

# 共同翻訳プロトコルの改善と性能評価

## Improvement and Performance Evaluation of Collaborative Translation Protocol

森田 大翼\*<sup>1</sup> 石田 亨\*<sup>1</sup>  
 Daisuke Morita Toru Ishida

\*<sup>1</sup>京都大学情報学研究科社会情報学専攻  
 Department of Social Informatics, Kyoto University

In this paper, we present the protocol for collaborative translation, where two non-bilingual people who use different languages collaborate to perform the task of translation using machine translation (MT) services, whose quality is imperfect in many cases. This protocol 1) improves MT quality; and 2) terminates successfully only when the translation result becomes accurate and fluent. The key idea of this protocol is that one person, who handles the source language (source language side) and another person, who handles the target language (target language side), play different roles: the target language side modifies the translated sentence to improve its fluency, and the source language side evaluates its adequacy. The experiment results show that when the protocol terminates successfully, the percentage of the sentences whose adequacy is evaluated as perfect increases about 83 percent in Japanese-English translation and of 91 percent in Japanese-Chinese translation.

### 1. はじめに

国際化やインターネットの普及等により、我々は多くの言語を見聞きする状況に遭遇している。それに伴い、メンバーの母語が様々であるコミュニティも数多く発生している。従来、そのようなコミュニティはある一つの共通言語を利用しており、多くの場合は英語であった。しかし、非母語の理解やそれを用いたコミュニケーションは困難となる傾向がある [Takano and Noda 1993, Aiken et al. 1994, Kim and Bonk 2002]。

そのようなコミュニティにおいて機械翻訳は有用なツールである。機械翻訳を用いることで母語で自分の意思を伝えることができる。実際、機械翻訳を用いて活動する多くのコミュニティが存在している。

一般に機械翻訳は誤訳を多く生成する。翻訳品質が完全でなくても、いくつかの単語から意図を汲み取ることができ、アウェアネスの向上につながる [Nomura et al. 2003] という点では、機械翻訳は有用である。しかし、機械翻訳を用いたコラボレーションでは、その誤訳は双方に誤解を生じさせる。また、機械翻訳では訳語関係が必ずしも 1 対 1 に対応しないという非対称性の性質があり、誤訳の特定は困難である [Yamashita and Ishida 2006]。

この論文は、バイリンガルでない異なる言語を用いる 2 人の人間が機械翻訳を利用して協力して文書の翻訳作業を行う方式、共同翻訳のプロトコルを提案する。共同翻訳の方式は [森田, 石田 2008] において提案されているが、2 人の作業者のインタラクションの衝突を防止することを解決すべき主な問題としてプロトコルを提案している。しかし、この論文では、実際の異文化コラボレーションの現場で利用されている機械翻訳方式を応用し、1) 機械翻訳の品質を確実に向上させ、2) 正確で流暢な翻訳が得られた場合にそのプロトコルを終了する、共同翻訳プロトコルの存在を示し、そのプロトコルの改善を明らかにする。

また、本研究では、実験によりプロトコルの成功率と信頼度を評価する。本実験結果の分析を通して、上記で示した 1)、2) の性質を十分に保証することを実験的に明らかにした。

### 2. 人間の補助を含む機械翻訳方式

#### 2.1 異文化コラボレーション現場での実践

実際に、多くの異文化コラボレーションの現場ではコミュニケーションツールとして機械翻訳を利用している。国際的に活動している日本の NPO 団体について例を挙げる。この団体は、韓国、オーストリア、ケニアの団体と連携して活動している。それらの団体のメンバーの母語は日本語、韓国語、ドイツ語、英語と様々である。

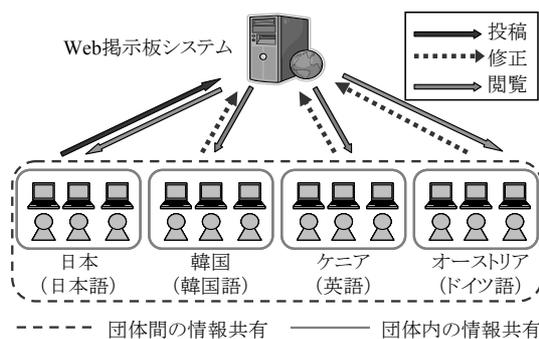


図 1: Web 掲示板システムのイメージ

国際間で情報の共有を促進して議論を活性化させるために、これらの団体間では機械翻訳を用いた Web 掲示板システムを独自に開発して利用している。図 1 にこの Web 掲示板システムのイメージを、日本語記事の投稿を例として示す。この Web 掲示板システムでは、各人は自分の母語で記事を編集する。その記事はシステム内部で翻訳され、各人はその記事を母語で閲覧する。しかし、機械翻訳の品質はしばしば低く、そのことがメンバー間での情報共有を困難にさせる場合がある。そこで、このシステムでは機械翻訳された文を人手によって自然な表現になるように修正する操作を可能にしている。また、文章中の

連絡先: 〒 606-8501 京都府京都市左京区吉田本町京都大学  
 大学院情報学研究科社会情報学専攻石田・松原研究室,  
 075-753-5396

文脈やコミュニティ内での共通知識を用いて文意を推測して修正することで、翻訳品質自体が改善される場合がある。これにより、団体内での適切な情報共有を可能にしている。例 1 に、翻訳文の修正による翻訳品質の改善の様子を示す。

#### 例 1) 修正による翻訳品質の向上

“その絵を見た子供たちはみんな驚きました” という文を機械翻訳すると “Everyone was surprised at the children who saw the picture.” と翻訳される。この文の意味は原文とは異なるが、修正者は文脈やコミュニティの背景から意味を推測し、“Children were surprised to look at the picture.” と修正した。この修正された文は原文の意味と全く正しくなっている。

### 2.2 翻訳文の修正における問題点

翻訳文の修正により、文意が明確になり、団体内での情報共有が容易になる。この点で、異文化コラボレーションの現場で有用な方式であると言える。しかし、この人の補助を含む機械翻訳方式の単純な実装には、主に 2 つ問題がある。以下にその問題を明らかにする。

#### 例 2) 修正者の誤解

“彼は風邪をなおすのに 1 週間必要でした” という文を機械翻訳すると、“He was necessary to correct a cold for 1 week.” というように翻訳される。この文は文法、語法に誤りがあり、自然な表現に修正されることになるが、修正者はこれを “He should recover from a cold within 1 week.” と修正した。これは、“彼は 1 週間以内に風邪をなおすべきです” という意味になる。

#### 例 3) 文意の理解不能

“彼はお腹が出ている” という文を機械翻訳すると、“A stomach has gone out to him.” と翻訳される。この文は修正者にとって理解不可能な文であるため、この文は修正されないまま残った。

人の補助を含む機械翻訳方式において、

- 修正された翻訳文が翻訳前の文の意味と正しいかどうか判定できない
- 翻訳文の意味が理解できない場合、修正されず品質の低いままになる

という問題がある。この方式は、翻訳文の品質を改善して情報を共有する手法として機械翻訳を用いたコラボレーションの現場で有用な手法であるが、これら 2 つの問題を解決できない限り、正しい情報共有をし、誤解を防ぐという課題を根本的に解決できない。

## 3. 共同翻訳

### 3.1 定義

参加者はバイリンガルでない 2 人の人間であり、文書の翻訳元言語を理解する人間(原文側)と、翻訳先言語を理解する人間(翻訳文側)に分かれる。その 2 言語間の翻訳は翻訳機が行う。作業者は端末を用い、ネットワークを介して作業する。共同翻訳は文書翻訳を対象とする。ここで、元の原文自体に手は加えられないが、原文側は原文の参考のために代替案を提示できる。

また、原文側と翻訳文側の役割を明確にする。翻訳文側では翻訳文と原文の意味的類似度を評価できないが、翻訳文の流暢

さは評価できる。翻訳文側は翻訳文から文意を推測し、修正して翻訳文を流暢にする。また、人が修正した文は常に流暢であるとする。翻訳文側と同様、原文側は翻訳文と原文の意味的類似度を直接評価できない。しかし、翻訳文側の修正翻訳文を、原文の言語に翻訳し直した文と原文が意味的に類似しているかは評価できる。共同翻訳では、その情報を利用し、翻訳文が原文の意味内容を正しく反映しているかを確認する。

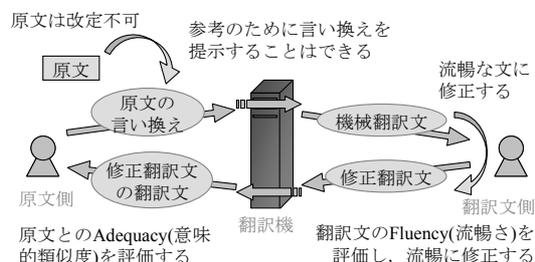


図 2: 共同翻訳のイメージ

図 2 に、以上の定義を表現した図を示す。本研究では上記に定義された共同翻訳を適切に実行するためのプロトコルを設計する。

### 3.2 基本プロトコル

効果的なコラボレーションの為には、相互理解の構築が不可欠である [Clark and Gibbs 1986]。相互理解の構築過程は、情報の提供を行う提案フェーズと、その情報を受け手が理解できたかを確認し合う受理フェーズからなる [Clark and Brennan 1991]。共同翻訳では、提案フェーズは原文を翻訳して送信することであり、受理フェーズは、翻訳文の受信以降の、その翻訳文の正しさの確認までの処理全般を意味する。この受理フェーズにおいて翻訳文の正しさが適切に確認できるようにプロトコルを設計する必要がある。

図 3 に共同翻訳を実現するための最小のプロトコルである基本プロトコルの形式的記述を示す。

このプロトコルでは、上記の定義に加え、機械翻訳文の可読性判定を行っている。翻訳文側の流暢さの評価、原文側の意味的類似度の評価は元の機械翻訳文が可読でない限り評価することができないためである。

### 3.3 有効性

以下に 2.2 章の例 2 にプロトコルを適用した場合に起こる一連の処理を示す。原文は「彼は風邪をなおすのに 1 週間必要でした」という文である。翻訳文側は最初の機械翻訳文 “He was necessary to correct a cold for 1 week.” を誤解して解釈し、“He should recover from a cold within 1 week.” と修正している。原文側はその修正文の折り返し翻訳文 “彼は 1 週間以内に風邪から回復するべきである” を読んで原文の意味と異なっていると判断する。そこで、原文を「彼は風邪をなおすのに 1 週間かかった」と修正して再度翻訳文側に送信する。翻訳文側は “It took 1 week for him to correct a cold.” という翻訳文を見、前回とは異なる解釈をして “It needs 1 week to recover from his cold.” と翻訳文を修正する。原文側はその文の折り返し翻訳文 “彼の風邪から回復するために、それに 1 週間が必要である” は、十分に原文の意味を表していると判断する。プロトコルは成功で終了する。

以下に 2.2 章の例 3 にプロトコルを適用した場合に起こる一連の処理を示す。原文「彼はお腹が出ている」が入力され翻訳文側に送信される。翻訳文側は最初の翻訳文 “A stomach

原文を  $s$ 、文の意味 (content) が理解できる範囲か否かの真偽値を返す関数を  $c$ 、原文  $s$  との正確さ (accuracy) が許容される範囲か否かの真偽値を返す関数を  $a$ 、文の流暢さ (fluency) が許容される範囲か否かの真偽値を返す関数を  $f$  とする。

また、人間による文の修正を示す関数を  $m$ 、翻訳機による文の翻訳を示す関数を  $t$  とする。例えば原文  $s$  の翻訳文の修正文は  $m(t(s))$  のように表す。

原文側の修正回数を  $p$  その上限を  $P$  とする。また、翻訳側の修正回数を  $q$  その上限を  $Q$  とする。

- 1) [原文側：原文の送信]  
原文側の修正回数  $p := 0$  とする。  
原文  $s$  を翻訳文側に送信する。
- 2) [翻訳文側：翻訳文の可読性判定]  
 $s$  の機械翻訳結果を  $t(s)$  とする。  
もし、 $c(t(s))$  が偽なら：  
原文側に原文  $s$  の修正を依頼し 6) へ。  
もし、 $c(t(s))$  が真なら：  
翻訳文側の修正回数  $q := 0$  とする。
- 3) [翻訳文側：翻訳文の修正]  
翻訳文側の修正回数  $q := q + 1$  とする。  
もし、 $q \geq Q$  なら：  
プロトコルを終了し失敗を返す。  
もし、 $q < Q$  なら：  
流暢性  $f(m(t(s)))$  を満たすよう翻訳文  $t(s)$  を修正した結果を  $m(t(s))$  とし原文側に送信する。
- 4) [原文側：折り返し翻訳文の可読性判定]  
 $m(t(s))$  の機械翻訳結果を  $t(m(t(s)))$  とする。  
もし、 $c(t(m(t(s))))$  が偽なら：  
翻訳文側に  $m(t(s))$  の修正を依頼し 3) へ。  
もし、 $c(t(m(t(s))))$  が真なら：5) へ。
- 5) [原文側：折り返し翻訳文の正確性判定]  
もし、 $a(t(m(t(s))))$  が真なら：  
プロトコルを終了し成功を返す。  
もし、 $a(t(m(t(s))))$  が偽なら：6) へ。
- 6) [原文側：原文の修正]  
原文側の修正回数  $p := p + 1$  とする。  
もし、 $p \geq P$  なら：  
プロトコルを終了し失敗を返す。  
もし、 $p < P$  なら：  
原文を修正した結果を  $s$  とし 2) へ。

図 3: 基本プロトコルの形式的記述

has gone out to him.”の意味が理解できないため、原文側に原文の修正を依頼する。原文側によって原文が「彼はすこし太っている」と修正された後、再度得られた機械翻訳文は“He’s a little overweight.”となる。この翻訳文は理解可能であるため、翻訳文側は翻訳文が流暢になるように“He’s a little bit overweight.”と修正する。原文側はその文の折り返し翻訳文「彼はわずかに太りすぎである」は十分に原文の意味を表している」と判断する。プロトコルは成功で終了する。

共同翻訳の基本プロトコルでは、作業者は修正翻訳文が原文の意味と正しいかどうかを判断することができる。また、機械翻訳文が理解できない場合に、他方の作業者に翻訳元の文を再度修正させることができる。このプロトコルは人の補助を含む機械翻訳方式における 2 つの問題点を解決する。

しかし、語の多義性や構文解析の曖昧性等の言語学的な要因に伴う機械翻訳の性能の不完全さが存在する。更に、判断の誤りや、「理解できる」の判断基準が人によって異なるという人的要因も存在する。それらの性質より、共同翻訳が 100% 正確な結果とならない可能性がある。プロトコルの実験を行い、

プロトコルの有効性を実験的に確認する。

## 4. プロトコルの評価実験

### 4.1 実験設定

基本プロトコルの評価実験を行った。実験に用いた機械翻訳のモジュールとして、独立行政法人情報通信研究機構 (NICT)、京都大学を中心として研究開発、及び運営されている言語グリッドプロジェクト [Ishida 2006] にて提供される Web サービスを用いた。

NTT のコミュニケーション科学基礎研究所で提供されている機械翻訳機能試験文 [池原他 1994] 3,718 文から日本語特有の固有名詞を含む文を除き、日本文 100 文を無作為抽出して、実験に用いられた。翻訳先言語は英語と中国語とされた。また、人的な影響を緩和するため、実験の参加者はそれぞれの言語対について 3 組で行われた。また、本実験では原文側及び翻訳文側の修正回数の上限  $P$ 、 $Q$  はそれぞれ 3 と設定された。

また、異文化コラボレーションの現場で機械翻訳を改善する手法として広く用いられている折り返し翻訳についても同様に実験が行われた。共同翻訳と同様に修正回数の上限は 3 回と設定し、上限回数以内の修正で翻訳文に改善が見られない場合は失敗とみなす。

それぞれ的方式で得られた翻訳文は、対となる 2 言語のバイリンガルによって、それぞれ意味の正確さに関して 5 段階 (All, Most, Much, Little, None) で評価が行われた<sup>\*1</sup>。評価は 3 人の評価者によって行われ、その中央値を最終的な評価値として用いた。人間によって修正された文章は完全に流暢であると仮定しているため、流暢さに関する評価は行わない。

### 4.2 評価結果

#### 4.2.1 成功率

日英翻訳では折り返し翻訳の成功率は約 67%、共同翻訳の成功率は約 70%であった。一方、日中翻訳では折り返し翻訳の成功率は約 78%、共同翻訳の成功率は約 62%であった。折り返し翻訳と共同翻訳の成功率の間に相関は見られず、ほぼ同程度の成功率であった。

#### 4.2.2 翻訳品質

図 4 は日英と日中の翻訳について、機械翻訳、折り返し翻訳、共同翻訳の結果の比較であり、それぞれの方式が成功した場合にその翻訳結果の評価が“All”となった割合と、失敗した場合も含む全データに対して翻訳結果の評価が“All”となった割合を示している。共同翻訳が成功した場合、日英で約 83%、日中で約 91%が完全な翻訳品質となった。共同翻訳プロトコルが成功していれば、その文の正しさに関して高い信頼をおけるといえる。前章に、理論的にはプロトコルが成功しても正しい翻訳文が生成されない可能性があることを示したが、本実験によりその可能性は小さいことが確認された。また、折り返し翻訳と比較した場合、翻訳品質が完全となる割合は、成功時のみのデータ、失敗した場合を含めた全データのそれぞれと比較して共同翻訳が大きいことが示されている。

表 1 は機械翻訳結果に対する共同翻訳結果の評価値の推移を示している。3 組の被験者により実験が行われたため、300 文の共同翻訳による結果が得られている。機械翻訳より評価値が下がった文が日英で 5%、日中で 1%程度と若干見られたが、プロトコルが成功で終了している場合、その翻訳結果の評価は機械翻訳より概ね高くなっている。この結果は共同翻訳は機械

\*1 <http://projects ldc.upenn.edu/TIDES/Translation/TransAssess02.pdf>

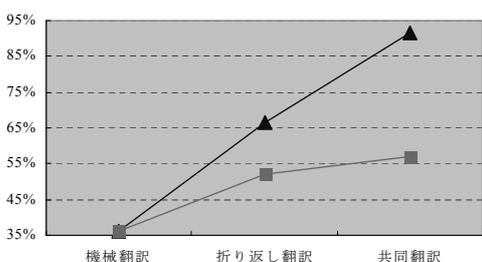
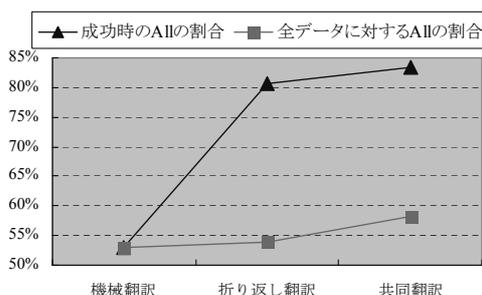


図 4: 機械翻訳, 折り返し翻訳, 共同翻訳の評価結果の比較

翻訳の品質を低下させることがほとんど無く, 確実に翻訳品質を向上させていることを示している.

## 5. むすび

近年, 多くのコミュニティで機械翻訳を用いたコラボレーションが行われている. 品質の低い機械翻訳を利用する上で, 機械翻訳文を修正することは, 情報共有のために有用である. しかし, その単純な実装では, 修正翻訳文の正しさを保証できない欠点がある.

本論文の貢献はバイリンガルのいない状態で, 1) 機械翻訳の品質を確実に向上させ, 2) 正確で流暢な翻訳が得られた場合にプロトコルを終了する, 共同翻訳プロトコルの存在を示したことである. 翻訳文側では, 機械翻訳文から意味を推測して流暢な翻訳文に修正できることを明らかにした. 原文側では, その折り返し翻訳文が, 原文の意味を表しているかを評価し,

表 1: 機械翻訳結果に対する共同翻訳結果の評価値の推移

		共同翻訳					失敗
		5	4	3	2	1	
機械翻訳	5	117	10	0	2	3	27
	4	21	5	0	0	0	22
	3	15	5	4	0	0	18
	2	11	1	2	0	0	16
	1	11	1	1	0	1	7

		共同翻訳					失敗
		5	4	3	2	1	
機械翻訳	5	84	2	1	0	0	21
	4	30	6	0	0	0	18
	3	39	1	1	0	0	22
	2	16	2	1	2	0	51
	1	2	0	0	0	0	1

翻訳の正しさを確認できることを明らかにした.

機械翻訳のみを適用した場合に, 完全に正確な意味に翻訳された文の割合は日英翻訳で 53%, 日中翻訳で 36%であったにも関わらず, 共同翻訳を適用してプロトコルが成功で終了すれば, その割合は日英翻訳で約 83%, 日中翻訳では約 91%にまで改善されたことが, プロトコルの初期的評価より示された. 即ち, プロトコルが成功した場合の翻訳文の信頼度がかかり高くなることが実験的に示された.

謝辞 京都大学グローバル COE プログラム「知識循環社会のための情報学教育研究拠点」並びに総務省戦略的情報通信研究開発推進制度から助成を受けた.

## 参考文献

- [Takano and Noda 1993] Y. Takano, and A. Noda, "A temporary decline of thinking ability during foreign language processing," Journal of Cross-Cultural Psychology, vol.24, no.4, pp.445-462, 1993.
- [Aiken et al. 1994] M. Aiken, C. Hwang, J. Paolillo, L. Lu, "A group decision support system for the asian pacific rim," Journal of International Information Management, vol.3, no.2, pp.1-13, 1994.
- [Kim and Bonk 2002] K.J. Kim, and C.J. Bonk, "Cross-cultural comparisons of online collaboration," Journal of Computer Mediated Communication, vol.8, no.1, Oct 2002.
- [Nomura et al. 2003] S. Nomura, T. Ishida, N. Yamashita, M. Yasuoka, and K. Funakoshi, "Open source software development with your mother language: Intercultural collaboration experiment 2002," vol.4, pp.1163-1167, 2003.
- [Yamashita and Ishida 2006] N. Yamashita, and T. Ishida, "Effects of machine translation on collaborative work," International Conference on Computer Supported Cooperative Work(CSCW-06), pp.512-523, Nov 2006.
- [森田, 石田 2008] 森田大翼, 石田亨, "共同翻訳のためのプロトコルと支援システムの開発," 情報科学技術フォーラム FIT, pp.417-420, 2008.
- [Clark and Gibbs 1986] H.H. Clark, and D. Wilkes-Gibbs, "Referring as a collaborative process," Cognition, vol.22, pp.1-39, 1986.
- [Clark and Brennan 1991] H.H. Clark, and S.E. Brennan, "Grounding in communication," In L.B. Resnick, J.M. Levine, & S.D. Teasley (Eds.), Perspectives on socially shared cognition, APA Books, Washington, 1991.
- [Ishida 2006] T. Ishida, "Language grid: An infrastructure for intercultural collaboration," IEEE/IPSJ Symposium on Applications and the Internet(SAINT-06), pp.96-100, 2006.
- [池原他 1994] 池原悟, 白井諭, 小倉健太郎, "言語表現体系の違いに着目した日英機械翻訳機能試験項目の構成," 人工知能学会誌, vol.9, no.4, pp.569-579, 1994.