

# クラウドソーシングを用いた機械翻訳適応文作成における 複数翻訳機の効果の検証

山本 里美<sup>1</sup> 福島 拓<sup>2</sup> 吉野 孝<sup>3</sup>

**概要**：クラウドソーシングを用いた多言語データの収集の研究は多く行われている。現在、我々は、クラウドソーシング上の単言語話者を用いた用例対訳作成手法に関する研究を行っている。用例対訳とは、医療分野などの正確な情報の共有が求められる現場における、多言語間コミュニケーションに用いられる多言語の対のことである。これまでの実験により、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、単言語話者であっても一部の用例について正確な対訳作成が行えることがわかった。しかし、翻訳対象の用例によって機械翻訳の精度に差があり、すべての用例について対訳作成を行うことは困難である。そこで、低精度な機械翻訳文への対応として、原文をもとに適切な機械翻訳が行えるような機械翻訳適応文を作成し、クラウドソーシングを用いた用例対訳作成に用いることによって訂正精度を向上させる手法を提案した。機械翻訳適応文を用いた用例対訳作成実験において訂正精度向上の効果は見られなかった。これは、翻訳文と折り返し翻訳文での翻訳精度に不一致があったためであると考えられる。本稿では、この翻訳精度の不一致に対応するために、機械翻訳適応文作成タスクにおいて複数の翻訳機を用いる手法を提案し、複数の翻訳機を用いる効果の検証実験を行った。検証の結果、機械翻訳適応文作成タスクに複数の翻訳機を用いることで、約97%の原文において、原文の意味を推測しやすい機械翻訳文が作成できることを明らかにした。

**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、折り返し翻訳

## 1. はじめに

世界的なグローバル化によって多言語間コミュニケーションの機会は増加している。インターネットを用いた多言語間コミュニケーション支援の方法として、機械翻訳や多言語用例対訳の利用などが挙げられる。多言語用例対訳とは、正確性の確保された多言語の対のことで、正確な情報の共有が重要となる医療分野などで利用される。しかし、正確な情報の共有のために必要とされる用例対訳の数は多く、十分な数の用例対訳を収集することは困難である [1]。

現在、我々はクラウドソーシングと機械翻訳を用いることで、単言語話者であっても用例対訳が作成できる手法に関する研究を行っている。これまでの実験より、クラウドソーシング上の単言語話者に対して、翻訳対象

の用例を含む会話文の機械翻訳文や、その用例の使用者の属性情報<sup>\*1</sup>を提示することで、一部の用例において用例対訳や応答用例対<sup>\*2</sup>を作成できることを明らかにした [2]。この手法では、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を取得する。しかし、用例によって機械翻訳の精度は異なり、低精度の機械翻訳の場合はクラウドソーシングの作業者は原文の意味を推測できず、適切な訂正が行えないという課題が存在した。

そこで、原文を精度の高い機械翻訳が行えるような文(機械翻訳適応文)に書き換えることで、訂正精度を向上させる手法を提案した [3]。機械翻訳適応文とは、原文と折り返し翻訳文との一致度が高くなるような文のことである。クラウドソーシングの作業者が入力文の折り返し翻訳文を確認しながら、入力文の修正をしていくことで折り返し翻訳文との一致度が向上するような文を作成する。しかし、文献 [3] における機械翻訳適応文を用いた用例対訳作成実験の結果、機械翻訳文の訂正精度向上への効果がないことがわかった。分析の結果より、機械翻訳文と折り返し翻訳文での翻訳精度の不一致が影響して

<sup>1</sup> 和歌山大学大学院システム工学研究科  
Graduate School of System Engineering, Wakayama University Sakaedani 930, Wakayama, 640-8510 Japan

<sup>2</sup> 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

<sup>3</sup> 和歌山大学システム工学部  
Faculty of Systems Engineering, Wakayama University Sakaedani 930, Wakayama, 640-8510 Japan

<sup>\*1</sup> 医療従事者や病院利用者のこと。

<sup>\*2</sup> 質問とその回答の対、またその類似文からなる用例対訳のことであり、会話の支援に用いられる。

いる可能性がある。例えば、「今主人がこちらに向かっているの、着いたら払ってもらいます。」という原文に対して作成された機械翻訳適応文「現在主人がこちらに向かっている為、彼が到着した後、彼がその支払いをします。」は、折り返し翻訳文との一致度は高いが、英語に機械翻訳する場合“Since the current master is heading here,after he arrived, he will make that payment.”という結果になり、原文とは意味が異なっている。

本稿では、このような機械翻訳の精度の不一致に対応するために、機械翻訳適応文作成において複数の翻訳機を用いる手法を提案し、機械翻訳適応文作成タスクに複数の翻訳機を用いることの効果について検証を行った。

## 2. 関連研究

クラウドソーシングを用いることで非専門家による多言語データの作成や収集を行う研究は多く行われている。Omar らによる、翻訳対象文やクラウドソーシング作業者の特徴をもとに翻訳文を分析し、品質の良い翻訳文を作成する研究 [4] や、複数の属性の作業者を組み合わせによって精度の高い翻訳文を作成する研究 [5] などがある。これらの研究により、クラウドソーシングを用いることで、専門家による翻訳の質に近い翻訳文を作成できることが分かったが、これらの研究では、クラウドソーシングの作業者が2言語以上を理解できることを前提としている。しかし、クラウドソーシング上の多言語話者の数は少なく、特に、話者の少ない言語を理解できる多言語話者をクラウドソーシング上で確保することは困難である。

クラウドソーシング上の単言語話者を対象にした多言語データ収集の研究に、児童向け図書の翻訳を目的とした、Chan らによる“Crowdsourced Monolingual Translation”がある [6]。これは、翻訳対象文の言語を母語とする作業者と翻訳対象言語を母語とする作業者が、文への注釈付けや機械翻訳文の修正を繰り返し行うことで翻訳文の作成を目指す手法である。複数の単言語話者によって繰り返し注釈付けと機械翻訳文の修正を行うことで、1人の多言語話者による修正と同程度の質の翻訳文を取得することができる。しかし、この手法では、翻訳対象文と同じ意味を翻訳文が示すかどうかの評価をしておらず、正確性の確保が必要とされる用例対訳作成に用いることは困難である。

また、宮部らは機械翻訳を用いたコミュニケーションにおいて複数の翻訳機による折り返し翻訳文を用いることで、翻訳精度の不一致を検出する手法を提案した [7]。文献 [7] では、折り返し翻訳文と原文の一致度は高いが、対象言語翻訳文の精度が低い場合に起こる翻訳精度の不一致の検出手法を提案している。この手法では、まず、折り返し翻訳文と原文について人手による一致度評価を行う。一致度が高い場合は1度目とは別の翻訳機を用い

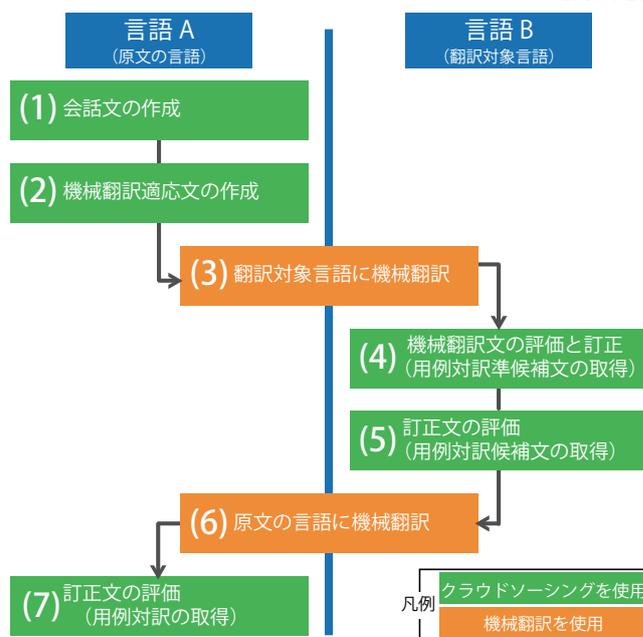


図 1 クラウドソーシングを用いた用例対訳作成の流れ

て対象言語翻訳文への折り返し翻訳を行い、複数の折り返し翻訳文から採用する折り返し翻訳文を抽出することで翻訳精度の不一致検出を行う。この手法により、翻訳精度の不一致の7割を検出できることを明らかにした。

## 3. クラウドソーシングを用いた用例対訳作成手法

本章では、我々が提案しているクラウドソーシング上の単言語を用いた用例対訳作成手法について述べる。クラウドソーシング上の単言語話者を用いた用例対訳作成手法について述べる。図 1 に、用例対訳作成の大まかな流れを示す。本手法では、原文の言語(言語 A)を母語とする作業者(グループ A)と、翻訳対象言語(言語 B)を母語とする作業者(グループ B)の2つの作業者グループによって用例対訳作成を行う。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。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)で作成された会話文を「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。

本稿では、図 1(2)の機械翻訳適応文作成に複数の翻訳機を用いて用例対訳作成実験を行い、複数の翻訳機による効果の検証を行う。

## 4. 機械翻訳適応文作成タスク

### 4.1 タスクの概要

機械翻訳適応文作成タスクでは、グループ A(原文の

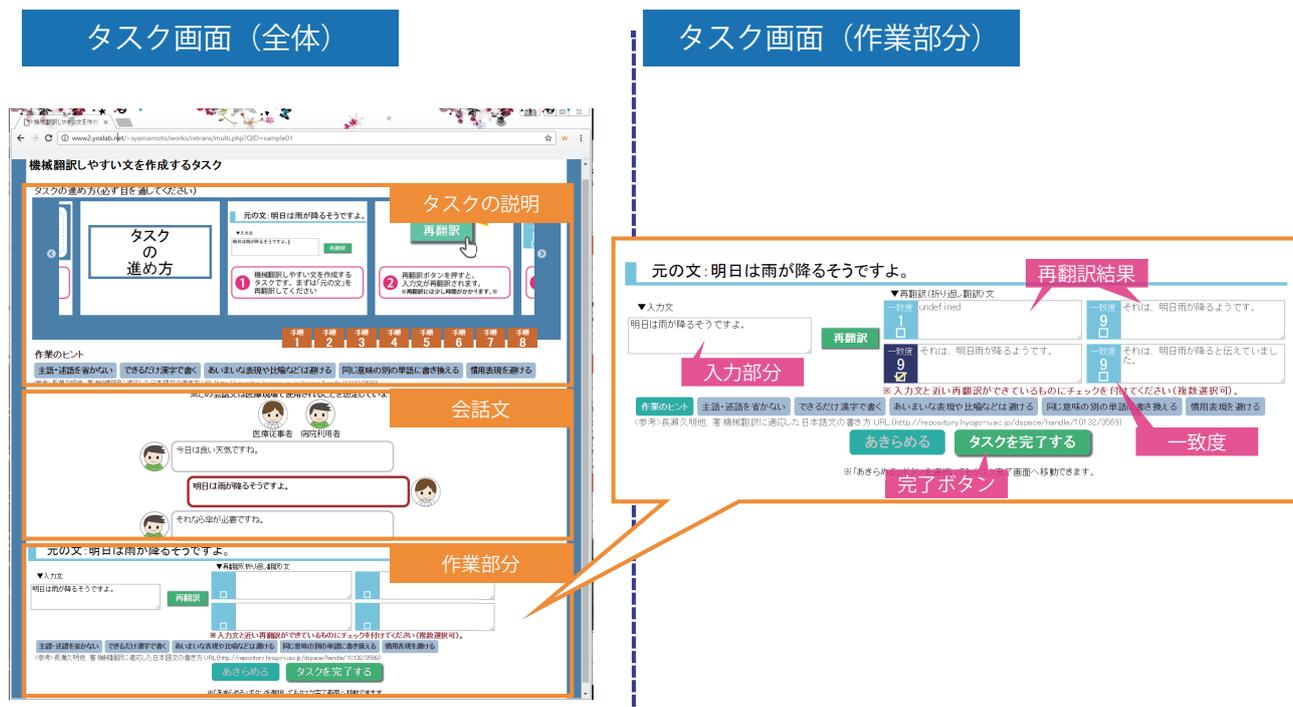


図 2 機械翻訳適応文作成タスク画面例

言語を母語とする作業グループ) が作業を行う。作業者は、提示される原文の会話文の 2 文目に対して機械翻訳適応文の作成を行う。機械翻訳適応文とは、折り返し翻訳文との一致度が高い文のことであり、適切な機械翻訳が行われる可能性が高いと考えられる。

#### 4.2 タスクの流れ

機械翻訳適応文作成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」をクリックすることで再翻訳する。
- (2) 4 種類の再翻訳結果 (再翻訳文と入力文との一致度) を確認する。
- (3) 再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、再翻訳結果を確認する。
- (5) 4 つの再翻訳文のうち 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 つの再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。

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。

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 [8] をいくつか提示した\*3。

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[9] の値を 10 段階で表したものである。RIBES は、機械翻訳の結果を自動的に評価するための尺度のひとつであり、人間による評価と相関が高い評価値を算出することが可能である。折り返し翻訳文との RIBES 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、折り返し翻訳文との RIBES の値が高い文であっても、文献 [7] で述べられている翻訳精度の不一致を考慮する必要がある。本稿では、機械翻訳適応文作成タスクに複数の翻訳機を用いることで、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。このことで、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。

また、作業者は提示される複数の再翻訳結果のうち、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

\*3 作業には、「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 [8] を参考にした。

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。

### 4.3 機械翻訳文の評価と訂正に用いる文の抽出

本節では、図 1(4) の機械翻訳文の評価と訂正に用いるために図 1(3) で機械翻訳を行う文を抽出する方法について述べる。

機械翻訳適応文の利用を提案した文献 [3] の実験では、1 つの機械翻訳サービス<sup>\*4</sup>を用いて作成した機械翻訳適応文を用いて、図 1(4) の対準候補文作成実験を行った。この実験では、図 1(3) で機械翻訳する文として、図 1(2) で作成された機械翻訳適応文の中で最も RIBES の値が高いものを採用した。そのため、文献 [3] の実験では折り返し翻訳における翻訳精度の不一致を考慮していない。本稿では、翻訳精度の不一致に対応するために機械翻訳適応文作成に複数の翻訳機を用いることを提案している。

4.1 節で述べたタスクにおいて、作業には、RIBES の値に関係なく、主観的に入力文と折り返し翻訳文の意味が同じと感じる折り返し翻訳文を選択することも依頼している。これにより、1 つの翻訳機のみを用いていた場合と異なり、RIBES の値以外での機械翻訳適応文に関する評価を得ることができる。

そこで、今後は、図 1(4) で作業者に提示する機械翻訳文について単純に RIBES の値が高い機械翻訳適応文の機械翻訳文を提示するだけでなく、RIBES の値に関係なく作業者が適切だと評価した機械翻訳適応文も用いることを考えている。これまでの実験により、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 [2]。そのため、機械翻訳文の訂正を行う際に数種類の機械翻訳文を提示することで、作業者がより原文の意味を推測しやすくなり、適切な訂正が行える可能性がある。

図 1(4) で作業者に提示する機械翻訳文として、以下の 3 種類を考えている。

- 原文の機械翻訳文
- 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 4.2 節の手順 (6) でチェックを付けた機械翻訳適応文の中で最も RIBES の値が低い機械翻訳適応文の機械翻訳文

作業者が 4.2 節の手順 (6) でチェックを付けた機械翻訳適応文は、RIBES の値に関係なく、入力文とその折り返し翻訳文の意味が似ている文として扱える可能性が高い。チェックを付けられた機械翻訳適応文のうち RIBES の値が最も低い文を用いるのは、RIBES の値が最も高い機械翻訳適応文とは異なる文を機械的に取得するためである。

表 1 機械翻訳適応文作成タスクの設定

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 <sup>*5</sup>

<sup>\*5</sup> 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、カルチュア・コンビニエンス・クラブが展開するポイントサービスである。

## 5. 機械翻訳適応文作成実験

### 5.1 実験用データセット

本節では、実験に用いるデータセットについて述べる。本実験では、クラウドソーシングを用いて作成した日本語の会話文 40 組を用いた。会話文は 3 文からなり、1 文目は疑問文、2 文目は 1 文目に対する回答、3 文目は 2 文目に続く応答となっている。なお、会話文作成を行う作業には、医療現場における会話を想定して作成することを指示した。会話文には、医療従事者が質問をする場合と病院利用者が質問をする場合の 2 通りあり、本実験で用いた 40 組のうち 22 組が、医療従事者が質問をする会話文である。

### 5.2 機械翻訳適応文作成タスクの設定

本稿では、5.1 節で述べた 40 組の会話文の 2 文目を対象とし実験を行った。表 1 にタスクの設定を示す。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、節で述べた 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利用したのは、Yahoo!クラウドソーシング<sup>\*6</sup>であり、作業には、Yahoo!クラウドソーシングのタスク画面から、図 2 の折り返し翻訳機能を持つ機械翻訳適応文タスク作成用 Web ページへ移動し、タスクを行うように指示した。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<sup>\*7</sup>、Yahoo!翻

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

<sup>\*7</sup> 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。

<sup>\*4</sup> Google 翻訳 (<https://translate.google.com/>)

訳<sup>\*8</sup>, Weblio 翻訳<sup>\*9</sup>, J-Server<sup>\*10</sup>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 [10] を介して利用している。

### 5.3 機械翻訳適応文の評価方法

原文の機械翻訳文と、実験で作成する機械翻訳適応文の機械翻訳文について主観評価を行い、比較を行う。比較に用いる機械翻訳適応文とは、4.3 節で述べる抽出手法を用いて、各原文において取得する2種類の機械翻訳適応文のことである。主観評価には Walker らの適合性評価 (5段階評価)[11]を用いる。評価段階は、1: 全く違う意味, 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: 文法などに他称問題はあがるが、大体同じ意味, 5: 同じ意味, である。この評価段階を用いて、原文の機械翻訳文と実験で作成する機械翻訳適応文それぞれの機械翻訳文について、原文との意味の比較を行う。本実験における評価者は、日本人大学院生3名である。

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、本稿では、機械翻訳文の意味が推測できると考えられる、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」として扱い、2(雰囲気は残っているが、もとの意味は分からない)未満の文を、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。

## 6. 実験結果と考察

### 6.1 クラウドソーシングの作業者

本実験で機械翻訳適応文作成タスクに参加した作業者の総数は82名であり、総作業件数は534件だった。なお、この534件には、チェック設問に対する回答も含む。また、チェック設問に対する作業件数を除いた総作業件数は468件だった。本稿では、この468件における回答を用いて検証を行う。

作業者総数の82名とは、1件でも回答作業を行った作業者の数である。82名中、タスク<sup>\*11</sup>を完了させた作業者は69名だった。タスクを中断した作業者もいたため、5.2 節で述べた総設問数よりも多い作業結果を取得した。本稿では、作業者がタスクを完了しているかどうかにかかわらず、すべての作業結果を用いて検証を行う。

なお、作業者が1問あたりにかけた作業時間<sup>\*12</sup>の平均は101秒だった。なお、標準偏差は174秒であり、最も

表2 機械翻訳文の評価の分布

文の種類	機械翻訳文の評価の分布					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 <sup>*13</sup>	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 <sup>*14</sup>	2	4	8	6	20	40
機械翻訳適応文(全体)	1	0	2	7	30	40

\*13 機械翻訳文(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のうち、評価値が高い方を採用したものである。

長い作業時間は32分、最も短い作業時間は4秒と、作業者によって作業時間は大きく異なっていた。また、折り返し翻訳を1度だけ行っている作業は534件中234件あった。これらのことから、不真面目な作業者の存在があったことが考えられる。しかし、本実験において使用した原文40文すべてにおいて、原文と原文の折り返し翻訳文とのRIBESの値よりも、作成された機械翻訳適応文とその折り返し翻訳文とのRIBESの値の方が高いという結果を得た。このことから、5.2 節の設定を用いたタスクにおいて、作業対象の原文すべてにおいて適切な作業結果を取得できたと考えられる。

### 6.2 機械翻訳適応文の評価

5.3 節で述べた評価方法による機械翻訳適応文の評価結果について述べる。表2に、原文や機械翻訳適応文それぞれの機械翻訳文に対する5段階評価の分布を示す。なお、本実験では、3名の評価者に評価を依頼しているため、機械翻訳文の評価値は3名の評価者による評価の中央値を用いた。また、4.3 節で述べたように、図1(4)では、作業者に原文の機械翻訳文と2種類の機械翻訳適応文それぞれの機械翻訳文を同時に提示して作業を依頼する。そのため、作成された機械翻訳適応文のうちどちらか1文でも、作業者に意味が通じる機械翻訳文が作成されていれば良いと考えられる。そこで、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、機械翻訳適応文(全体)として表2に示した。機械翻訳適応文(全体)では、2種類の機械翻訳適応文の機械翻訳文において、評価値が高い方を評価結果として採用している。

また、機械翻訳文の評価結果をもとに、機械翻訳適応文の評価を行った。表3に、機械翻訳適応文の評価結果を示す。5.3 節で述べる評価方法より、本稿では、評価値が3以上(評価値3: 意味は何となくつかめる)の文を機械翻訳に適していると判断し、図1(4)で作業者が機械翻

\*8 <http://honyaku.yahoo.co.jp/>

\*9 <http://translate.weblio.jp/>

\*10 <http://kodensha.jp/>

\*11 チェック設問を含む5問への回答で1タスクとしている。

\*12 タスク用Webページにアクセスしてからタスク完了ボタンを押すまでの時間

\*13 折り返し翻訳文とのRIBESの値が最も高い機械翻訳適応文のこと。

\*14 折り返し翻訳文とのRIBESの値は低いですが、作業者による主観評価によって同じような意味を示すと判断された機械翻訳適応文のこと。

表 3 機械翻訳適応文の評価結果

	機械翻訳文から 原文の意味が推測できない (機械翻訳文の評価値が 2 以下)	機械翻訳文から 原文の意味が推測できる (機械翻訳文の評価値が 3 以上)	適切な機械翻訳が行える (機械翻訳文の評価値が 5)
原文	11 文 (27%)	29 文 (72%)	14 文 (35%)
機械翻訳適応文 (全体)	1 文 (2%)	39 文 (97%)	30 文 (75%)

訳文から原文の意味を推測できる可能性があるとしている。本実験において評価値が 3 以上<sup>\*15</sup>となった文は、原文では 40 文中 29 文、折り返し翻訳文との RIBES の値が最も高い機械翻訳適応文では 40 文中 35 文、作業者がチェックを付けた中で最も折り返し翻訳文との RIBES の値が低い機械翻訳文では 40 文中 34 文、機械翻訳適応文 (全体) では、40 文中 39 文だった。よって、表 3 より、機械翻訳文から原文の意味が推測できる文は、原文では 40 文中 29 文 (72%) だが、機械翻訳適応文を用いる場合では 40 文中 39 文 (97%) となる。

このことから、4 章で述べた、複数の翻訳機を用いた機械翻訳適応文作成を行うことは、従来の課題であった翻訳精度の不一致に対応し、従来手法よりも機械翻訳文の訂正精度を向上させる可能性がある。

また、機械翻訳文の評価値が 5 (評価値 5: 同じ意味) となった機械翻訳適応文は、原文では 14 文 (35%)、機械翻訳適応文 (全体) では、30 文 (75%) となった。このことから、本稿で提案した、複数の翻訳機を用いた機械翻訳適応文作成手法では、図 1(4) の機械翻訳文の評価において高い評価を得る文を作成できる可能性がある。図 1(4) において、機械翻訳文の評価値が高い場合には、機械翻訳文をそのまま対訳候補文として利用することが可能なため、図 1(5) の訂正文の評価にかかるコストを削減することが可能である。

### 6.3 適切な機械翻訳適応文を取得できなかった例

表 2 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を作成できなかったことがわかった。その原文と作業結果を表 4 に示す。表 4 の原文は、医療従事者からの「どんな痛みですか?」という質問で始まる、「しくしくするんです」「じゃあみてみましょう」という会話文のものである。そのため、原文の「しくしく」は痛みの種類を表現している。本実験では、「しくしく」を泣くときの表現として扱った作業者が 12 名中 2 名存在した。

作業者の 1 人は、泣く時に用いる表現である「べそをかきます」を機械翻訳適応文としてタスクを行った。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と折り返し翻訳文が完全に同じ文となり、RIBES の値が 1 になっていた。

また、もう 1 人の作業者が、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ため、適切な機械翻訳適応文を取得することができなかった。この不適切な作業を行ったのは、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である。また、この作業者は 1 設問あたりの作業時間の平均が 14 秒だった。この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、図 1 で機械翻訳文の訂正を行う際に、「I sob」と翻訳されてしまう「しくしくするんです」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。

これらのことから、図 1(3) に使用する機械翻訳適応文の抽出を行う際、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。また、作業傾向として作業時間と作業回数を用いることで、不真面目な作業者は検出できる可能性がある。

### 6.4 翻訳機による比較

表 5 に、実験で取得した機械翻訳適応文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、表で示す順番に優先して、1 つのみを選択するようにした。例えば、Google 翻訳と Yahoo!翻訳で同じ RIBES の値だった場合には、機械翻訳適応文として Google 翻訳を用いた場合の翻訳結果が採用される。

表 5 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。なお、RIBES の値が高い文と、作業者がチェックをつけた中で最も RIBES の値が低い文で、入力文が同じだった文は 40 文中 6 文あった。しかし、同じ翻訳機を使っているものはそのうちの 1 文のみだった。他の 5 文では、RIBES の値が高い文で用いている翻訳機と作業者が選択した文で用いている翻訳機は異なっていた。このことから、複数翻訳機

\*15 機械翻訳に適していると判断された文のこと。

表 4 適切な機械翻訳適応文を取得できなかった文

	RIBES の値が最も高い 機械翻訳適応文*13	チェックされた中で RIBES の値が最も低い 機械翻訳適応文*14
原文		しくしくするんです
入力文	べそかきます	しくしくするんです
再翻訳文	べそかきます	私はむせび泣きます。
翻訳文	べそかきます*16	I sob.
翻訳機	Yahoo!翻訳	J-Server
RIBES の値	1	0

\*16 Yahoo!翻訳を用いた翻訳文が「べそかきます」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「べそかきます」となっている。

表 5 取得した機械翻訳適応文が各翻訳機を利用した回数

翻訳機	RIBES の値が 最も高い*13	チェックされた中で RIBES の値が最も低い*14
Google 翻訳	23	14
Yahoo!翻訳	9	18
Weblio 翻訳	0	2
J-Server	8	6
合計	40	40

を用いて機械翻訳適応文作成を行い、4.3 節で述べた方法で機械翻訳適応文を取得することで、用例対訳作成における図 1(4) で作業者に提示する情報を増やすことが可能となる。これは、機械翻訳文の訂正精度の向上につながる可能性がある。

## 7. おわりに

本稿では、クラウドソーシング上の単言語を用いた用例対訳作成手法について、従来手法における機械翻訳精度の不一致に対応するために、機械翻訳適応文作成において、作業者に複数の折り返し翻訳文を提示する手法を提案し、複数の翻訳機を利用する効果の検証を行った。実験より、機械翻訳適応文作成タスクにおいて複数の翻訳機を利用することで、97%の原文において、機械翻訳に適した機械翻訳適応文を取得できることを明らかにし、機械翻訳文の訂正精度向上につながる可能性を示した。

本稿では、検証実験により以下のことを明らかにした。

- (1) 機械翻訳適応文作成タスクに複数の翻訳機を用いることで、1つの翻訳機を用いる場合よりも多く、適切な機械翻訳適応文を取得できる。
- (2) 人手による主観評価を用いた機械翻訳文適応文の評価\*17を行うことで、折り返し翻訳文における翻訳精度の不一致に対応できる可能性がある。
- (3) 作業傾向によって抽出できる不真面目な作業者による作業結果を除去して、機械翻訳適応文の抽出を行うことで、すべての原文において適切な機械翻訳適応文を作成できる可能性がある。
- (4) 機械翻訳適応文の抽出結果に影響を与えるような不真面目な作業者は、作業時間や作業回数より検出で

\*17 折り返し翻訳文についてチェック機能を用いて評価を行うこと。

きる可能性がある。

今後は、本実験で取得した機械翻訳適応文を用い、用例対訳作成手法における図 1 での用例対訳準候補文作成実験を行い、4.3 節で述べた 3 種類の機械翻訳文を提示することで、クラウドソーシング上の作業者による機械翻訳文の訂正精度が向上するか確認を行う。

## 参考文献

- [1] 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).
- [2] 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文を用いた応答用例対訳作成手法の提案, 情報処理学会論文誌, Vol. 56, No. 3, pp. 1080-1089 (2015).
- [3] 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).
- [4] Omar F. Zaidan and Chris Callison-Burch: Crowdsourcing Translation: Professional Quality from on-Professionals, '11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Vol. 1, pp. 1220-1229 (2011).
- [5] Vamshi Ambati, Stephan Vogel, Jaime Carbonell: Collaborative Workflow for Crowdsourcing Translation, CSCW '12 Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, pp. 1191-1194 (2012).
- [6] Chang Hu, Philip Resnik, and Benjamin B. Bederson.: Crowdsourced Monolingual Translation, ACM Trans. Comput.-Hum. Interact., vol. 21, No. 4, pp. 1-35 (2014).
- [7] 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).
- [8] 長瀬 久明, 橘 守亮: 機械翻訳に適応した日本語文の書き方, (<http://hdl.handle.net/10132/3569>) (2010).
- [9] 平尾 努, 磯崎 秀樹, 須藤 克仁, Kevin Duh, 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).
- [10] Toru Ishida: Language Grid: An Infrastructure for Intercultural Collaboration, IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06), pp. 96-100 (2006).
- [11] Walker, K. et al: Multiple-Translation Arabic(MTA) Part 1, Linguistic Data Consortium, Philadelphia(2003).