

# 構文解析と類義語情報提示に基づく翻訳リペア支援システム

曾 暉<sup>†</sup> 井上 潮<sup>‡</sup>

<sup>† ‡</sup> 東京電機大学大学院 工学研究科 〒101-8457 東京都千代田区神田錦町 2-2

E-mail: <sup>†</sup>10kmc18@ms.dendai.ac.jp <sup>‡</sup>inoue@c.dendai.ac.jp

あらまし 機械翻訳を介したコミュニケーションにおいて、翻訳精度を向上するために、折り返し翻訳を用いた翻訳リペアが注目されている。これは、母国語の知識しか持っていないユーザが機械翻訳に適する文章を作成する方法として、重要な役割を担っている。しかし、文章を繰り返しリペアする作業がユーザにとって大きな負担となる。本論文では、ユーザの負担を軽減するために、「日本語文の成分に関するルール」に基づいた翻訳不適箇所の検出とリペア情報の提供を行う翻訳リペア支援システムを提案する。またそれを用いて翻訳リペア実験を行い、その有用性について評価を行った結果を述べる。

キーワード 折り返し翻訳, 構文解析, 翻訳リペア支援

## Translation repair support system based on syntactic analysis and synonym information presentation

Hui ZENG<sup>†</sup> and Ushio INOUE<sup>‡</sup>

<sup>†‡</sup> Tokyo Denki University 2-2 Kanda-Nishiki-cho, Chiyoda-ku, Tokyo, 101-8457 Japan

E-mail: <sup>†</sup>10kmc18@ms.dendai.ac.jp, <sup>‡</sup>inoue@c.dendai.ac.jp

**Keyword** back-translation, syntactic analysis, translation repair support

### 1. はじめに

近年、インターネットの普及により、ネットワークを介した異文化間コミュニケーションの機会が増加している。一般に異文化の人たちとコミュニケーションを行うには、相手国の言語また共通言語を学習する。しかし、多くの人にとって、外国語を学習することは容易ではない。そのため、機械翻訳技術を利用した取り組みが行われている。しかし、現在の翻訳精度はまだ不十分であり、異文化間コミュニケーションによる相互理解が困難な状況にある。

このような機械翻訳精度の問題を克服するための手段として、翻訳リペア技術が注目されている。翻訳リペアとは、よい翻訳結果を得られるまで、ユーザが原文を修正する作業を繰り返すものである。例えば、野村らの異文化コラボレーションの実験[1]では、日中韓馬英の5言語での多言語コミュニケーションが行われた。その中、ユーザがメッセージを投稿する際に、英語の翻訳結果を参照し、原文の修正が繰り返された[2]。しかし、英語を習得していないユーザは、翻訳リペアを行うことができない。この問題に対し、母国語の知識だけで、翻訳結果を確認し、翻訳リペアを行える方法として、折り返し翻訳が研究されている[3,4,5]。折り返し翻訳とは、ある言語から目的

言語に翻訳し、さらに元の言語に翻訳することにより、翻訳精度を向上させる方法である。

翻訳リペアにおいて、どのように修正するかを考え出すこと、また何回も文章の修正を繰り返すことは、ユーザにとって負担の大きい作業である。そのため、修正作業の支援が必要である。そこで、我々は、折り返し翻訳を用いた翻訳リペア作業において、ユーザの負担を軽減する方法の開発を目標とした研究を行っている。

本論文では、最初に有効な支援方法を検討するために行った予備実験について述べる。続いて、我々の新たな支援方法を提案し、構築した翻訳リペア支援システムを紹介する。最後に、構築したシステムを用いて翻訳リペア実験を実施し、その有効性について検証した結果を報告する。

### 2. 関連研究

宮部らは原文にある単語が折り返し翻訳結果文の中に出現していない場合、その単語を翻訳不適箇所として指摘する手法を提案した[6]。翻訳不適箇所の指摘は、ユーザが修正箇所を決めるのに役立つことが示されている。しかし、翻訳不適箇所を指摘するだけで、ユーザがその箇所をどのように修正すればよい

かの支援がないため、翻訳リペアの負担を大きく軽減することができない。

山下らは日本語文法体系などに従って、総計 70 のルールを整理した[7]。これを「よい翻訳結果を得るためのルール集合」と呼ぶ。また、このルール集合を教示することによって、実際にユーザのリペア作業の削減に役立つかどうかを検証するための実験が行われた。ルール集合の教示とは、実験前にユーザが 70 個のルールを 5 分間で目を通すというものである。これにより、リペア作業が大幅に削減できることが確認された。しかし、実際の翻訳リペア作業において、どの箇所が翻訳不適なのか、また翻訳不適箇所をどのルールを用いてどのように修正すればよいのかをリアルタイムにユーザに呈示することはできない。

### 3. 翻訳リペアの予備実験

#### 3.1. 実験概要

我々は、実際の翻訳リペア作業によって、どのくらい精度を向上できるか、またそのためにどのくらいコストを要するか、さらにコストを削減するためにどのようなリペア支援方法が有効なのかを検証するために、折り返し翻訳機能だけの翻訳ツールを用いて、3 人の被験者による日英間の翻訳リペア実験を行った。

実験で使用したツールは、言語グリッド<sup>1</sup>が提供している折り返し翻訳サービスを用いて構築したものである。被験者は、学部生 2 名、大学院生 1 名の合計 3 名である。日本語文は、機械翻訳機能試験文<sup>2</sup>のうち、15 文字以上 45 文字以下の文を 90 文選択したものである。試験文の一部を表 1 に示す。

実験の手順は、以下の通りである。なお、翻訳リペアの終了判断は、ユーザに任せる。

- ① 日本語文を入力エリアに入力し、翻訳する。
- ② 折り返し翻訳結果を見ながら、原文と同じ意味になるまで原文を繰り返し修正する。
- ③ 折り返し翻訳結果が入力文と同じ意味になったと判断したら、翻訳リペア作業を終了する。

#### 3.2. 実験結果

##### 3.2.1 主観評価による翻訳精度

最初に、被験者による翻訳精度の主観評価を行った。

評価方法は、翻訳リペア前の試験文とその折り返し翻訳結果、また試験文と翻訳リペア後の折り返し翻訳結果がそれぞれ、同

じ意味になっているかどうか比較した。

評価基準は Walker らの適合性評価(5 段階評価)[8]を利用した。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 (意味が何となくつかめる) 2:Little (雰囲気は残っているが、もとの意味が分からない) 1:None (まったく違う意味) である。

主観評価の平均値を表 2 に示す。リペア前は 2.92 であったものが、リペア後は 4.15 になった。従って、翻訳リペア作業により翻訳精度を向上させることができたといえる。

##### 3.2.2 翻訳リペアコスト

今回の実験における翻訳リペアコストを表 3 に示す。翻訳リペアコストは、一つの文の平均修正回数と平均修正時間とした。なお、被験者がリペアを放棄した文は除外した。

一つの文を修正するに当たって、被験者 3 人の平均修正回数は 5.2 回で、平均修正時間が 2 分 2 秒であった。3 人の被験者中、一番少ない修正回数は 3.9 回に対し、一番多い修正回数は 6.9 回であった。被験者のアンケートから、単語の翻訳不適に対応する類義語などの言い換えを考え出すことが容易ではないことが分かった。

##### 3.2.3 主なリペア方法

今回の実験で被験者が用いた主なリペア方法を表 4 に示す。またそれらが使用された文の数を表 5 に示す。これらの分類は、山下らの「よい翻訳結果を得るためのルール集合」に基づいたものである。

翻訳リペア実験において、被験者が最も多く行ったリペア方法は単語および表現の言い換えである。適切な言い換えを行えば、翻訳精度を向上できると考えられる。しかし、被験者の知識の差により、適切な言い換え表現を考え出すことが容易ではない。これを支援するために、単語の言い換え候補として、類義語辞書を用いた類義表現を提示することにより、ユーザの負担を軽減できると考えられる。

他に用いられたリペア方法は、主語の補完、文の分割、語順の変更、文の簡略化などである。実験のログデータから、ユーザがどのように修正すればよいか分らず、いろいろな方法で何回も試していたことがわかった。このような状況について、小倉らは、「文法的制約が弱く、文脈依存性が高い」ことから生じる翻訳の問題と指摘している[9]。例えば、主語が省略された場合、また文節間の係り受けが複雑で正しく認識されない場合である。

このような文法的な原因による翻訳不適を見つけ出すには、

<sup>1</sup> <http://langrid.org/jp/>

<sup>2</sup> <http://www.kecl.ntt.co.jp/>

文中の主語と述語の関係、修飾節と被修飾節の係受け関係などを明確にする必要がある。仮にユーザが素早く正確に文の構造を判断できれば、翻訳不適箇所を見つけやすくなり、翻訳リペア作業コストの軽減に繋がる。そのため、文章の係り受け解析を行い、各文節間の係り受け情報を提示することが有効と考えられる。

表 1 試験文の例

1.現地の人々の気持ちも踏まえて行動したいものだ.
2.この提案は、それをさらに進めて、最も合理的なものになった.
3.年齢別で見ますと若い層では歌謡番組がよく見られています.
4.君は細かい所まで注意を払わなければいけない.
5.各店の意見を聞いたが、いずれも後向きだった.

表 2 主観評価平均値

被験者	リペア前		リペア後	
	平均	標準偏差	平均	標準偏差
A	2.89	1.18	4.14	0.78
B	3.19	1.25	4.09	0.78
C	2.75	1.26	4.25	0.62
平均	2.92		4.15	

表 3 翻訳リペアコスト

被験者	修正回数(回)		修正時間(分:秒)	
	平均	標準偏差	平均	標準偏差
A	3.90	3.40	1:52	2:27
B	6.87	5.80	2:23	2:53
C	4.80	3.01	1:46	1:21
平均	5.19		2.02	

表 4 被験者による主なリペア方法の例

リペア方法	文の種類	リペア内容
(A)単語および表現の言い換え	原文	この提案は、それをさらに進めて、最も合理的なものになった。
	書換え文	この提案は、それをさらに進めて、最も適切なものになった。
(B)文簡略化	原文	私はひとりで手を揉みながら、三郎をも迎えた。
	書換え文	私は、三郎を迎えた。
(C)語順変更	原文	この映画は、地元民の協力を得て撮影された。
	書換え文	地元民の協力を得て、この映画は撮影された。
(D)文の分割	原文	彼はバスに乗って学校へ行くが、私は川に沿って歩いて学校へ行く。
	書換え文	彼はバスに乗って学校へ行く。しかし、私は川に沿って歩いて学校へ行く。
(E)主語補完	原文	あなたの予定を考慮して計画を作成しました。
	書き換え文	私は、あなたの予定を考慮して計画を作成しました。
(F)表現の補完	原文	子供は父に東京へ連れて行ってくれるよう頼んだ。
	書換え文	子供はその子の父に子供を東京へ連れて行ってくれるよう頼んだ。

表 5 被験者によるリペア方法の割合

リペア方法	文数(文)	割合(%)
A) 単語および表現の書き換え	46	60.5%
B) 文簡略化	10	13.2%
C) 語順変更	6	7.9%
D) 文の分割	2	2.6%
E) 主語補完	3	3.9%
F) 表現補完	4	5.3%
G) その他	5	6.6%
合計	76	

#### 4. 提案手法

本研究では、翻訳精度低下の原因となる主語の欠落と翻訳不適単語の検出を行う。また、文法的な翻訳不適を見つけやすくするため、文節間の係り受け情報を提示する。さらに、翻訳不適単語を修正しやすくするため、単語の類義語情報を提示する。

##### 4.1. システム構成

システムの構成を図 1 に示す。ユーザはクライアント部のユーザインターフェースを通じて本システムを利用する。なお、機械翻訳、類似度の計算、形態素解析、係り受け解析、類義語辞書などの処理ツールは、言語グリッドが提供している Web サービス API を介して利用する。

システムの実行画面を図 2 に示す。ユーザが領域 A の翻訳結果（翻訳不適指摘）を読んで、領域 B のリペア情報を参考し、領域 C の入力文をリペアする。

システム処理の流れを以下に示す。

- ① ユーザが文章を入力し、翻訳ボタンを押す。
- ② システムが翻訳システムを介して、入力文を翻訳する。
- ③ システムが入力文と折り返し翻訳文との類似度を計算する。
- ④ 類似度が一定値以上の場合、システムが「正しく翻訳できている」というメッセージと②の結果を画面に表示する。  
(以後の処理はしない)
- ⑤ 類似度が一定値未満の場合、まず、システムが入力文の主語が欠落しているかどうかのチェックを行う。その結果を画面に表示する。
- ⑥ 次に、システムが入力文の係り受け解析を行う。また、入力文の連文節を抽出し、画面に表示する。
- ⑦ システムが翻訳不適単語の検出およびその言い換え候補として、類義語情報を検索して、画面に表示する。
- ⑧ ユーザが画面に表示された情報を参考し、入力文のリペアを行う。

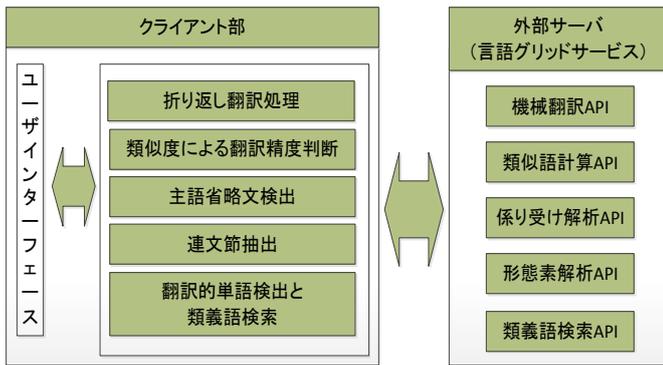


図 1 システム構成

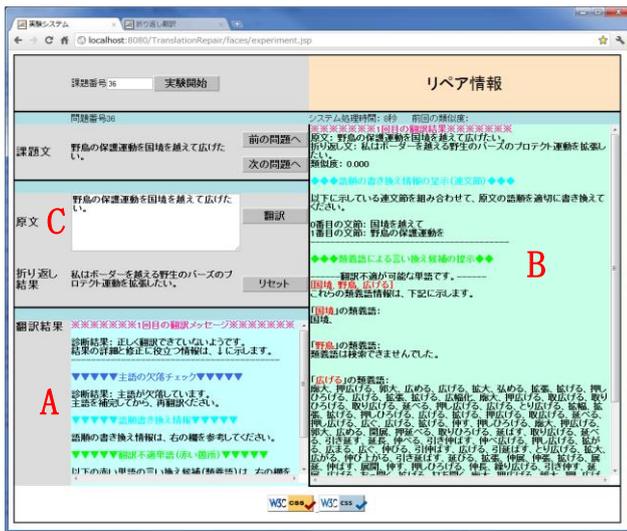


図 2 実行画面

## 4.2. 類似度閾値

本提案システムでは、翻訳精度を自動的判斷するための一つの基準として、入力文と折り返し翻訳文の類似度を用いる。類似度とは、二つの文が似ている度合いを表している。類似度が高ければ、翻訳精度が高いと判断できるが、類似度が低ければ、必ずしも翻訳精度が低いとは限らない。そのため、類似度閾値を設定する。

そこで、本研究では、翻訳リペアの予備実験で、被験者によりリペア不要と判断された文の類似度の平均値である 0.5 を類似度閾値とした。

## 4.3. 機械翻訳不適箇所の検出手法および情報提示

### 4.3.1 主語省略文判定

日本語では、「これは」や「私が」のように、名詞の後に助詞「は」と「が」があれば主語となることが多い。そこで本研究では、名詞と助詞「は」や「が」の句を主語とする。

主語の判定は、以下のような手順で行う。

- ① まず、句点を持つ文節を述語とする。
- ② 述語文節にかかる文節を主語候補として、抽出する。

- ③ 主語候補から主語である文節に限定するために、格助詞「は」、「が」を主語候補の中から見つけ出す。
- ④ 格助詞「は」、「が」を持つ主語候補の中から、述語に係る文節を主語として判定する。
- ⑤ 主語候補の中に格助詞「は」、「が」を持つ文節が存在しなければ、主語が欠落していると判定する。
- ⑥ 主語候補の中に格助詞「は」、「が」を持つ文節が 1 個だけ存在する場合、主語を欠落していないと判定する。
- ⑦ 主語候補の中に格助詞「は」、「が」を持つ文節が 2 個以上存在する場合、主語が複数存在していると判定する。

手順⑤で主語の欠落を検出した場合、ユーザに「主語の補充」というメッセージを表示する。また手順⑦で主語を複数検出した場合に、それらの主語候補を表示し、ユーザに「主語を明記するように」というメッセージを表示する。

### 4.3.2 翻訳不適単語の検出およびその類義語検索

文の中で翻訳不適単語を検出し、その言い換え候補として、類義語情報をユーザ提示する。なお、今回は検出対象を名詞と動詞とした。以下の手順で、翻訳不適単語を判定し、類義語情報を提示する。

- ① 形態素解析を行い、入力文と折り返し文のそれぞれ単語を抽出する。
  - ② 入力文中の単語が折り返し文中に存在していなければ、その単語を翻訳不適単語とする。
  - ③ その単語を赤字で強調表示する。
  - ④ その単語の類義語を WordNet で検索し、ユーザに提示する。
- なお、手順③で、すべての単語の類義語があるとは限らない。そのため、類義語がない場合、「類義語を検索できなかった」というメッセージを表示する。

### 4.3.3 連文節の抽出

ユーザが正確に文法を判断するために、文節の係り受け情報を提示する。係り受け構造が複雑な文の場合、どの文節とどの文節が関連しているかを判断するのが難しいため、修飾節単位で文を分割し、連文節としてまとめる。連文節は、二つ以上の連続した文節が一まとまりとなつて、一つの文節と同様の働きをするものである。以下の条件で、文頭の文節から検査していき、連文節を抽出する。例を用いて説明する。

**条件：** 文頭から一つずつの文節の係先が述語であるかを検査していき、係先が述語でなければ文節を連結する。係先が述語である文節が来たら、それまで連結したものは連文節となる。

例えば図 3 では、述語文節は「作成しました。」である。まず、「私は」の係先は「作成しました。」であるので、「私は」が単独の連文節となる。続いて、「あなたの」の係先は、「予定を」であり、「あなたの」と連結し、「あなたの予定を」となる。さらに、「予定を」の係先は、「考慮して」であり、「あなたの予定を」と連結し、「あなたの予定を考慮して」が連文節となる。「考慮して」の係先は、述語文節であるため、これまで連結した文節が一つの単独な連文節となる。「計画を」の係り先は述語文節であるので、一つの単語連文節となる。

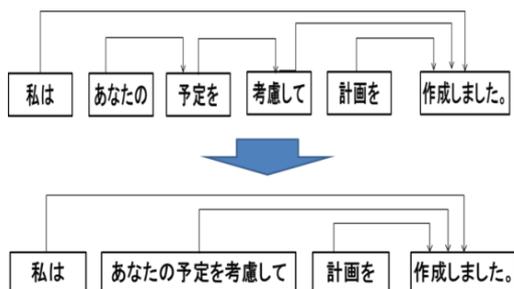


図 3 連文節の抽出

## 5. 評価実験

### 5.1. 実験概要

提案したシステムを用いることより、翻訳リペア作業におけるコストをどのくらい軽減できるかについて検証するために、予備実験でかかったコストとの比較実験を行った。

実験方法は、予備実験と同じである。ただし、本システムの有効性を検証するために、以下の二つの実験を行った。

**実験 1** 3人の被験者に同じ試験文を与える。

**実験 2** 2人の被験者ごとに異なる試験文を与える。

それぞれの実験では、1人ずつの被験者は翻訳リペア経験がある。実験で使用した試験文は、機械翻訳機能試験文の中から選択したものであり、二つの実験とも 30 文ずつ用いた。

なお、実験 1 で用いた試験文は、予備実験で用いた試験文のうちの 30 文である。また、実験 2 で用いた 60 文の試験文は、予備実験と実験 1 で用いた試験文とは異なるものである。

## 5.2. 実験結果

### 5.2.1 予備実験との翻訳精度の比較

評価方法は、予備実験と同じである (3.2.1 節)。実験 1 の翻訳精度を表 6 に、実験 2 の翻訳精度を表 7 に示す。実験 1 では、リペア前の平均主観評価値は 3.10、リペア後は 3.76 となった。また実験 2 では、平均主観評価値は 2.81、リペア後は 3.75 となった。

表 8 に示したように、翻訳精度が向上してはいるが、予備実

験の結果 (表 2) と比べると、向上率が少し低い。これは、被験者によって原文と折り返し翻訳結果の意味合いの判断基準が異なること、また実験で用いた試験文も異なることによるものと考えられる。

### 5.2.2 翻訳リペアコスト

実験 1 の翻訳リペアコストを表 9 に示す。実験 2 の翻訳リペアコストを表 10 に示す。実験 1 では、平均修正回数は 4.3 回、平均修正時間は 1 分 39 秒となった。実験 2 では、平均修正回数は 3.9 回、平均修正時間は 1 分 37 秒となった。

表 6 主観評価値 (実験 1)

被験者	リペア前		リペア後	
	平均	標準偏差	平均	標準偏差
A	3.23	1.02	3.73	1.00
B	3.00	1.02	3.67	1.22
C	3.07	1.26	3.89	1.30
平均	3.10		3.76	

表 7 主観評価値 (実験 2)

被験者	リペア前		リペア後	
	平均	標準偏差	平均	標準偏差
A	2.96	1.73	3.91	1.06
B	2.66	1.27	3.59	1.03
平均	2.81		3.75	

表 8 予備実験との翻訳精度の比較

	リペア前	リペア後	向上率 (%)
予備実験	2.92	4.15	42.1
実験1	3.10	3.76	21.3
実験2	2.81	3.75	33.5

表 9 実験 1 のリペアコスト

被験者	修正回数(回)		修正時間(分:秒)	
	平均	標準偏差	平均	標準偏差
A	3.43	1.96	1:41	1:44
B	4.04	2.20	1:25	1:10
C	5.32	3.37	1:50	1:28
平均	4.26		1:39	

表 10 実験 2 のリペアコスト

被験者	修正回数 (回)		修正時間 (分:秒)	
	平均	標準偏差	平均	標準偏差
A	3.78	2.15	01:26	01:15
B	4.03	1.81	01:49	01:07
平均	3.91		01:37	

### 5.3. 予備実験との翻訳リペアコストの比較

提案システムを用いた翻訳リペア作業におけるコストをどのくらい削減できるかを削減率で評価した。削減率は次の式で求めた。それぞれの結果を表 11 と表 12 に示す。

$$\text{削減率} = \frac{\text{予備実験のリペアコスト} - \text{評価実験のリペアコスト}}{\text{予備実験のリペアコスト}}$$

実験 1 では、翻訳リペアに要する平均修正回数を予備実験より 18.1%、平均修正時間を 18.9%減らせた。実験 2 では、予備実験より平均修正回数を 24.9%、平均修正時間を 20.5%減らせた。これにより、本提案システムの有効性を示している。しかし、今回の実験は、少ない被験者で実験を行ったため、限定的な結果である。今後は、多数の被験者による検証を行う必要であると考えられる。

表 11 翻訳リペアコスト削減率 (実験 1)

	修正回数 (回)	修正時間 (分:秒)
予備実験	5.19	02:02
実験 1	4.26	01:39
削減率	18.1%	18.9%

表 12 翻訳リペアコスト削減率 (実験 2)

	修正回数 (回)	修正時間 (分:秒)
予備実験	5.19	02:02
実験 2	3.91	01:37
削減率	24.9%	20.5%

## 6. まとめ

本稿では、翻訳リペア作業のコスト (修正時間と修正回数) を軽減するために、文法ルールに基づく翻訳リペア支援システムを提案した。また、提案システムの有効性を検証するために、3 人の被験者による翻訳リペア実験を実施した。その結果、ユーザへの支援を行うことにより、平均修正回数および平均修正時間を 20%程度減らすことができた。

今後は、類義語を用いた翻訳不適単語の検出、また単語の類義語からの有用な類義語の選別について検討する。また今回は、主語の欠落問題についてのみ検討したが、目的語や述語の欠落についても検討する必要がある。

## 謝辞

本研究の実験システムを構築するにあたって、必要な言語処理ツールや言語資源を提供していただいた言語グリッドプロジェクトのメンバに心より感謝いたします。

## 参考文献

- [1] 野村,船越,山下,安岡,石田 “機械翻訳を介したオープンソースソフトウェア開発: Intercultural Collaboration Experiment 2002”, 情報処理学会第 65 回全国大会, 2T6-3, 2003
- [2] 小倉健太郎, 林 良彦, 野村早恵子, 石田 享: 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).
- [3] 安岡美佳, 中小路久美代, 大平雅雄, 石田享, 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
- [4] 宮部真衣, 吉野孝: 折り返し翻訳における中間言語の精度評価, 情報処理学会, デジタルドキュメント研究会, Vol.2009-DD-71, No.4, pp.1-6 (2009).
- [5] 宮部真衣, 吉野 孝, 重信智宏, “機械翻訳を用いた翻訳リペアの効果”, 信学論(D), vol. J90-D, no. 12, pp.3141-3150, Dec. 2007
- [6] 宮部真衣, 吉野 孝: 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, 情処論, vol.50, NO.4, pp.1390-1398 (2009) .
- [7] 山下直美, 坂本知子, 野村早恵子, 石田享, 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
- [8] K. Walker, M. Bemba, 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.
- [9] 小倉健太郎, 林良彦, 野村早恵子, 石田享, “目的指向の異文化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-, ” 情報処理学会第 65 回全国大会, 2T6-4, 2003.