

クラウドソーシング上の単言語話者による 複数の機械翻訳を用いた用例対訳作成手法の提案

山本 里美¹ 福島 拓² 吉野 孝^{3,a)}

受付日 2017年4月14日, 採録日 2017年10月3日

概要: 従来研究より, クラウドソーシングを用いて単言語話者に機械翻訳文の評価と訂正タスクを依頼することで, 作業者が単言語話者であっても, 一部の用例について正確な用例対訳作成が行えることが分かっている. しかし, 翻訳対象の用例によって機械翻訳の翻訳精度には差があり, 多くの用例について精度の高い対訳作成を行うことは困難であった. そこで, 本論文では, 従来の用例対訳作成手法に機械翻訳適応文と複数の機械翻訳機を利用する手法を提案する. 機械翻訳適応文とは, 原文とほぼ同じ意味であるが, 適切な機械翻訳が行われる文のことである. 機械翻訳適応文を用いることで, 機械翻訳の精度が低い用例における機械翻訳文の訂正精度向上を目指す. 本論文の貢献は, 従来の用例対訳作成手法に対して, 機械翻訳適応文と複数の機械翻訳を用いる手法の提案, および機械翻訳適応文評価手法として人手による評価と計算機による評価を組み合わせることを提案し, 従来手法より 25 ポイント高い 40 文中 39 文 (97%) の原文において, 適切な機械翻訳文の作成が可能であることを示した点である.

キーワード: クラウドソーシング, 多言語間コミュニケーション, 用例対訳, 機械翻訳

Proposal of Parallel Texts Creation Method Using Multiple Machine-translation by Monolingual on Crowdsourcing Workers

SATOMI YAMAMOTO¹ TAKU FUKUSHIMA² TAKASHI YOSHINO^{3,a)}

Received: April 14, 2017, Accepted: October 3, 2017

Abstract: Our former study showed that crowdsourcing workers, even they are monolingual speaker, can make correct parallel texts to request tasks. There is a difference in the translation accuracy of the machine translation according to the sentences of the translation target. It is difficult to create bilingual sentences for many sentences. In this paper, we propose a method to use a sentence adapted to machine translation and to use multiple machine translation engines for the conventional method. A sentence adapted to machine translation is a sentence that is almost the same as the original, but the appropriate machine translation is done. We aim to improve the accuracy of machine translation sentences in the sentences of a machine translation with a low accuracy. The contribution of this paper is as follows. We proposed a method to use a sentence adapted to machine translation and to use multiple machine translation engines in the conventional method for the conventional method. We also proposed a combination of evaluation by the person and the evaluation by the computer as a method of evaluation of a sentence adapted to machine translation. We have shown that the proposed method is capable of making the appropriate machine translation sentences in the original 39 of 40 sentences (97%) with of 25 points higher than the conventional method.

Keywords: crowdsourcing, multilingual communication, parallel text, machine translation

¹ 和歌山大学大学院システム工学研究科
Graduate School of System Engineering, Wakayama University, Wakayama 640-8510, Japan

² 大阪工業大学情報科学部
Faculty of Information Science and Technology, Osaka Institute of Technology, Hirakata, Osaka 573-0196, Japan

³ 和歌山大学システム工学部
Faculty of Systems Engineering, Wakayama University, Wakayama 640-8510, Japan

a) yoshino@sys.wakayama-u.ac.jp

1. はじめに

世界的なグローバル化により, 多言語間コミュニケーションの機会が増加している. 日本でも, 2013 年以降, 継続して年間訪日外国人数が過去最高を記録していること [1] や, 日本政府が 2020 年に訪日外国人数 4,000 万人という目標を掲げていること [2] などから, 今後も訪日外国人数の増加は継続すると考えられる. 多言語間コミュニケーション

ンの機会の増加にともない、多言語間コミュニケーション支援の必要性も高まる。

現在、我々は、多言語用例対訳共有システム TackPad (タックパッド)*1によって、医療に関する用例対訳の収集を行っている [3]。TackPad には多くの用例や用例対訳が収集されているが、医療現場で用いるために必要とされる数*2には足りておらず、また、正確性評価が十分にされていないものも多い。そこで、我々はクラウドソーシング上の単言語話者を用いた多言語用例対訳作成手法に関する研究も行っている。我々の提案する手法では、クラウドソーシング上の単言語話者に機械翻訳文の評価と訂正を依頼することで、対訳の作成を行う。通常の場合の対訳作成には、多言語話者が関わるために大きなコストが必要となる。本提案手法は、各言語の単言語話者と機械翻訳の利用のみで用例対訳を作成する手法である。

これまでの研究として、文献 [4] では、機械翻訳文の訂正タスクにおいて、作業者に「用例の使用者の属性情報」と「翻訳対象文を含む会話文」を提示する手法を提案した。提案手法により、適切な訂正文が作成される割合が、機械翻訳文 1 文のみを提示する場合と比較して 16 ポイント増加することを示した。また、文献 [5] では、機械翻訳文の訂正精度向上のために機械翻訳適応文を利用する手法を提案した。機械翻訳適応文とは、原文と同じ意味を示すが、適切な機械翻訳が行われるような文のことである。

本論文の目的は、文献 [4] や文献 [5] の実験より明らかになった、以下の 2 点の課題の解決である。

(1) 低精度の機械翻訳

クラウドソーシング上の作業員には、作業員が母語とする言語に翻訳された機械翻訳文のみが提示される。そのため、機械翻訳の精度が低すぎる場合には、作業員が機械翻訳文から原文の意味を推測することが困難となり、適切な訂正が行えない場合が存在した。

(2) 折り返し翻訳の精度不一致

文献 [5] で提案する機械翻訳適応文作成手法では、折り返し翻訳を用いる。これは、折り返し翻訳文との一致度が高い文が作成できる場合、適切な機械翻訳が行われている可能性が高いと考えたためである。しかし、文献 [5] で行った評価実験では、機械翻訳が不適切であっても折り返し翻訳文が原文と似る場合や、機械翻訳が適切であっても折り返し翻訳文が原文と大きく異なる場合が存在し、適切な機械翻訳適応文を作成することができなかった。文献 [6] でも、折り返し翻訳時の精度不一致の問題に着目し、複数翻訳機を用いることによる、精度不一致の検出を試みている。本研究では、複数の機械翻訳を用いるだけでなく、機械翻訳適応文の評価方法に人手と計算機による評価を組み合わ

せる手法を提案している。

2 章で関連研究について、3 章で提案手法について述べる。4 章では対訳作成実験について述べ、5 章では実験結果と考察について述べる。6 章において、本研究のまとめについて述べる。

2. 関連研究

クラウドソーシングを用いて多言語データを収集する研究は多く行われている。Zaidan らは、クラウドソーシングを用いて複数の作業員に翻訳作業を依頼し、その作業結果の統合を行うことで、専門家による翻訳結果と同等の翻訳結果を取得できることを明らかにした [7]。また、Callison-Burch らによる研究では、クラウドソーシングによる翻訳文の評価手法では、一般的に機械翻訳の評価に用いられている評価尺度 BLEU [8] を用いる場合よりも専門家による評価との相関が高い評価結果が取得可能であることを明らかにした [9]。なお、これらの研究では多言語話者を作業員として設定している。そのため、クラウドソーシング上の作業員の確保や、話者の少ない言語への対応に課題がある。

クラウドソーシング上の単言語話者を対象とした研究も行われている。Hu は、機械翻訳文に対して後編集を行うことで、正確な翻訳文の取得を目指している [10]。後編集とは、機械翻訳文に対して、手作業で低精度な翻訳部分を修正することである。後編集を行うことで、単純な機械翻訳文よりも正確な翻訳文を取得することが可能となる [11], [12]。文献 [10] で提案する手法では、2 種類の単言語話者グループを作業員としている。それぞれのグループの作業員が、母語に翻訳された機械翻訳文に対して後編集を行う。2 つのグループ間で繰り返し後編集を行うことで、正確な翻訳文な翻訳文の取得を目指している。文献 [4] や文献 [10] での評価実験より、クラウドソーシング上の単言語話者が機械翻訳文の訂正を行うことで、適切な対訳を取得できることが明らかにされている。

しかし、従来の対訳作成手法では、機械翻訳精度が低すぎる場合に対応できていない。

本論文では、この問題に対応するために、機械翻訳が適切に行われる機械翻訳適応文の利用を提案する。また、機械翻訳適応文の作成や評価に、複数の機械翻訳や折り返し翻訳を用いる手法を提案した。複数の機械翻訳を用いることで、折り返し翻訳の精度不一致へ対応する。

3. 提案手法

3.1 提案手法の概要

文献 [4] の手法について、以下の変更を行ったものを提案手法とする。文献 [4] の手法は、用例対訳として作成したい文の機械翻訳文を、会話文の一部に含めることで、用例対訳を作成する手法である。なお、本論文では文献 [4]

*1 <http://med.tackpad.net/>

*2 1 言語あたり 3 万~5 万文 [3]。

の手法を従来手法とする。

- (1) 機械翻訳適応文作成の工程を追加
- (2) 複数の機械翻訳を利用
- (3) 機械翻訳文の訂正タスクで 3 種類の機械翻訳文を提示

従来手法では、低精度の機械翻訳が行われた場合に作業者が適切な訂正を行えない場合が存在した。本論文では、低精度の機械翻訳に対応するために機械翻訳適応文を用いることで、訂正精度の向上を目指す。

3.2 提案手法を用いた用例対訳作成手法の流れ

図 1 に提案手法における用例対訳作成の流れを示す。本手法では、原文の言語（言語 A）を母語とする作業（以下、グループ A）と、翻訳対象言語（言語 B）を母語とする作業（以下、グループ B）の 2 つの作業グループによって用例対訳作成を行う。作業者は、「タスク上一方の言語のみ扱う話者」であり、必ずしも「2 つの言語のうち一方しか使えない話者」ではない。提案手法では、ある 1 つの言語（母語）を理解していれば、評価や訂正作業が可能である。

図 1(1), (2), (7) はグループ A の作業者が作業を行う。図 1(4), (5) はグループ B の作業者が作業を行う。図 1(2) が、提案手法で新しく追加した工程である。グループ B の作業者が、図 1(4) で機械翻訳文の評価と訂正を行うことで、単純な機械翻訳よりも正確な対訳の取得を行うことができる。なお、図 1(1) で作成された会話を「原文の会話文」とし、図 1(2) で作成された文を「機械翻訳適応文」、

図 1(4) で作成される機械翻訳文の訂正文を「対訳準備候補文」、図 1(5) で取得できる文を「対訳候補文」、図 1(7) において評価された対訳候補文を「用例対訳」としている。

3.3 機械翻訳適応文作成タスク

機械翻訳適応文作成（図 1(2)）のためのタスク画面を図 2 に示す。タスク画面は、タスクの説明を行う「説明部分」と、機械翻訳適応文作成の対象となる原文を含んだ会話文を提示する「会話文部分」、作業者が機械翻訳適応文作成を行う「作業部分」からなる。

- 作業には、以下の手順で作業を行うように指示をした。
- (1) あらかじめ入力欄に入力されている「元の文」を「再翻訳ボタン」*3 をクリックすることで折り返し翻訳する。
 - (2) 4 種類*4 の折り返し翻訳結果（折り返し翻訳文と入力文との一致度）を確認する。
 - (3) 折り返し翻訳結果が不自然だと感じたり、提示される一致度が低いと感じたりした場合、入力文の書き換えを行う。
 - (4) 「再翻訳ボタン」をクリックし、折り返し翻訳結果を確認する。
 - (5) 4 つの折り返し翻訳文のうち 1 つでも入力文と同じような文になるように入力文の修正を繰り返す。
 - (6) 4 つの折り返し翻訳文のうち最も良いと思う文にチェックをつける（複数選択可）。
 - (7) 「タスク完了ボタン」をクリックしてタスクを完了する。
- 本研究では 4 種類の機械翻訳サービスを利用している。この数は、宮部ら [6] が利用している機械翻訳機の数（宮部らは 3 種類を利用）を参考に決定した。

なお、指示文中の「同じような文」というのは、同じ文や類似文、同様の意味を示している文などのことであり、作業者が主観的に判断を行う。

作業者は、提示された会話文を参考に、指定された原文の意味を変えない程度に修正していくことで、折り返し翻訳文との一致度が高くなるような文の作成を行う。作業者に提示する一致度は、入力文と折り返し翻訳文における RIBES [13] の値を 10 段階で表したものである。折り返し翻訳文との RIBES の値が高い場合、その入力文は機械翻訳に適応した文である可能性があると考えられる。しかし、折り返し翻訳の精度不一致 [6] により、機械翻訳が不適切であっても折り返し翻訳文が原文と似る場合や、機械翻訳が適切であっても折り返し翻訳文が原文と大きく異なる場合が存在する。文献 [6] の分類から、精度不一致の種類には、第 1 種と第 2 種の大きく 2 種類ある。表 1 に、第 1 種と第 2 種の精度不一致の具体例を示す。

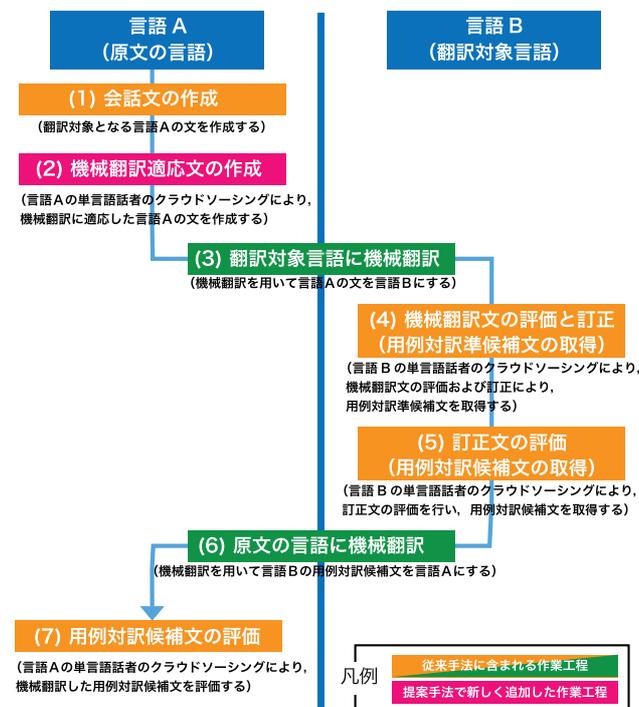


図 1 提案手法を用いた用例対訳作成の流れ

Fig. 1 A flow of parallel translation using the proposed method.

*3 「再翻訳」と「折り返し翻訳」は同じ意味である。実験時の画面では、分かりやすさのために「再翻訳」を用いている。

*4 本研究では、後述の 4.1 節で述べるように、折り返し翻訳に 4 種類の機械翻訳サービスを利用している。



図 2 機械翻訳適応文作成タスク画面例
 Fig. 2 Example screen of a sentence adapted to machine translation task.

そこで、提案手法では、精度不一致への対応のために、複数の機械翻訳を用いることで、1つの入力文から複数種類の折り返し翻訳文を確認できるようにした。これにより、作業者が翻訳精度の不一致を考慮した機械翻訳適応文の作成を行うことを期待する。

作業者は提示される複数の折り返し翻訳結果のうち、適切に折り返し翻訳がされていると判断できるものにチェックを付けることができる。これにより、RIBESの値は低いが入力文の意味をほとんど変えることなく折り返し翻訳を行える文を取得することができる可能性がある。

また、機械翻訳適応文作成の手順のほかに、折り返し翻

訳を行う際のコツ [14] をいくつか提示した*5。

3.4 機械翻訳文の訂正タスク

従来手法では、機械翻訳文の訂正タスク (図 1(4)) で1種類の機械翻訳文を提示していた。そのため、その機械翻訳文の翻訳精度が低い場合に、作業者が原文の意味を推測できず、適切な訂正が行われなかった場合があった。また、文献 [5] では、折り返し翻訳文との一致度が最も高い機械翻訳適応文1文を用いた機械翻訳文の訂正タスクを作成した

*5 作業者には、「主語・述語を省かない」「できるだけ漢字で書く」「曖昧な表現・難しい表現 (比喩など) は避ける」「意味が変わらない程度に) 単語を書き換える」「慣用表現を避ける」の5つを提示する。これは、文献 [14] を参考にした。

表 1 精度不一致の例

Table 1 Examples of an accuracy mismatch between a target-translated sentence and its back-translated sentence.

種類	原文	翻訳文	折り返し翻訳文
第 1 種	住友商事は米国の CATV 局を買収するため、四月をメドに投資グループを結成する。	To bribe an American CATV bureau, Sumitomo Corp. forms an investment group around April.	アメリカ CATV 局を買収するために、住友商事は 4 月ごろに投資グループを結成する。
第 1 種	수염이 난다. (日本語訳: 髭が生える)	ひげが飛ぶ。	수염이 난다. (日本語訳: 髭が生える)
第 2 種	日本の経済は発展した。	Japanese economy is developed.	日本経済は開発される。
第 2 種	何でもいいですよ。	不管什么都好哟。 (日本語訳: 何でも構いませんよ。)	何もすきに関わらず。

第 1 種: 折り返し翻訳文の精度が高いが、翻訳文の精度が低い, 第 2 種: 折り返し翻訳文の精度が低い, 翻訳文の精度が高い [6]. 一部のテキストは, 日英機械翻訳機能試験文 (<http://www.kecl.ntt.co.jp/icl/lirg/resources/mt-test-set-1.txt>) を利用。

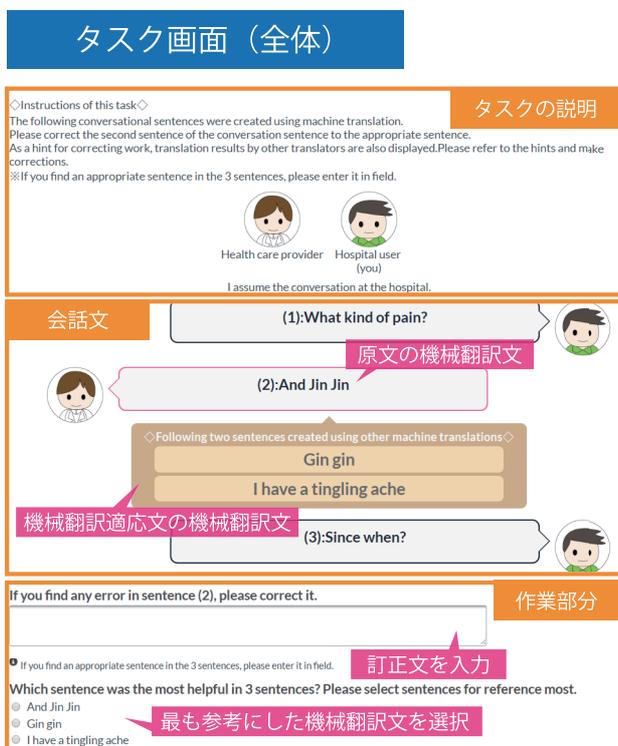


図 3 機械翻訳文の訂正タスク画面例

Fig. 3 Example screen of machine translation correction task.

が, 実験の結果, これは文献 [6] における折り返し翻訳の精度不一致に対応できていないことが明らかになった。

そこで提案手法では, 機械翻訳文の訂正タスクにおいて, 以下に示す 3 種類の機械翻訳文を提示する。また, タスク画面の例を図 3 に示す。

- 原文の機械翻訳文
 - 折り返し翻訳文との一致度 (RIBES の値) が最も高い機械翻訳適応文の機械翻訳文
 - 作業者が適切と判断し, チェックを付けた機械翻訳適応文の中で, 最も折り返し翻訳文との一致度が低い機械翻訳適応文の機械翻訳文
- 文献 [4] より, 作業者に提示する情報が多いほど, 機械翻

訳文の訂正精度が向上することが明らかになっている。そこで, 提案手法では異なる 3 種類の機械翻訳文を提示し, それを訂正作業のヒントとして用いることで, 機械翻訳文の訂正精度の向上を目指す。

機械翻訳文の訂正タスクにおいて, 提示する文の数を 3 種類に限定している理由を示す。

- クラウドソーシングのコスト
クラウドソーシングの利用を前提しており, すべての機械翻訳適応文の提示はコストが掛かる。一般にクラウドソーシングにおいては, 同一作業を複数人で実施することにより, 精度を維持する。詳細は 4 章で述べるが, 本研究では, 1 つの文の作業を 10 名に依頼する。
- クラウドソーシングの作業への負担
同時に多数の文を提示した場合, クラウドソーシングにおいては十分な精度が期待できない可能性が高い。
- 機械翻訳の精度
もともと機械翻訳文そのものは必ずしも精度が低いわけではないため, 多数の候補を示す必要はない。

折り返し翻訳文との一致度が低い機械翻訳適応文を提示する理由を述べる。通常は, 原文と折り返し翻訳文との一致度は高い。そのなかで, 作業者が適切な文と判断したもので, 一致度が低いものは, 精度不一致の可能性があり, 確認が必要である。つまり, 第 2 種の精度不一致の問題の解消が期待できる。

4. 対訳作成実験

クラウドソーシングを用いて作成した日本語 (言語 A) の会話文 40 組を用いて用例対訳作成実験を行う。本実験の目的は, 提案手法で新たに追加した「機械翻訳適応文の作成」により, 機械翻訳文の訂正精度の向上を確認し, 提案手法の有用性を確認することである。翻訳対象言語 (言語 B) は英語である。4.1 節で機械翻訳適応文作成実験 (図 1(2)) について述べ, 4.2 節で機械翻訳文の訂正文作成実験 (図 1(4)) について述べる。

4.1 機械翻訳適応文作成実験

3.3 節で述べたタスクを用いて機械翻訳適応文作成実験を行った。40 組の会話文の 2 文目を機械翻訳適応文作成の対象とし、表 2 に示す設定で実験を行った。実験に使用した会話文の例を表 3 に示す。また、機械翻訳適応文作成の対象となる 2 文目の例を、表 4 に示す。会話文は、クラウドソーシングを用いて作成した、医療現場での使用を想定した文であり、治療に関する会話以外に、受付や会計に関する会話などが含まれる。表 4 のように、適切な主語が含まれていない文や、オノマトペや比喩などが使われている文が作成された。40 組の会話文の 2 文目の平均は 13 文字、標準偏差は 5.6 文字である。最大文字数は 27 文字（表 4 の 26）で、最小文字数は 4 文字（表 4 の 24）である。

作業者は 1 タスクあたり 5 組の会話文について作業を行う。なお、5 組のうち 1 組は検証の対象としている 40 組の会話文に含まれていない会話文である。これは、作業者

表 2 機械翻訳適応文作成タスクの設定

Table 2 Settings of a sentence adapted to machine translation task.

設定項目	設定値
対象文数	40 文
1 文あたりのタスク実行回数	10 回
1 タスクあたりの設問数	5 問
チェック設問の有無	有
1 人あたりのタスク回数上限	5 回
総設問数	500 問
総タスク数	100 件
1 タスクあたりの報酬	10 ポイント*

* 作業者に支払われる報酬は T-ポイントとして支払われる。T-ポイントは、カルチャ・コンビビニエンス・クラブが展開するポイントサービスである。

表 3 実験に使用した会話文の例

Table 3 Example of conversational sentences used in experiments.

	原文
1 文目	どんな靴をはけばよいですか？
2 文目	足に負担のかからない物にして下さい。
3 文目	わかりました。

表 4 作成の対象となる文の例

Table 4 Examples of target sentences used in experiments.

番号	原文
1	頭がチクチクします。
7	この辺がズキズキするんですが。
10	しくしくするんです
15	健康保険はききますか？
24	あります
26	患部を触った手で他の場所に触れないようにしてください。
27	足に優しい靴をはいてください

が正確に作業をしているかを確認するためのチェック設問として用いる設問である。そのため、作業者は 1 タスクあたり 5 問について作業を行うが、1 タスクで取得できる作業結果はチェック設問の 1 件を除いた 4 件である。本実験では各文について 10 回ずつ機械翻訳適応文作成タスクを行ったため、チェック設問を含まない総設問数は 400 問であり、総タスク数は 100 件となる。クラウドソーシングサービスとして Yahoo!クラウドソーシング*6 を利用した。

原文の対象言語への翻訳には、Google 翻訳*7,*8 を利用した。また、折り返し翻訳に利用する機械翻訳サービスとして、Google 翻訳、Yahoo!翻訳*9、Weblio 翻訳*10、J-Server *11 の 4 つの翻訳機を用いた。なお、Google 翻訳と J-Server は言語グリッド [15] を介して利用している。

なお、機械翻訳適応文の評価は、Walker らの適合性評価 (5 段階評価) [16] を用いた機械翻訳文の評価結果を用いる。これは、日本人大学院生 3 名によって行う。評価段階は、1:まったく違う意味、2:雰囲気は残っているが、もとの意味は分からない、3:意味は何となくつかめる、4:文法などに多少問題はあるが、大体同じ意味、5:同じ意味、である。この評価段階を用いて、原文と原文の機械翻訳による折り返し翻訳文、および、原文と実験で作成する機械翻訳適応文とについて、それぞれ意味の比較を行う。機械翻訳適応文の評価は、本提案手法の評価のために実施しており、本手法の通常の手順では行わない手順である。

機械翻訳適応文作成の目的は、図 1(4)の作業者にとって、機械翻訳文から原文の意味を推測しやすくなるようにすることである。そのため、本実験では、評価段階 3 (意味は何となくつかめる) 以上の評価を得た文を、図 1(4)において作業者が原文の意味を推測しやすい「機械翻訳に適している文」として扱う。評価段階 2 (雰囲気は残っているが、もとの意味は分からない) 以下の文を、原文の意味を推測することが困難な「機械翻訳には適しない文」とする。

また、提案手法では、図 1(3)の機械翻訳文の訂正タスクで 3 種類の機械翻訳文を同時に提示する。そのため、3 種類の文のうち 1 文でも適切に機械翻訳が行われている文が存在すれば、作業者は原文の意味が推測でき、正確な訂正が行えると予想される。そのため、原文 1 文に対して作成される 2 種類の機械翻訳適応文の評価をまとめて、機械翻訳文 (全体) として表 6 に示した。機械翻訳文 (全体) は、2 種類の機械翻訳文それぞれの評価値から、高い方を評価結果として採用している。

*6 <http://crowdsourcing.yahoo.co.jp/>

*7 <https://translate.google.com/>

*8 Google 翻訳は 2016 年 11 月 16 日にニューラルネットに基づく機械翻訳を導入したが、本実験ではそれ以前のバージョンの翻訳結果を利用した。

*9 <http://honyaku.yahoo.co.jp/>

*10 <http://translate.weblio.jp/>

*11 <http://kodensha.jp/>

表 5 機械翻訳文訂正タスクの設定

Table 5 Settings of machine translation sentence correction task.

設定項目	設定値
使用会話文数	40 組
作業対象文数	40 文
1 文当たりのタスク実行回数	10 回
1 タスクあたりの設問数	1 問
チェック設問の有無	無
総設問数	400 問
総タスク数	400 件
作業者の所在地	United States
1 タスクあたりの報酬	2 セント

表 6 機械翻訳文の評価の分布

Table 6 Distribution of the evaluations of machine translation sentences.

文の種類	機械翻訳文 (英) の評価の分布					合計
	1	2	3	4	5	
原文	3	8	5	10	14	40
RIBES の値が最も高い 機械翻訳適応文 (日)	3	2	2	14	19	40
チェックされた中で RIBES の値が最も低い 機械翻訳適応文 (日)	2	4	8	6	20	40
機械翻訳適応文 (全体)	1	0	2	7	30	40

- ・評価段階は、1: まったく違う意味, 2: 雰囲気は残っているが、もとの意味は分からない, 3: 意味は何となくつかめる, 4: 文法などに多少問題はあるが、大体同じ意味, 5: 同じ意味, である。
- ・機械翻訳適応文 (全体) は、1つの原文に対して作成される2種類の機械翻訳適応文 (日) のうち、評価値が高い方を採用したものである。なお、評価値が高い方を採用した理由は、機械翻訳作成の際に、作業者はより参考になる方を参照するので、評価の高い方で考える。

4.2 機械翻訳文の訂正文作成実験

3.4 節で述べたタスクを用いて機械翻訳適応文作成実験を行った。従来手法を用いた機械翻訳文の訂正文作成結果と比較し、提案手法の妥当性の検証を行う。実験には、4.1 節で作成した機械翻訳適応文を用いる。また、表 5 に示すタスクの設定を用い、クラウドソーシングサービスである CrowdFlower^{*12}において、タスクを依頼した。本実験では、従来手法の原文の機械翻訳文のみを提示する場合と、提案手法の3種類の機械翻訳文を提示する場合での機械翻訳文の訂正タスクの結果を比較する。表 5 で示すタスクの設定は1種類のタスクについての設定である。そのため、作業対象文数の40文について、2種類のタスクをそれぞれ実行するため、本実験で取得する訂正文の総数は800文となる。なお、4.1 節の実験では、後述する不真面目な作業者が1名存在した。本実験では、その不真面目な作業

^{*12} <http://crowdfunder.com/>

者の作業結果を除いている。

機械翻訳文の訂正タスクにおいて、従来手法では機械翻訳文の5段階の流暢性評価を指示し、提案手法では3種類の機械翻訳文の中から訂正に最も参考になった文を選択することを指示している。

5. 実験結果と考察

5.1 機械翻訳適応文の評価

機械翻訳適応文 (日) の機械翻訳文 (英) について評価を行うことで、機械翻訳適応文作成手法の妥当性の検証を行う。実験では、英語の翻訳されている機械翻訳文 (英) と機械翻訳適応文 (日) との評価を行う。機械翻訳文 (英) は、機械翻訳適応文 (日) として選択されたものに用いられている機械翻訳サービスによって生成されたものを用いている。

表 6 に、原文 (日) や機械翻訳適応文 (日) それぞれの機械翻訳文 (英) に対する5段階評価の分布を示す。なお、本実験では、3名の評価者 (日本人大学生) に評価を依頼しているため、機械翻訳文 (英) の評価値は3名の評価者による評価の中央値を用いた。ここでは、評価のために「両方の言語に対する知識をある程度持っている人」に評価を依頼しているが、提案手法そのものは、どちらか一方の言語を理解する作業で実施する。

機械翻訳適応文 (日) を利用する目的は、言語 B の作業者が原文の意味を推測できる機械翻訳文 (英) を作成することである。そのため、評価値が3以上 (評価値3: 意味は何となくつかめる) の場合、作業者が原文の意味を推測できる機械翻訳適応文 (日) であると判断できる可能性がある。本実験において評価値が3以上となった文は、原文では40文中29文 (72%)、機械翻訳適応文 (全件) では40文中39文 (97%) だった。

今回は、40文を用いた実験を行った。表 4 に示すように、40文には、主語が含まれていなかったり、オノマトベや比喩が使われていたりするため、必ずしも機械翻訳機が翻訳しやすい文ではない [14]。機械翻訳適応文 (日) の作成においては、文献 [17] から、文字数としては折り返し翻訳で修正できる範囲に含まれている。日本語の原文で30文字程度以下の場合には、本提案手法により、同様の効果が得られると考えられる。

一般に、機械翻訳機の翻訳結果は、文のわずかな違いにより変わる。たとえば、同じ機械翻訳機でも、入力文を少し修正ただけで、精度の高い翻訳文が得られたり、逆に翻訳が適切に行われなかったりする。また、機械翻訳機 A では、慣用句やオノマトベが適切に翻訳できたからといって、別の慣用句やオノマトベの場合うまくいかずに、機械翻訳機 B がうまくいく場合もある。今回、評価が向上した10文に対して、個々の文に提案手法が効果を示しているわけではなく、用いた40文のほとんど (39文) に対し

て、一定程度以上（全体的に評価が向上し、ほとんどの文の評価値が3以上）の高い評価が得られた点が重要であると考えている。

5.2 適切な機械翻訳適応文を取得できなかった例

提案手法は、機械翻訳機の特徴（出力された文の正確性が不明、慣用句やオノマトペ、特殊な言い回し、例外的な使い方、主語のない文への対応が苦手）に対応できると考えているにもかかわらず、評価が低い文が存在した。そこで、どのような問題が発生していたかについて分析する。

表6の機械翻訳適応文（全体）より、原文40文のうち1文のみ、適切な機械翻訳適応文（日）を作成できなかったことが分かった。その原文と作業結果を表7に示す。

作業者の1人は、泣くときに用いる表現である「べそをかきます」を機械翻訳適応文（日）としてタスクを完了した。この「べそをかきます」は、Yahoo!翻訳では翻訳できない文だったため、作業者の入力した機械翻訳適応文（日）と折り返し翻訳文が完全に同じ文となり、RIBESの値が1となっていた。

また、もう1人の作業者は、泣くときの表現としての「しくしく」と近い意味を持つ折り返し翻訳文にチェックを付けていた。これらのことより、この1文については適切な機械翻訳適応文（日）を取得することができなかった。なお、不適切な文に対してチェックをつけたのは、本実験において唯一、1人あたりの作業回数上限の5タスク（計25問）を行っていた作業者だった。また、この作業者は1設問あたりの作業時間の平均が14秒である。このことから、この作業者は不当に報酬を得ようとする不真面目な作業者であると推測できる。この不真面目な作業者による回答を除いたところ、作業者がチェックした機械翻訳適応文（日）のなかで最もRIBESの値が低いのは、「ずきずきするんです」となり、機械翻訳文（英）は“*It throbs.*”となった。“*It throbs.*”という文自体は、「しくしくする痛み」を表現する

表7 適切な機械翻訳適応文（日）を取得できなかった文

Table 7 A sentence that could not obtain a sentence adapted to machine translation.

	RIBESの値が最も高い機械翻訳適応文（日）	チェックされた中でRIBESの値が最も低い機械翻訳適応文（日）
原文	しくしくするんです	
入力文	べそをかきます	しくしくするんです
折り返し翻訳文	べそをかきます	私はむせび泣きます。
翻訳文	べそをかきます*	I sob.
翻訳機	Yahoo!翻訳	J-Server
RIBESの値	1	0

* Yahoo!翻訳を用いた翻訳文が「べそをかきます」という日本語を翻訳できなかったため、翻訳文も日本語の「べそをかきます」となっている。

には不適切だが、“*throb*”，痛みの種類を表す意味もある。そのため、不真面目な作業者による作業結果を用いる場合よりも適切な訂正文を取得できる可能性が高いと考えられる。

これらのことより、翻訳結果がきちんと対象言語に翻訳されているかどうかの確認と、作業者の作業傾向から不真面目な作業者を検出し、その作業者の回答は結果に含めないことが必要であると考えられる。なお、作業傾向として作業時間と作業回数を用いることで、不真面目な作業者を検出できる可能性がある。

5.3 タスクごとの機械翻訳文（英）の評価結果

4.2節における各手法における機械翻訳文（英）の評価結果を表8と表9に示す。表8は、クラウドソーシング上の作業者（英語の単言語話者）による機械翻訳文（英）の流暢性の5段階評価の結果である。表9は、クラウドソーシング上の作業者（英語の単言語話者）に、翻訳対象文40文それぞれにおいて、3種類の機械翻訳文（英）のうち訂正文を作成する際に、訂正に最も参考になった文を選択してもらった結果である。

提案手法において原文の機械翻訳が最も参考になったと回答されたのは10文だった。従来手法で流暢性の評価が5であり、提案手法で原文の機械翻訳が最も参考になったと評価された原文の数は7文だった。この7文については、原文が適切に機械翻訳を行える文であったと推測できる。

しかし、従来手法では機械翻訳文（英）の評価値は低い

表8 従来手法における機械翻訳文（英）の評価結果

Table 8 Evaluation results of machine translation sentences in the conventional method.

評価値	評価された機械翻訳文の数
1	4文
2	14文
3	7文
4	3文
5	12文

・評価段階は、1：まったく違う意味、2：雰囲気は残っているが、もとの意味は分からない、3：意味は何となくつかめる、4：文法などに多少問題はあがるが、大体同じ意味、5：同じ意味、である。

表9 提案手法における機械翻訳文（英）の評価結果

Table 9 Evaluation results of machine translation sentences in the proposed method.

最も参考になった機械翻訳文	該当する原文の数
原文	10文
RIBESの値が最も高い機械翻訳適応文（日）	20文
チェックされた中でRIBESの値が最も低い機械翻訳適応文（日）	10文

表 10 「訂正が困難である」という意味のコメントが入力された訂正文の数

Table 10 The number of corrections that have been entered with comments that mean “difficult to correct.”

手法	「訂正が困難」という意味のコメントの数
従来手法	24 文/400 文 (6%)
提案手法	0 文/400 文 (0%)

が、提案手法では、原文の機械翻訳文（英）が最も参考になったと判断された文が 2 文存在した（機械翻訳文（英）自体には誤りがあるが、原文の意味を推測できる程度である文）。2 文のうち 1 文は、従来手法において評価値 4 を得た機械翻訳文（英）である。そのため、この 1 文については、ある程度の精度を保った機械翻訳が行える原文であったと推測できる。

残りの 1 文は、従来手法における原文の機械翻訳文（英）の評価値が 3 にもかかわらず、提案手法では、原文の機械翻訳文（英）が最も参考になったと評価されていた。これは、この 1 文が 3 種類の機械翻訳文（英）がすべて同じ文だったためだと考えられる。また、この 1 文は “It is a thing of the order to carry out the inspection smoothly.” という、不適切な機械翻訳が行われた文であった。

本論文での提案では、異なる 3 種類の機械翻訳文（英）を提示することで機械翻訳文（英）の訂正精度向上を目指している。そのため、このように 3 種類の機械翻訳文（英）が同じ文になる場合には、他の機械翻訳を利用した機械翻訳文（英）を提示することや、他の機械翻訳適応文（日）を用いることを検討する必要がある。

5.4 作成された訂正文の考察

従来の機械翻訳文（英）の訂正タスクでは、作業者が機械翻訳文（英）の意味を推測できなかった場合に、訂正文ではなく “I have no idea.” のようなコメントが入力されることがあった。また、本実験においても、従来手法において取得した訂正文には、“It is not possible to tell what sentence 2 means, it is utterly incomprehensible.” や、“It takes no sense.” といったコメントがあった。

従来手法と提案手法それぞれにおける、このような「訂正が困難である」という意味のコメント数を表 10 に示す。「訂正が困難」という意味のコメントの数は、従来手法ではのべ 400 文中 24 文であった。なお、作業対象とした 40 文のうち、従来手法において「訂正が困難」という意味のコメントがされた文は 13 文だった。表 10 に示すように提案手法では、「訂正が困難である」といった意味のコメントを入力した作業者はいなかった。これは、機械翻訳適応文（日）を用い、複数の機械翻訳文（英）を提示した提案手法では、作業者に原文の意味が推測しやすくなり、適切な訂正文が作成されている可能性が高いと考えられる。

表 11 機械翻訳適応文（日）の作成に各翻訳機が利用された回数
Table 11 Number of times a sentence adapted for machine translation was used by each translation machine.

翻訳機	RIBES の値が最も高い	チェックされた中で RIBES の値が最も低い
Google 翻訳	23	14
Yahoo!翻訳	9	18
Weblio 翻訳	0	2
J-Server	8	6
合計	40	40

5.5 必要な機械翻訳サービスの数

表 11 に、実験で取得した機械翻訳適応文（日）がどの翻訳機を用いていたかを示す。RIBES の値が高い文と低い文それぞれについて、各翻訳機が選択された文の回数を示している。なお、各翻訳機で同じ RIBES の値であった場合には、表の順番を優先して、1 つのみを選択するようにした。たとえば、Google 翻訳と Yahoo!翻訳で同じ RIBES の値だった場合には、機械翻訳適応文（日）として Google 翻訳を用いた場合の翻訳結果が採用される。

表 11 より、RIBES の値が高い文の場合には Google 翻訳が多く用いられているが、作業者がチェックをつけた機械翻訳適応文（日）の中で最も RIBES の値が低い文では Yahoo!翻訳が多く用いられている。今回の実験では、機械翻訳機は、3 種類あれば、今回の提案手法は十分な効果が得られることが分かった。

これらのことから、複数の機械翻訳を用いて機械翻訳適応文（日）作成を行うことで、原文の意味を推測しやすい機械翻訳文（英）を作成することができ、従来手法よりも訂正精度が向上する可能性があると考えられる。

6. おわりに

本論文では、従来の用例対訳作成手法の課題である低精度の機械翻訳に対応するために、複数の翻訳機を用いた用例対訳作成手法を提案した。また、提案手法の妥当性を検証するために、機械翻訳適応文作成実験や機械翻訳文の訂正文作成実験を行った。

本研究の貢献は以下の 3 点である。

- (1) 複数の機械翻訳を用いた機械翻訳適応文作成手法を提案した。作業者に複数の折り返し翻訳文を提示することで、折り返し翻訳の精度不一致による低精度な機械翻訳適応文作成を防げる可能性を示した。
- (2) 機械翻訳適応文の評価において、人手による機械翻訳適応文の評価と、計算機による折り返し翻訳文との一致度を組み合わせる手法によって、機械翻訳文の評価の高い機械翻訳適応文の抽出が行えることを示した。
- (3) 提案手法により、機械翻訳文の訂正タスクを行う作業者が原文の意味を推測しやすい機械翻訳文を作成可能な機械翻訳適応文作成が行えることを明らかにした。

今後は、提案手法を用いた用例対訳作成サービスの実現について検討する。

参考文献

- [1] 日本政府観光局 (JNTO): 訪日外客数, 統計データ, 日本政府観光局 (オンライン), 入手先 (http://www.jnto.go.jp/jpn/statistics/since2003/_tourists.pdf) (参照 2017-01-20).
- [2] 国土交通省観光庁: 明日の日本を支える観光ビジョン, 国土交通省観光庁 (オンライン), 入手先 (<http://www.mlit.go.jp/common/001126598.pdf>) (参照 2017-01-20).
- [3] 福島 拓, 吉野 孝, 重野亜久里: 正確な情報共有のための多言語用例対訳共有システム, 情報処理学会論文誌, コンシューマ・デバイス&システム, Vol.2, No.3, pp.23–33 (2012).
- [4] 山本里美, 福島 拓, 吉野 孝: クラウドソーシングにおける会話をういた応用例対作成手法の提案, 情報処理学会論文誌, Vol.56, No.3, pp.1080–1089 (2015).
- [5] 山本里美, 福島 拓, 吉野 孝: 機械翻訳適応文を用いた単言語話者による機械翻訳文の評価と訂正, 2016 年度情報処理学会関西支部支部大会, No.G-22, pp.1–4 (2016).
- [6] 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).
- [7] Zaidan, O.F. and Callison-Burch, C.: Crowdsourcing Translation: Professional Quality from on-Professionals, *Proc. 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Vol.1, pp.1220–1229 (2011).
- [8] Papineni, K., Roukos, S., Ward, T. and Zhu, W.-J.: BLEU: A Method for Automatic Evaluation of Machine Translation, *ACL '02*, pp.311–318, Association for Computational Linguistics (2002).
- [9] Callison-Burch, C.: Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk, *EMNLP '09*, pp.286–295, Association for Computational Linguistics (online) (2009).
- [10] Hu, C.: Crowdsourced Monolingual Translation, PhD Thesis, College Park, MD, USA (2012).
- [11] Mitchell, L., O'Brien, S. and Roturier, J.: Quality Evaluation in Community Post-editing, *Machine Translation*, Vol.28, No.3-4, pp.237–262 (2014).
- [12] Green, S., Heer, J. and Manning, C.D.: The Efficacy of Human Post-editing for Language Translation, *Proc. SIGCHI Conference on Human Factors in Computing Systems, CHI '13*, pp.439–448, ACM (2013).
- [13] 平尾 努, 磯崎秀樹, 須藤克仁, Duh, K., 塚田 元, 永田昌明: 語順の相関に基づく機械翻訳の自動評価法, 自然言語処理, Vol.21, No.3, pp.421–444 (2014).
- [14] 長瀬久明, 橘 守亮: 機械翻訳に適応した日本語文の書き方 (2010), 入手先 (<http://hdl.handle.net/10132/3569>).
- [15] Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration, *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp.96–100 (2006).
- [16] Walker, K. et al.: Multiple-Translation Arabic (MTA) Part 1, *Linguistic Data Consortium*, Philadelphia (2003).
- [17] 宮部真衣, 吉野 孝, 重信智宏: 折返し翻訳を用いた翻訳リペアの効果, 電子情報通信学会論文誌 D, Vol.J90-D, No.12, pp.3141–3150 (2007).



山本 里美

1992 年生。2015 年和歌山大学システム工学部卒業。2017 年同大学大学院システム工学研究科博士前期課程修了。在学中、多言語用例対訳収集に関する研究に従事。



福島 拓 (正会員)

1986 年生。2008 年和歌山大学システム工学部中退。2013 年同大学大学院システム工学研究科博士後期課程修了。博士 (工学)。現在、大阪工業大学情報科学部特任講師。CSCW の研究に従事。



吉野 孝 (正会員)

1969 年生。1992 年鹿児島大学工学部卒業。1994 年同大学大学院工学研究科修士課程修了。博士 (情報科学)。現在、和歌山大学システム工学部教授。CSCW, HCI, コミュニケーション支援の研究に従事。