

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年6月6日現在

機関番号：12601

研究種目：研究活動スタート支援

研究期間：2011～2012

課題番号：23800014

研究課題名（和文） 利用場面に最適化した翻訳リペア支援に関する研究

研究課題名（英文） A Study on Translation Repair Method Optimized for Use Situation

研究代表者

宮部 真衣 (MIYABE MAI)

東京大学・知の構造化センター・特任研究員

研究者番号：00613499

研究成果の概要（和文）：

本研究は、信頼性のある多言語間コミュニケーションの実現を目指し、利用場面に最適化した翻訳リペア支援手法を構築することを目的としている。本研究により、以下の成果が得られた。(1)特に正確性が要求される分野として、医療分野における機械翻訳の利用可能性および医療分野において翻訳文が満たすべき必要要件を調査した。調査の結果、必要要件が満たされれば、機械翻訳文でも利用可能であることがわかった。(2)類語を用いた言い換え文自動生成手法をもとに、良い翻訳結果を得るためのルールに違反するかどうかを検証する仕組みを構築し、プロトタイプシステムを構築した。また、折り返し翻訳文と対象言語翻訳文の精度不一致を検出する手法として、複数の翻訳システムを用いる手法を構築した。構築した手法により、約71%の精度不一致を検出できることが確認できた。

研究成果の概要（英文）：

The purpose of this research is to develop translation repair method optimized for use situation, and improve the reliability of multilingual communication via machine translation. The contribution of this research is as follows: (1) this research revealed that machine-translated sentences are available in the medical field if they have information of their original sentences without exaggeration and without omission. (2) We developed a prototype system using detection methods of breaking the rules for accuracy translation. Moreover, we proposed a method for detecting the mismatch between a target-translated sentence and its back-translated sentence using multiple translation systems. This method can detect approximately 71% of the mismatches.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2011年度	1,300,000	390,000	1,690,000
2012年度	1,200,000	360,000	1,560,000
年度			
年度			
年度			
総計	2,500,000	750,000	3,250,000

研究分野：工学

科研費の分科・細目：メディア情報学・データベース

キーワード：多言語間コミュニケーション, 機械翻訳, 多言語処理

1. 研究開始当初の背景

インターネットが世界的に普及し、2010年のインターネット利用者数は20億人を突破した。インターネット上で最も用いられている言語は英語である。一方、2010年のインターネットにおける使用言語の割合は、英語が27.3%、英語以外が72.7%となっており、英語以外の言語の使用率が年々増加している。また、日本における外国人登録者数は約218万人に上り、日本国内における多言語対応の必要性も高まっている。特に医療のような場面ではコミュニケーションが文字どおり生死を分けることがあり、日常生活にとどまらず、様々な場面での多言語対応が喫緊の課題となっている。このような背景から、多言語間コミュニケーションを支援するために、機械翻訳などの様々な言語資源を用いた取り組みが行われてきている。

母語の文章を自動的に他言語へと翻訳可能な機械翻訳システムは、他言語を十分に習得していない人々による多言語間コミュニケーションを可能にする。一方、機械翻訳システムによる完全な翻訳の生成は容易ではなく、これまでに開発された、機械翻訳を用いた多言語間コミュニケーション支援システムにおいては、機械翻訳システムの翻訳精度によるコミュニケーションへの悪影響が報告されている。他言語話者間で正しく意思伝達を行うためには、翻訳精度の確保が不可欠である。

2. 研究の目的

従来、機械翻訳システムの翻訳精度を向上するためには、システム提供者側での対応が不可欠である。一方、完璧な翻訳が可能な機械翻訳システムの実現は困難である。これまでに、機械翻訳システムの利用者が入力内容を工夫することにより、翻訳精度を向上させる手法に関する研究を行ってきた。折り返し翻訳（対象言語への翻訳結果を、再度原言語へと翻訳しなおすこと。流れの例を図1に示す）を用いた入力文の書き換え（翻訳リペア）により、現状の完璧な翻訳が不可能な機械翻訳システムでも、ユーザが入力文を書き換えること（翻訳リペア）により、翻訳精度の向上が可能であることを示した。しかし、翻訳リペアでは、ユーザ自身が修正内容を考えなければならない。そのため、利用者はどのように入力文を書き換えればよいか分からず、適切に翻訳リペアができないという問題がある。また、翻訳文を利用する場面によって、要求される精度が異なり、それらを考慮する必要がある。

そこで、信頼性のある多言語間コミュニケーションの実現を目的とし、利用場面に最適化した翻訳リペア支援手法について研究を

入力文

それじゃあ、よろしくお願いします。

対象言語翻訳文

那么，谢谢你。

折り返し翻訳文

さて、あなたに感謝します。

原言語から対象言語への翻訳

対象言語から原言語への翻訳

図1 折り返し翻訳の流れ

行う。本研究で扱う課題は、(課題1)利用場面に合わせた支援手法に関する研究、および(課題2)利用者による翻訳リペア作業の負担軽減に関する研究、の2点である。

課題1については、特に高精度な翻訳が要求される場面において、翻訳リペアを用いることで機械翻訳の適用が可能かどうかを調査する。

課題2については、翻訳リペア作業におけるユーザの負荷として、翻訳精度の確認および修正作業という2つを扱うこととし、それぞれを軽減させる手法について検討・提案する。

3. 研究の方法

本研究では、以下の手順により研究を行う。

<STEP 1>利用場面に応じた機械翻訳の適用可能性の検証

まず、より高精度な翻訳が要求される場面における翻訳リペアの適用を想定し、利用場面に応じた要件について調査する。特に正確性が要求される分野として医療分野を想定し、医療分野における機械翻訳システム利用の問題点の抽出を行う。実際の医療現場で用いられる文章を、機械翻訳システムを用いて多言語化し、問題点の抽出および要件を整理し、機械翻訳の適用可能性を検証する。

<STEP 2>翻訳リペア支援システムの構築

翻訳リペアを支援するシステムとして、ユーザに提供する翻訳文の質を向上させる手法およびユーザ自身の書き換え支援手法を構築する。

ユーザによるリペア作業の負荷軽減手法の一つとして、折り返し翻訳が用いられているが、折り返し翻訳では、対象言語翻訳文との不一致が生じ得るという問題を抱えている(表1)。表1に示す例では、原文中の「行なう」という表現の対訳として、韓国語翻訳文では「指揮する」を意味する語(表1の韓国語翻訳文における下線部)が用いられている。そのため、表1の韓国語翻訳文は文としておかしいと判定される。しかし、システムAを用いて折り返し翻訳文を生成すると、原

表 1 折り返し翻訳文と対象言語翻訳文の精度不一致の発生例

原文	研究会は第五教室において <u>行なう</u> 。
韓国語翻訳文 (システム A)	연구회는 다섯째 교실에서 <u>지휘한다</u> .
折り返し翻訳文 (システム A)	研究会は、五番目の教室で <u>行なう</u> 。
折り返し翻訳文 (システム B)	研究会は五つ目教室で <u>指揮する</u> 。
折り返し翻訳文 (システム C)	研究会は五番目教室で <u>指揮する</u> 。

※表中の対象言語翻訳文では、原文中の「行なう」という語が、「指揮する」を意味する表現（下線部）になっている。

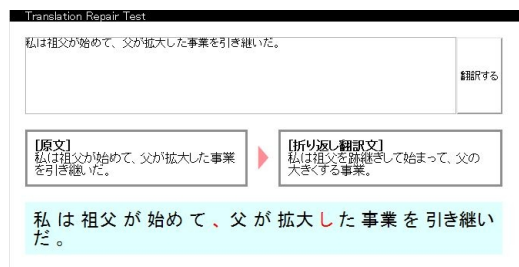


図 2 プロトタイプシステムの画面例

文と同じ「行なう」という表現に戻っており、折り返し翻訳文の内容は原文と一致し、正しく翻訳されていると判断されてしまう。ユーザ自身に対象言語の知識がない場合、この問題の発生に気づくことが困難であるため、折り返し翻訳文と対象言語翻訳文の精度不一致を検出し、適切なリペア作業を支援する手法を構築する。

4. 研究成果

本研究の成果を以下に示す。

(1) 医療分野における翻訳の必要要件と機械翻訳の適用可能性

特に正確性が要求される分野として医療分野を想定し、機械翻訳が利用できるかどうか（翻訳リペアの適用可能性）および医療分野における翻訳文が満たすべき必要要件について調査を行った。医療通訳者へのインタビューにより、従来、医療通訳者はどのような指針で通訳を行っているのかを確認し、医療分野で用いられる翻訳文に対する必要要件を整理した。その結果、必要要件を満たしていれば機械翻訳を利用することも可能であり、翻訳リペアの適用可能性を確認できた。また、現在の機械翻訳は、医療分野での翻訳文に対する必要要件を満たしているのかを

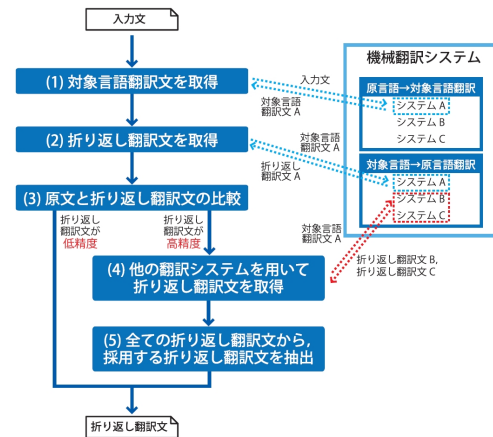


図 3 提案手法における精度不一致検出手法の流れ

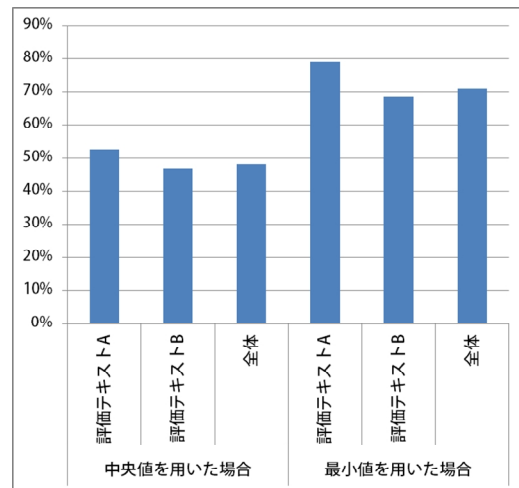


図 4 提案手法による精度不一致検出率

確認するために、英語翻訳文、中国語翻訳文、韓国語翻訳文の精度評価を行った。

(2) 翻訳リペア支援システムの構築

これまでに構築してきた、類語を用いた言い換え文自動生成手法をもとに、良い翻訳結果を得るためのルールに違反するかどうかを検証する仕組みを構築し、プロトタイプシステム（図 2）を構築した。

また、折り返し翻訳文と対象言語翻訳文の精度不一致を検出する手法として、複数の翻訳システムを用いる手法を構築した。提案手法の流れを図 3 に示す。図 3 のように、提案手法では生成した折り返し翻訳文の翻訳精度が高精度である場合、複数翻訳機を用いて折り返し翻訳文を再生成する。それらの翻訳精度から総合的に精度を判定し、ユーザに提示する折り返し翻訳を選択することによ

て、精度不一致を検出している。2種類の評価用テキスト（評価テキスト A、評価テキスト B）を用いて、構築した手法の精度評価実験を行った結果を図4に示す。実験の結果、約71%の精度不一致を検出できることが確認できた。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計1件）

- ① 宮部真衣，吉野孝：折り返し翻訳を用いた高精度なコミュニケーションのための複数翻訳機による精度不一致検出サービスの提案，情報処理学会論文誌，Vol. 53, No. 8, pp. 2036-2043 (2012).

〔学会発表〕（計2件）

- ① 宮部真衣，吉野孝：翻訳精度評価における原文参照の影響分析，電子情報通信学会技術報告，人工知能と知識処理，AI2012-39, pp.103-108 (2013).
- ② 東拓央，宮部真衣，吉野孝：在日外国人を対象としたやさしい日本語文と機械翻訳文間の理解のしやすさの比較，電子情報通信学会技術報告，人工知能と知識処理，AI2012-37, pp.91-96 (2013).

〔図書〕（計0件）

〔産業財産権〕

○出願状況（計0件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計0件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕

ホームページ等

6. 研究組織

(1) 研究代表者

宮部 真衣 (MIYABE MAI)

東京大学・知の構造化センター・特任研究員

研究者番号：00613499

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：