

# 翻訳不適箇所指摘による翻訳リペア効率の改善効果 Improvement Effect of Translation Repair Efficiency on Providing Translation Inadequacy Part

宮部真衣<sup>†</sup>  
Mai Miyabe

吉野 孝<sup>†,‡</sup>  
Takashi Yoshino

## 1. はじめに

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．多言語を十分に習得することは困難であるため，機械翻訳などの技術が利用されている [1]．近年翻訳技術は急速に進展しているが，完璧な翻訳を行うことは困難であり，不適切な翻訳箇所の少ない文を作成する必要がある．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．

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る翻訳精度改善効果の検証を行った [2]．検証により，15 文字以上 32 文字以下の文章では，一部の文章を除いて，約 7 回の翻訳リペアにより 3 言語（英語，中国語，韓国語）の翻訳結果の改善が可能であることがわかった．しかし，適切なりペアを行えるかどうかは個人の能力に依存する可能性が高いことがわかった．

そこで，形態素解析を利用した翻訳不適箇所の指摘による翻訳リペア実験を行った．本稿では翻訳不適箇所指摘の有無による翻訳リペア効率の改善効果の検証を行う．

## 2. 形態素解析を利用した翻訳不適箇所指摘

翻訳リペアを行う場合，修正すべき箇所がわからなければ，適切に修正することができない．林田らの研究により，翻訳不適箇所の提示はリペア効率の向上に有効であると示されている [3]．また，入力文と折り返し翻訳文の類似度計算により翻訳不適箇所の自動推定を行う，入力言い換え支援の研究が行われている [1]．

本研究では，入力文と折り返し翻訳文の形態素解析を行い，単語の比較により翻訳不適箇所を推定する．翻訳不適箇所指摘の流れを図 1 に示す．本研究では，以下の手順により翻訳不適箇所の指摘を行う．

1. 機械翻訳により，入力文（「ここで私は体積の変化は考えないものとする」）の折り返し翻訳文（「私達は，私がここでボリュームの変化を考慮しないと思う」）を取得する．
2. 形態素解析器 Mecab[4] 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し，赤字で強調表示する（「体積」「考え」「もの」を赤字で表示）

## 3. 翻訳リペア実験

### 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，翻訳リペア実験を実施した．実験の

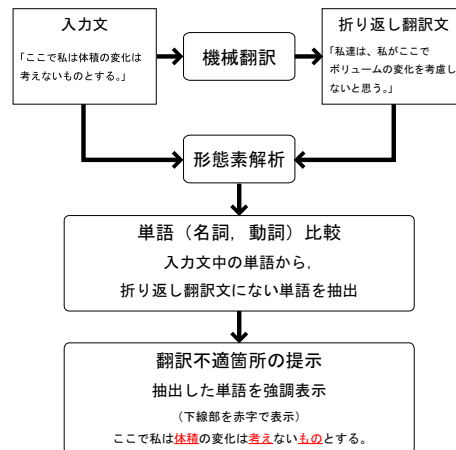


図 1: 翻訳不適箇所指摘の流れ

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．実験を行う際，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．各被験者で 2 項目の実施順序は異なる．

本実験で使用したツールの画面例を図 2 に示す．入力エリア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に表示される．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に表示される．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．翻訳システムは，言語グリッド [5] を介して高電社の J-Server[6] を使用した．

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．実験で利用した文は，機械翻訳機能試験文 [7] のうち，20 文字以上 30 文字以下である文を 200 文選択したものである．

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，確定ボタン（図 2 (3)）を押し次の文を表示する．
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．

1 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

<sup>†</sup>和歌山大学

<sup>‡</sup>情報通信研究機構

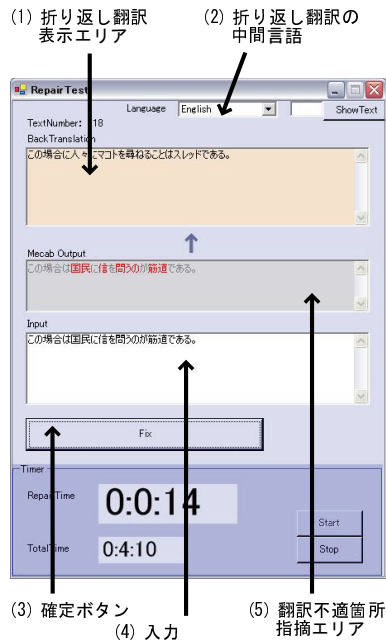


図 2: 実験ツールの画面例

表 1: 修正放棄数

	1 回目 (文)	2 回目 (文)
平均	20.7	16.8
標準偏差	11.3	10.6
有意確率	0.007	

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 6～12 時間を要し、実験中は適宜休憩をとるように指示した。

### 3.2 実験結果

翻訳リペアにおいて、修正すべき箇所がわからなければ適切に修正することができず、(1) 修正の放棄、(2) 修正回数の増加、(3) 修正結果の精度低下が発生する。本稿では、(1) 修正の放棄をリペア効率の指標とし、指摘の有無による修正放棄数の差異について検証を行う。

指摘の有無の切り替えによる 100 文ずつの実施段階 (1～100 文, 101～200 文) をそれぞれ 1 回目, 2 回目とする。各実施段階の修正放棄数およびウィルコクソンの符号付順位検定による 1 回目と 2 回目の有意確率を表 1 に示す。表 1 の平均値は被験者 30 名の平均値である。

1 回目, 2 回目の平均放棄数はそれぞれ 20.7 文, 16.8 文となっている。1 回目と 2 回目には有意水準 1% で有意差があり、2 回目のリペアで放棄数が減少している。

また、指摘の有無における放棄数、マン・ホイットニーの検定による指摘の有無における有意確率を表 2 に示す。各実施段階における指摘なし、指摘ありの被験者はそれぞれ 15 名である。表 2 より、各実施段階において、指摘の有無における放棄数には有意差が見られない。すなわち、指摘の有無は修正放棄数の減少に影響していないと考えられる。

2 回目のリペアにおける放棄数の減少の要因としては、次の項目が考えられる。

表 2: 各実施段階における放棄数

	1 回目		2 回目	
	指摘なし (文)	指摘あり (文)	指摘なし (文)	指摘あり (文)
平均	20.8	20.6	16.1	17.5
標準偏差	11.8	11.1	9.7	11.8
有意確率	0.771		0.934	

- 被験者の翻訳リペアへの慣れ
- 実験で使用した入力文の修正し易さの偏り

2 回目のリペアにおける放棄数の減少にこれらの項目が影響しているか、今後検証する必要がある。

## 4. 考察

林田らはリペア支援機能として折り返し翻訳、翻訳不適箇所の強調表示、修正教示を提案し、理想的な表示を行った場合の効果を検証している [3]。林田らの研究では、強調表示だけでは適切な修正方法まではわからないものの、リペア作業の効率化において強調表示は有効であることが示されている。

本実験では、翻訳不適箇所指摘の有無を切り替え、実験を行ったが、修正放棄数において指摘の有無による差異は見られなかった。今後、修正回数、修正結果の精度における翻訳不適箇所指摘の有無による差異の検証および翻訳不適箇所指摘の精度検証を行う必要がある。

## 5. おわりに

本稿では、翻訳不適箇所指摘による翻訳リペア実験を行い、リペア効率の改善効果について修正放棄数により考察を行った。今回の実験では、翻訳不適箇所指摘による修正放棄数の減少は見られなかった。

今後は、修正回数および修正結果の精度による分析を行う。また、翻訳不適箇所指摘の精度を検証し、ユーザの翻訳リペア効率を改善するための機能を検討する。

謝辞

本研究は、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] 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,” 情報処理学会会誌, Vol.47, No.3, pp.269-275, March 2006.
- [2] 宮部真衣, 吉野 孝, 重信智宏, “折り返し翻訳を用いた翻訳リペア効果の評価,” 電子情報通信学会, 思考と言語研究会, TL2006-38, pp.43-48, Nov. 2006.
- [3] 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, Sep.2005.
- [4] Mecab, <http://mecab.sourceforge.jp/>
- [5] Toru Ishida, “Language Grid: An Infrastructure for Intercultural Collaboration,” IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06), pp.96-100, 2006.
- [6] 高電社, <http://www.kodensha.jp/>
- [7] NTT Natural Language Research Group, <http://www.kecl.ntt.co.jp/icl/mtg/resources/index.php>