情報処理学会研究報告, 2004-NL-165, pp.73-78 (2005).
- 11) 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, 情報処理学会論文誌, Vol.47, No.4, pp.1276-1286 (2006).
- 12) 日本語検定委員会: 日本語検定公式模擬・練習問題集, 東京書籍 (2007).
- 13) Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration, *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp.96-100 (2006).
- 14) 高電社. <http://www.kodensha.jp/>
- 15) Kudo, T., Yamamoto, K. and Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis, *Proc. 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004)*, pp.230-237 (2004).
- 16) Shigenobu, T.: Evaluation and Usability of Back Translation for Intercultural Communication, *Proc. Lecture Notes in Computer Science*, Vol.4560, pp.259-265 (2007).
- 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, 情報処理学会論文誌, Vol.35, No.6, pp.1018-1028 (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, 情報処理学会論文誌, Vol.33, No.8, pp.1022-1031 (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, 情報処理学会論文誌, Vol.42, No.12, pp.1234-1246 (2001).
- 20) NTT Natural Language Research Group.
<http://www.kecl.ntt.co.jp/icl/mtg/resources/index.php>

- 21) Walker, K., Bamba, M., Miller, D., Ma, X., Cieri, C. and Doddington, G.: Multiple-Translation Arabic (MTA) Part 1, Linguistic Data Consortium (LDC) catalog number LDC2003T18 and ISBN 1-58563-276-7.
- 22) 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, 情報処理学会論文誌, Vol.47, No.6, pp.1963-1975 (2006).
- 23) Papineni, K., Roukos, S., Ward, T. and Zhu, W.-J.: BLEU: A Method for Automatic Evaluation of Machine Translation, *Proc. 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp.311-318 (2002).
- 24) NIST: Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics, Technical Report, NIST (2002).
- 25) Uchimoto, K., Hayashida, N., Ishida, T. and Isahara, H.: Automatic Rating of Machine Translatability, *10th Machine Translation Summit (MT Summit X)*, pp.235-242 (2005).

(平成 20 年 6 月 13 日受付)

(平成 21 年 1 月 7 日採録)



宮部 真衣 (学生会員)

昭和 59 年生。平成 18 年和歌山大学システム工学部デザイン情報学科中退。平成 20 年同大学大学院システム工学研究科システム工学専攻博士前期課程修了。現在、同大学院システム工学研究科システム工学専攻博士後期課程在学中。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 (正会員)

昭和 44 年生。平成 4 年鹿児島大学工学部電子工学科卒業。平成 6 年同大学大学院工学研究科電気工学専攻修士課程修了。現在、和歌山大学システム工学部デザイン情報学科准教授。博士 (情報科学)。コラボレーション支援に関する研究に従事。