

# 折り返し翻訳を用いた文章推敲手法の提案

鈴木 諒輔<sup>1,a)</sup> 角 康之<sup>1,b)</sup>

**概要:** 本稿では、正しく簡潔に情報伝達を達成する文章が書けるようになるべく、大学生が文章作成を学ぶための手法を提案する。提案手法は機械翻訳システムによる折り返し翻訳を用いる。折り返し翻訳は主語と述語が明確で、短く曖昧な表現がない文章を作成する方法として有用であり、文章推敲の手段の一つとして重要であると考えられる。提案システムは、学習者による入力文に対して折り返し翻訳結果を提示する。提案システムによる文章推敲手法が効果的であるか、文章間の類似度を提示した場合としない場合で文章推敲の回数や内容に変化があるかについて検討するために実験を行った。実験結果から、類似度の提示によって文章の編集回数が増加傾向であることや文章推敲への有用性が示唆された。

## 1. はじめに

本研究は、正しく簡潔に情報伝達を達成する文章が書けるようになるべく、大学生が文章作成を学ぶための手法を提案するものである。提案手法は機械翻訳システムによる折り返し翻訳を用いる。提案システムは、学習者による入力文に対して折り返し翻訳結果を提示する。折り返し翻訳とは原言語を対象言語に翻訳し、その翻訳結果を再び原言語に翻訳することを指す。例えば、原言語で書かれた文章の折り返し翻訳の結果を参照することで、その文章がどの程度良い文章であるかの指標になると考えられる。なぜなら、書かれた文章が機械翻訳にとって折り返し翻訳された結果の良し悪しから、自分の書いた文章の評価を客観的に知ることができるからである。よって、折り返し翻訳文が自分の意図した通りの文章として提示されたならば、原言語で書かれた文章は良い文章であることが予想される。

折り返し翻訳結果が意味の通じない文章であったり入力文とはかけ離れた内容であった場合、それはそもそも学習者が入力した文章に問題があると考えられる。提案手法のねらいは、そのような考えから学習者自らが自分の文章を再考し、他者に伝わりやすい文章作成を試行錯誤を促すことである。本研究に通底する仮定は以下の通りである。他人に正しく情報伝達できる文章は、機械翻訳システムにも正しく処理できるものであると考えられる。言い換えると、機械翻訳システムに正しく処理できない文章は人間も正しく理解できる保証は無いと考えられる。現在の機械翻訳システムの能力は、この仮定を支援するに足るものにな

りつつあると考える。なお本研究では、原言語を日本語、対象言語を英語に設定している。折り返し翻訳の具体的な流れについては図1に示す。

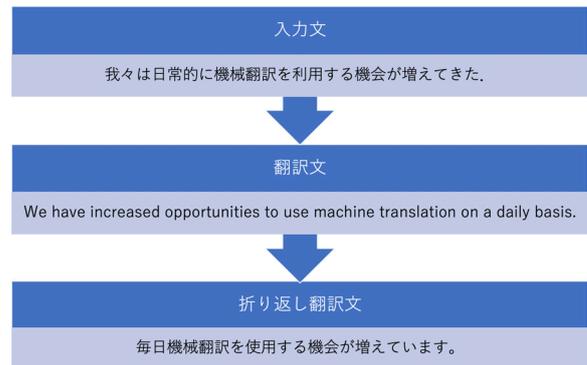


図1 折り返し翻訳の流れ

このような背景から機械翻訳を活用することで、英語論文を読んだり自分が書いた英文の意味が正しいか確かめることができる。また、折り返し翻訳を行うことで自分が理解できる言語で文章が妥当であるかを知ることができるようになったと考えられる。また一般的には、折り返し翻訳によって妥当性が高い文章は主語と述語が明確であり、短く曖昧表現がない文章が良いとされる。これは、論文やニュース記事、技術文書のような曖昧性を排除すべき情報伝達型テキストに求められる文章の要素でもある。論文や技術文書は多種多様な読者を想定して書かれるべきものである。このような文章は文章推敲が特に重要だとされる。なぜなら、これらの文章は客観的でわかりやすいものでなければならないからである。このことから、折り返し翻訳が情報伝達型テキストなどの文章推敲に寄与すると考えら

<sup>1</sup> 公立はこだて未来大学

<sup>a)</sup> r-suzuki@sumilab.org

<sup>b)</sup> sumi@acm.org

れる。

本研究では、折り返し翻訳が文章推敲手法として有用か検討するとともに、その手法を提案する。本研究の着眼点は、折り返し翻訳という機械翻訳の簡単な操作を通じて、文章構成や単語などの推敲前の文章における問題点を明らかにできるという点である。また提案手法の狙いとして、推敲者が折り返し翻訳文の提示によって推敲者が書いた文章を再考するきっかけを与え、他者に伝わりやすい文章を作成するために必要な試行錯誤を促すことがある。さらに本研究における仮定として、他人にも正しく情報伝達できる文章は機械翻訳を通じても正しく処理できるのではないかという点がある。本研究では、このような仮定を支援できると考える。本稿では、折り返し翻訳文の提示が文章推敲に有用であるか、推敲文と折り返し翻訳文との類似度を提示したときの推敲者へ与える影響や効果について検討する。

## 2. 関連研究

不適切な翻訳箇所が少ない文章を作成するために入力文を修正する作業を繰り返すことは翻訳リペアと呼ばれるが、宮部らは翻訳リペアの結果がどの程度の精度を持つかについて評価を行った [1]。この研究では、日本語から対象言語とされる英語・中国語・韓国語へ意味伝達が可能なレベルでの翻訳を目的として、折り返し翻訳を用いて対象言語への翻訳に適した日本語文章への修正を行い、その翻訳精度と修正コストについて評価した。本研究では、日本語の文章推敲支援を目的として、折り返し翻訳を活用したときの支援効果について検討・評価する。

林田らは機械翻訳システムの性能を測るインタラクティブティという新たな評価尺度を提案している [2]。また、折り返し翻訳結果と対象言語である英語の翻訳結果を同時に提示することによって、翻訳リペアの有効性が高まることも明らかにした。特に英語能力が十分でない場合に有用であることが述べられている。このような点を考慮して本研究では、文章推敲支援を目的とした折り返し翻訳文の提示を行うが、同時に順翻訳である英語訳された文章も提示している。

横林らは文章の書き換え候補の生成と文章推敲支援に役立つ情報提示を行うために、係り受けの複雑な文を抽出することができる係り受け解析モデルを作成した [3]。提案手法では構文情報から係り受けを推定することで係り受けの複雑な文を抽出し、係り受けの複雑さの指標が高い文章に対して、書き換え候補を生成し書き手に提示している。この手法は係り受けの判断を書き手に提示することで、書き手に文章推敲を促している。本研究では、推敲したい文章の折り返し翻訳文を書き手に提示することで、書き手に文章推敲を促す。

折り返し翻訳の信頼性に関する研究として、Grunwald

らは医療現場におけるインフォームドコンセントフォーム (ICF) の品質を担保するために折り返し翻訳を活用している事例を挙げている [4]。通常、医療現場における ICF が被験者または代表者が理解できない言語で書かれている場合には、被験者または代表者が理解できる言語に翻訳しなければならないとされている。ICF は高度な技術文書であるため、認定された医療翻訳者による翻訳が不可欠とされている。しかし、認定された医療翻訳者であっても完璧な翻訳を行うことは難しいとされている。そこで、この事例では折り返し翻訳を行うことで翻訳文の正確性を担保することを試みている。結果として、文章の正確性向上に寄与するがコストがかかる点が指摘されている。また、翻訳文中の単語の置き換えが不適切になる危険性や言語間の文法構造の違いに影響されることで折り返し翻訳文の可読性が低下する点、特定の翻訳において予測不可能な誤りが発生する点を指摘している。本研究では、折り返し翻訳を行うことで、推敲したい文章の正確性を測ることができると考えている。これにより推敲したい文章とその折り返し翻訳文を比較することが文章推敲支援につながると考えている。

内藤らはライティング能力を向上させることを目的として、Microsoft Word のマーカ機能とコメント機能を組み合わせたレポートのフィードバックを行い、学生に文章を修正をさせた [5]。マーカ機能を活用した支援では、添削項目を設定し、それぞれの色を決めて修正に必要な箇所にマーカを引いた。その結果、マーカ機能を活用した支援では、引用表現の修正・ねじれ文・語彙・書き言葉・文法の修正の動機づけに有効であることがわかった。本研究では、マーカなどの視覚的な表示は行わないが、折り返し翻訳文の提示によって修正が必要である部分を可視化し、文章推敲の大きな動機づけになると考えている。

Flower らは文章を書くときにおける書き手の心的な活動を文章産出過程モデルとしてモデル化した [6]。具体的には、3つの過程に分けられており、「課題環境」、「書き手の記憶」、「文章産出過程」である。さらに、「文章産出過程」の下位過程として「プランニング」、「文章化」、「再考」、「モニター」が定義されている。また、推敲には書き手の意図と実際の文章の意味のズレを検出する能力と、そのズレを的確に認識する能力が重要であると述べている。本研究では、推敲者の意図と文章の意味のズレに対する認識を折り返し翻訳文の提示によって助けることができるものだと考えている。

宮部らは折り返し翻訳文と対象言語翻訳文の精度不一致に関する要因を分析した [7]。不適切な語句の選択や多義語の不適切な使用、語句の区切り方や語順の失敗などが大きな影響を与えると述べている。本研究では、本システムを運用した過程で得た知見を、文章推敲時に推敲者が参照する折り返し翻訳結果などに影響を与えないようにするために注意すべき点としてまとめ、3.1 節で述べている。

### 3. 文章推敲システムの実装と推敲方法

本研究では、折り返し翻訳を用いた文章推敲手法を提案する。そのために、折り返し翻訳を用いて文章推敲を支援するシステムを実装した。折り返し翻訳を行うために機械翻訳エンジンが必要である。本研究では利用のしやすさや拡張性を考慮して、Google 翻訳を機械翻訳エンジンとして利用した。また、Google 翻訳をシステムに組み込むことが容易なことから Google スプレッドシート上に文章推敲システムを実装した。実装には Google Apps Script を利用し、Google 翻訳を翻訳エンジンとして英語への順翻訳、日本語への折り返し翻訳、以下で述べる文字列類似度の計算と出力を Google スプレッドシート上で自動的に行えるように実装した。Google スプレッドシート上に翻訳したい文章（日本語）を入力する列、対象言語とする英語を出力する列、翻訳したい文章の折り返し翻訳文を出力する列、そして推敲したい文章とその折り返し翻訳文の類似度を出力する列を用意した。また、本システムで提示される文章間の類似度は文字列類似度である。今回は文字列類似度として、ダメラウ・レーベンシュタイン距離を採用する。ダメラウ・レーベンシュタイン距離とは、2つの文字列間の編集距離を測定するために用いられる文字列計量である [8]。これは一方の文字列を他方の文字列に変換するのに必要な最小の操作回数で表される。ここで1回の「操作」とは1文字の挿入、削除、置換、隣接文字交換を指す。文章間の差異を確かめる指標となる。文章推敲システムのインターフェースは図2に示す。以下では、実際にシステムを利用して文章推敲を行っていくプロセスについて述べる。

入力文 (日本語)	出力文 (英語)	折り返し翻訳文 (日本語)	類似度
我々は機械翻訳技術の向上とともに、機械翻訳を利用する機会が増えました。	With the improvement of machine translation technology, the opportunity to use machine translation has increased.	機械翻訳技術の改善により、機械翻訳を使用する機会が増えました。	0.6
ビッグデータを利用したデータサイエンスが現実のものとなるにつれて、これまで増われてきた人工知能技術の実用化が花開きつつある。	As the data science using big data becomes a reality, the practical application of the artificial intelligence technology cultivated up to now is blossoming.	ビッグデータを用いたデータサイエンスが現実のものになると、これまで増われてきた人工知能技術の実用化が花開いています。	0.75
本研究では、人に伝わる曖昧性の無い文章を書くことを支援する文章推敲システムを構築する。	In this research, we construct a text revision system that supports writing unambiguous text that is transmitted to humans.	この研究では、人間に送信される明確なテキストの作成をサポートするテキスト修正システムを構築します。	0.32

図2 文章推敲システムのインターフェース

#### 3.1 推敲したい文章の入力

まず最初に、推敲したい文章を定められたスプレッドシートのセルに入力する。図2を参照すると、2行A列のセルから順次入力していく形である。Google スプレッドシート上で文章推敲を行うので、下のセルに逐次推敲したい文章を入力しながら繰り返し文章推敲を行うことや文章を変更した箇所の差分を確認しながら文章推敲が行うことができる。次に、折り返し翻訳を用いた文章推敲時に気を

つけるべき点を以下に示す。これらの点は、筆者が本システムの運用を通じて得た知見である。下記のこととは情報伝達型テキストを書く際に必須の事柄であると考えられるとともに、この点を意識することで推敲作業が楽になると思われる。

- 長文になればなるほど翻訳精度が落ちるため、できる限り短文にすること
- 複文（述語が二つ以上ある文）や重文（単文を並列させた文）は避けること
- 余計な副詞や修飾語は避けること
- 理解しやすい、論理的な文章を書くこと
- 主語、述語、目的語が明確な文章にすること
- できるだけ漢字を用いること
- できるだけ意味を誤解されるような曖昧表現を避けること

#### 3.2 入力文と折り返し翻訳文の比較

次に推敲したい文章を入力すると、Google 翻訳を介して対象言語である英語の翻訳文と入力文の折り返し翻訳文が出力される。また、文字列類似度としてダメラウ・レーベンシュタイン距離が計算されて提示される。なお、提示される類似度はダメラウ・レーベンシュタイン距離を文字列長で標準化したものを用いており、0に近いほど類似度が低く、1に近いほど類似度が高いことを示す。図2を参照すると、2行B列から推敲したい文章の翻訳文、2行C列から折り返し翻訳文、2行D列から標準化された文字列類似度が出力される仕組みとなっている。推敲時に着目すべきこととして、推敲したい文章とその折り返し翻訳文の間に意味の相違や許容できない構文の変容、単語の置き換えなどが挙げられる。また、英語の翻訳文も確認することで、より文章推敲を促すことも期待される [2]。その他にも、入力文と折り返し翻訳文の文字列類似度を参照することで、文章間の差異を数値的に意識できると考えられる。文章を推敲する過程において、類似度を中心に考えて文章を修正する場合と折り返し翻訳文だけでなく英語の翻訳文なども参考にして文章の構造的欠陥などを見出そうとする場合には、文章推敲の効果に差が出るとも考えられる。例としてシステムを使用して一つの文章を入力すると、システムの出力例として図3のような結果が得られる。

#### 3.3 文章推敲の反復

図4は本システムを用いたときの文章推敲プロセスを表している。基本的にはプロセスに従って文章推敲を行う。まずはじめに3.1節で触れた通り、推敲したい文章を指定されたセルに入力する。入力後には翻訳文、折り返し翻訳文、類似度が自動的に出力される。その次は3.2節で触れた通り、英語の翻訳文や類似度も参考にしながら入力文と折り返し翻訳文の比較を行う。入力文と折り返し翻訳文の

間に大きな差異がなく問題点を書き手が感じなければ、入力文に対しての文章推敲を終了する。一方で問題があると感じた場合は、入力文の修正を行う。次の行にセルを移して推敲した文章を入力し同様に確認作業を行い、納得できるまで文章推敲を反復する。また、スプレッドシート上で共同編集が可能なので、複数人で編集アドバイスやコメントを送り合うことで、相互指導を行うことも可能である。文章推敲の例として、図5を示しておく。

入力文 (日本語)	翻訳文 (英語)	折り返し翻訳文 (日本語)	類似度 (0~1)
日常的に機械翻訳を利用する機会が増えてきた。	Opportunities to use machine translation on a daily basis have increased.	機械翻訳を毎日使用する機会が増えています。	0.5

図3 システムの出力例

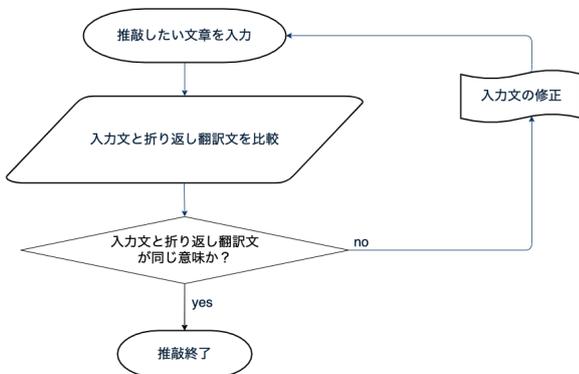


図4 システムを用いた文章推敲のプロセス

入力文 (日本語)	翻訳文 (英語)	折り返し翻訳文 (日本語)	類似度 (0~1)
本研究では、機械翻訳を用いて逆翻訳を行うことを通じて、どのくらい文章推敲に寄与するか議論する。	In this study, we discuss how much it contributes to sentence revision through reverse translation using machine translation.	この研究では、機械翻訳を使用した逆翻訳によって、文の修正にどれだけ貢献するかについて説明します。	0.31
本研究では、折り返し翻訳を活用することで文章推敲を支援するシステムを開発し、評価する。	In this study, we develop and evaluate a system that supports sentence review by using back translation.	この研究では、逆翻訳を使用して文のレビューをサポートするシステムを開発および評価します。	0.36

図5 文章推敲の例

#### 4. 提案手法による文章推敲実験

折り返し翻訳文を参照しながら文章推敲を行う場合、推敲された文章の改善効果や、類似度の有無による文章推敲の促進効果を検討するために、折り返し翻訳を用いた文書推敲の実験を実施した。実験の被験者は、公立はこだて未来大学および大学院の学生6人であった。被験者にはノートパソコンなどの手元でGoogleスプレッドシートを開い

て文章推敲作業ができる環境を用意してもらった。実験前には、折り返し翻訳を用いた文章推敲を行うためのGoogleスプレッドシートの使い方について説明した。各被験者には、折り返し翻訳を利用した文章推敲システムについての文章を推敲するタスクを与え、取り組んでもらった。各被験者の作業量の差や推敲過程の違いを観察する目的で、全員一律の文章を与えた。その際、類似度を提示する群と提示しない群の2群に分けて実験を行った。今回の実験は被験者が文章推敲し終え次第終了という形をとったため、実験に要した時間は各被験者1時間から1時間半程度であった。また文章推敲終了後には、簡単なヒアリングを行なった。文章推敲作業に使用したシステムの画面例を図2に示す。

実験の手順を以下に示す。

- (1) 推敲対象である文章をチェックする。
- (2) A列のセルに推敲したい文章(問題があると感じた文章)を入力する。
- (3) 文章入力後、翻訳文と折り返し翻訳文、対象の被験者には類似度も出力されるので、それらの情報を元に文章推敲を行う。
- (4) 入力文と折り返し翻訳文のが同じ意味になったと判断するなど被験者が納得のいく文章になった場合、その文章に対しての文章推敲を完了とする。
- (5) 1~4の手順を繰り返し、全ての文章に対して推敲を完了させる。

原文から推敲したい文章とその折り返し翻訳文の意味が同じになっているなど修正の必要がないと被験者が判断した場合には、そのまま完了して良いものとした。また、全体の文章構成に疑問があれば適宜変更して良いものとした。実験では、文章の入力や修正などの文章推敲作業記録について、Googleスプレッドシートの編集履歴を利用した。

#### 5. 結果と考察

システムを利用して6名の被験者に文章推敲を行ってもらった。実験終了後のヒアリングでは、実際に文章推敲後を振り返って推敲が簡単だと感じた箇所と逆に難しいと感じた箇所を答えてもらった。このシステムを利用した上で感じたことや不満点なども聞いた。また、システム利用時に類似度が出力された実験群の被験者には、どの程度類似度を意識したか、参考になったかについても答えてもらった。

##### 5.1 推敲過程と内容に対する考察

文章の推敲過程に着目すると各被験者共に、文章推敲が難しかったと感じた箇所については文章の編集回数が増えている。具体的には、接続詞や句点の使い方や単語の置き換え、構文の変更などによって編集回数が増加しているのがわかった。被験者に推敲してもらった文章の中には、一文

が長すぎるものや主従関係がはっきりしないもの、複文が含まれているものがあつた。各被験者はシステムを利用して、そのような文章を構成する上で問題となる要素を修正できていた。

しかしながら推敲内容に着目すると、4人の被験者が冗長表現を修正することができなかつた。例を挙げると、「折り返し翻訳によって、入力文と折り返し翻訳文から、曖昧性を排すべき文章の精度を確認することができる」と考えられる。」のような文章が修正されなかつた。この文には、「することができる」という冗長表現が含まれており、「できる」に修正した方がより良いことはわかる。このとき、被験者の入力文のセルには冗長表現が含まれている文章が入力されているが、その折り返し翻訳文には冗長表現が修正された文章が出力されていた。しかし、被験者は入力した冗長表現が含まれている文章を最終的に採用した。このことから、折り返し翻訳文の表現の内容まで踏み込み、文章改善の手立てを見出すことの重要性が垣間見られる。多くの被験者は自身が入力した文章の良し悪しについて、折り返し翻訳との比較などを通じて確認するのみである傾向があるので、そのような視点を事前に教示する必要性を感じた。

## 5.2 実験群別の推敲回数に対する考察

文章推敲時に類似度の提示がされた被験者 A と被験者 B は類似度を参考にした・気になったと答えたが、被験者 C は類似度を参考にはしなかつた・気にならなかつたと答えた。文章推敲システムを利用して文章の推敲が完了するまでの文章の編集回数をまとめた表を表 1 として示す。ここでの文章の編集回数とは、単語を書き換えるなどの文章修正を指す。しかし、一つのセルで複数回編集しセルを更新しなかつた（編集結果を折り返し翻訳文として出力させなかつた）場合は、それらを一回の編集としてカウントする。表 1 を見ると、類似度なしよりも類似度ありの場合の方が編集回数が多い傾向にあることがわかつた。類似度ありの実験群における被験者 C は 35 回と同一実験群の中では編集回数は少ないが、ヒアリングでは類似度を参考にはしなかつたと答えていた。このことから、類似度などの数値的な指標を提示した方が推敲回数が増えることが示唆された。基本的には、文章推敲を繰り返すほど文章は良くなっていく。さらに、折り返し翻訳を通じて文章推敲を行うことで、より具体的な修正点に意識を持って文章推敲に取り組めると考えられる。例えば、単語の意味が適切であるかどうかや文章の構造がおかしくなっていないか、意味がズレていないかという点に気づきやすくなる。このような、ゲーミフィケーション要素を文章推敲に取り入れることによって、文章推敲へのモチベーションや意欲を上昇させることに役立つと考えられる。しかしながら、提示する類似度の改善も必要と考えられる。

表 1 実験群別の文書推敲回数

	類似度あり	類似度なし
被験者 A	103 回	—
被験者 B	78 回	—
被験者 C	35 回	—
被験者 D	—	63 回
被験者 E	—	42 回
被験者 F	—	22 回

## 5.3 実験協力者からのヒアリング

ヒアリングにおいて多数のコメントをもらった。特に重要なコメントを 2 つ挙げる。1 つ目は、このシステムではあくまでの一つ一つの文章の良し悪しを確かめるには有用であるが、文章全体の良し悪しを確かめることができないというコメントであつた。特に、接続詞や代名詞のような前後関係が重要となる部分に対してフォローができないといった意見があつた。2 つ目は、内容のつながりが考慮されないというコメントであつた。文章のつながりなどは推敲者が考慮しなければならず負担になっている。このシステムの欠点として、一つの文章についてしか折り返し翻訳によって文章の良し悪しを確認することができない点がある。文章間の内容の繋がりや文章全体のまとまりは推敲者が考慮しなければならない。推敲者の不満や疑問も浮き彫りになつた。本システムはあくまでも一つの文章に対して、その文章の確かさを確認するためには有用であるが、文章全体の良し悪しの確認を保証するものではないことを提示する必要がある。

## 6. おわりに

本稿では、正しく簡潔に情報伝達を達成する文章が書けるようになるべく、大学生が文章作成を学ぶための手法を提案した。また、システムを利用して折り返し翻訳を利用した文章推敲手法が効果的であるか、文章間の類似度を提示した場合としない場合で文章推敲の回数や内容に変化があるかについて検討するために実験を行った。実験結果から類似度の提示によって文章の編集回数が増加傾向であることがわかつた。また、各被験者はシステムを利用することで 5.1 節で述べたような問題のある文章を修正できたことから文章推敲への有用性が示唆された。しかし、冗長表現を修正することができなかつた事例から、被験者への折り返し翻訳文に対する着目点を教示する重要性を確認した。本研究によって、折り返し翻訳という簡便な方法によって文章推敲の支援に資することができる。今後の課題として、より良い文章間の類似度指標を実装すること、その類似度を提示することによる効果と差分を確認すること、より実用的なシステムへの改変などを行う必要があると考えられる。

## 参考文献

- [1] 宮部真衣, 吉野孝, 重信智宏. 折返し翻訳を用いた翻訳リベアの効果. 電子情報通信学会論文誌 D, Vol. 90, No. 12, pp. 3141–3150, 2007.
- [2] 林田尚子, 石田亨. 翻訳エージェントによる自己主導型リベア支援の性能予測. 電子情報通信学会論文誌 D, Vol. 88, No. 9, pp. 1459–1466, 2005.
- [3] 横林博, 菅沼明, 谷口倫一郎ほか. 係り受けの複雑さの指標に基づく文の書き換え候補の生成と推敲支援への応用. 情報処理学会論文誌, Vol. 45, No. 5, pp. 1451–1459, 2004.
- [4] David Grunwald and Norman Goldfarb. Back translation for quality control of informed consent forms. Vol. 2, , 03 2006.
- [5] 内藤真理子, 小森万里. どんな手助けがあればレポートの自己修正ができるのか. 専門日本語教育研究, Vol. 15, pp. 41–46, 2013.
- [6] Linda Flower and John R Hayes. A cognitive process theory of writing. *College composition and communication*, Vol. 32, No. 4, pp. 365–387, 1981.
- [7] 宮部真衣, 吉野孝. 折返し翻訳文と対象言語翻訳文の精度不一致要因. 電子情報通信学会論文誌 D, Vol. 95, No. 1, pp. 11–18, 2012.
- [8] Fred J Damerau. A technique for computer detection and correction of spelling errors. *Communications of the ACM*, Vol. 7, No. 3, pp. 171–176, 1964.