

クラウドソーシングを用いた応答用例対作成手法の提案

山本 里美¹ 福島 拓² 吉野 孝¹

概要: 現在、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が増加している。しかし、多言語間での正確な共有は困難であり、医療分野などの正確な情報の共有が求められる場合には十分な正確性の確保された用例対訳が使用されている。しかし、必要な用例対訳は多く、十分な数の用例対訳の収集は困難である。現在我々は、機械翻訳とクラウドソーシングを用いて誤った機械翻訳が行われた用例の訂正を依頼することで、単言語話者でも正確な用例対訳を作成が可能となる用例対訳作成手法の研究を行っている。しかし、作業者が機械翻訳文から翻訳前の用例の意図を推測しなければならず、正確な訂正文の取得が可能となる用例はほんの一部である。そこで本稿では、訂正を行う多言語テキストペアを作業者に提示する際に、作業者が翻訳前の用例の意図をより推測しやすくするために、評価対象の多言語テキストペアを会話文として提示し、応答用例対を作成する手法を提案する。本稿の貢献は、クラウドソーシングを用いて作成した質問と回答の対をクラウドソーシングと機械翻訳に利用することで、一部の多言語テキストペアにおいて、会話の支援に有用な応答用例対の作成が行える可能性を示した点である。

Proposal of a Pair of Dialogic Parallel Texts Creation Method Using Crowdsourcing

SATOMI YAMAMOTO¹ TAKU FUKUSHIMA² TAKASHI YOSHINO¹

1. はじめに

現在、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。日本でも、外国人留学生数の増加 [1] や、2013 年の訪日外国人数が 1000 万人を突破したこと [2]、日本政府が 2020 年に訪日 2000 万という目標を掲げていること [3] から、今後も訪日外国人数は増加すると考えられる。そのため、今後も多言語間コミュニケーションの機会も増加すると考えられる。しかし、一般に多言語を十分に習得することは難しく、日本語を理解できない外国人と日本人との間で正確な情報を共有することは非常に困難である。

その影響が顕著に現れる分野の 1 つとして医療分野がある。医療分野では、わずかなコミュニケーション不足によって医療ミスが生じる恐れがある。現在、医療現場にお

いて日本語の理解できない外国人の支援は医療通訳者が行っている。しかし、慢性的な人員不足や、通訳者の身分保障、通訳者自身のメンタルケアなどの問題が存在している。また、入院患者が日本語の通じない外国人の場合は 24 時間の対応が必要となる。このような問題は外国人が多くない地域でも存在しており、対応する必要が出てきている。そのため、多言語間コミュニケーション支援として、アイコンを用いたコミュニケーションシステムの開発 [4] や、用例対訳や機械翻訳などの言語資源を組み合わせることで利用できる言語グリッド *1 の活動などが行われている [5]。

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)*2 による用例対訳の収集を行っている [6]。TackPad には多くの用例や用例対訳 *3 が収集されているが、医療現場で用いるために必要な正確性評価が十分にされてい

¹ 和歌山大学システム工学部
Faculty of Systems Engineering, Wakayama University
Sakaedani 930, Wakayama, 640-8510 Japan

² 静岡大学大学院工学研究科
Graduate School of Engineering, Shizuoka University Jo-
hoku 3-5-1, Naka-ku, Hamamatsu, 432-8561 Japan

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

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

*3 本稿では、正確性の確保が行われた多言語の用例の対用例群を「用例対訳」、正確性の確保が行われていない多言語の対の用例群を「多言語テキストペア」とする。

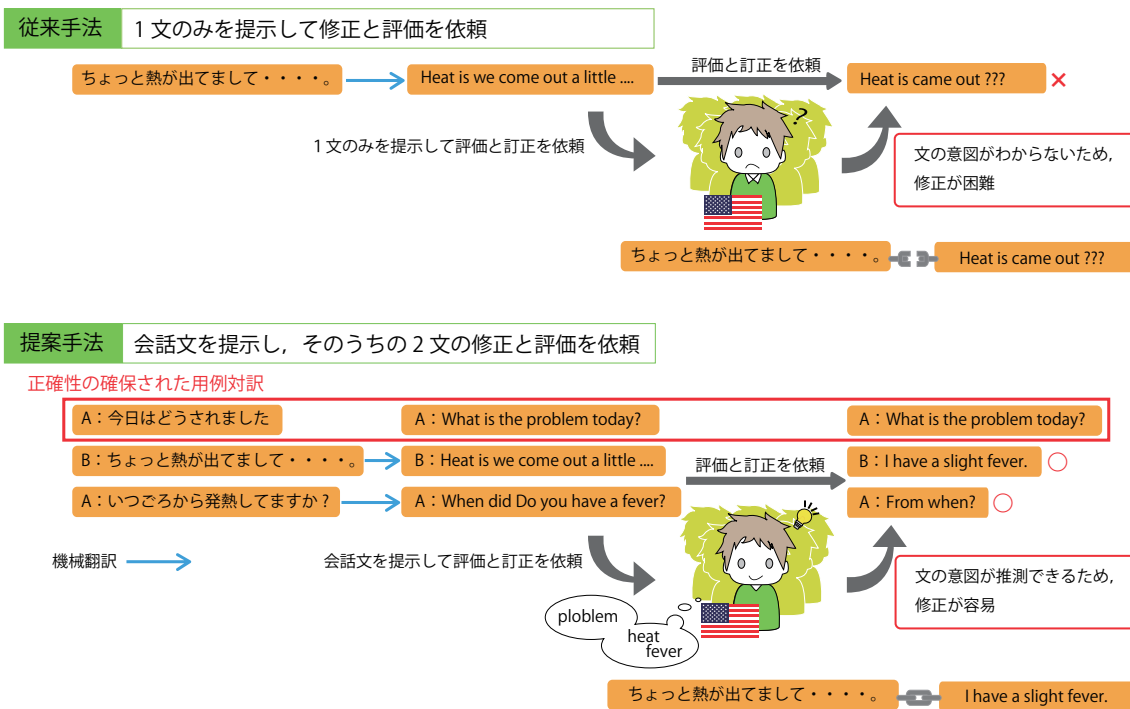


図 1 既存手法と提案手法の比較
 Fig. 1 Comparison of a conventional method and a proposed method.

ないものが多い。十分な正確性評価を得るためには、信頼度の高い人に作成や評価を依頼する方法があるが、医療分野で必要とされる用例の数に見合うだけの評価者を用意することは困難である。そのため、他の方法として、より多くの人から評価を取得する方法がある。

そこで本稿では、クラウドソーシングを用いて会話を作成し、機械翻訳によって別の言語に翻訳した後にクラウドソーシングによって評価と修正を依頼することで、専門家に依頼することなく応答用例対の作成を行う手法を提案する。また、本手法で作成する会話は、疑問文を会話の始点とした、質問とそれに対する回答、その回答に続く1文の3文で構成される。なお、応答用例対は文献 [7] での定義に従い、1個の質問と1個の回答、0個以上の類似文から構成されるものとする。

2. 関連研究

現在、多言語間コミュニケーションの支援を目的として、用例対訳や機械翻訳を用いた支援技術の研究が多く行われている。機械翻訳は自由に入力された文章を多言語に翻訳することが出来るため、多言語対応のチャットシステムの開発 [8] や、使用言語の異なる複数人グループ同士のコミュニケーション支援の研究 [9] などで利用されている。また、前処理や後編集、翻訳リペアなど、機械翻訳の精度を向上させるための手法の研究 [10], [11], [12] や、誤りがある機械翻訳文でもコミュニケーションを円滑に行う手法の研究 [13] が多く行われている。機械翻訳の精度は年々向

上しているが、現在の翻訳精度では医療分野で用いることが出来るだけの正確性を確保することができない。そのため、医療分野などの正確性が重要となる分野でのコミュニケーション支援では用例対訳を用いた方法が多く行われている。用例対訳を使用したシステムには、多言語医療受付支援システム M^3 (エムキューブ)[14] や、入院生活での多言語対話の支援を行うスマートフォン対応の多言語医療対話支援システムぶち通 [15] などがある。

また、計算機では判断が困難なデータに対してクラウドソーシングを用いることで、正確な評価を取得する研究が多く行われている。クラウドソーシングでは、計算機では困難とされる、見た目の印象や言葉の微妙な差異の判断、情報の収集などを行うことが出来るため、計算機を用いて情報を収集、判断するよりも正確な情報を得ることが出来る。このため、クラウドソーシングを用いた書籍や画像へのアノテーション、災害時の非難支援情報の収集などの研究が行われている [16], [17]。しかし、クラウドソーシングでは不特定多数の作業者に情報の評価や収集などを依頼するため、不適切に対価を得ようとする作業者の存在や、タスクの設計手法について考慮しなくてはならない。そのため、より精度の高い評価、情報を取得するための手法の提案や研究も行われている [18], [19], [20]。

また、我々は現在、用例対訳の収集に機械翻訳とクラウドソーシングを用いることで、専門家に依頼することなく、専門家に依頼するよりも安価に多言語用例対訳を作成する手法の研究を行っている。翻訳の依頼ではなく、機械翻

訳文の評価と訂正の依頼であれば、作業者が単言語話者であっても正確な用例対訳が作成できるのではないかと考えたためである。しかし、すべての機械翻訳文を正しく訂正することは難しい。そのため、評価と訂正を行う作業者に提示する情報によって訂正精度の向上をはかる研究を行っている [21]。

本手法では、作業者に正しい用例対訳を含む会話を提示することで、誤った機械翻訳文を翻訳前の文の意図に沿った文へ訂正することを目指す。作業者に評価する文が使用される会話を示すことに、正確な訂正が行われる可能性があると考えたためである。図 1 に既存手法 (文献 [21]) と提案手法についてを示す。本稿では、本手法との比較対象として、多言語テキストペアの評価、訂正を行う際に、評価対象の多言語テキストペアのみを作業者に提示する手法を従来手法とする。

3. 提案手法

本章では、クラウドソーシングを用いた質問と回答の対作成手法と、作成した質問と回答の対を含む会話文とクラウドソーシングを用いた多言語用例対訳作成手法について述べる。3.1 節でクラウドソーシングを用いた質問と回答の対作成手法について述べる。その後、3.2 節で、作成された質問と回答の対を含む会話を機械翻訳で翻訳し、翻訳文に対してクラウドソーシングを用いて評価、訂正を行うことで多言語用例対訳を作成する手法について述べる。

3.1 クラウドソーシングを用いた質問と回答の対作成手法

本節では、クラウドソーシングを用いた質問と回答の対作成手法について述べる。本手法は、クラウドソーシングを用いて、TackPad 内のある用例から開始されると考えられる会話文を作成することで、質問と回答の対を作成する。

図 2 に質問と回答の対を含む会話文作成の流れを示す。本手法は以下の 5 ステップで構成されている。

Step 1 会話の始点となる用例対訳の取得

図 2 の (1) で、会話の始点となる用例対訳を取得する。取得する用例対訳は疑問文とする。

Step 2 質問と回答の対の作成

図 2 の (2), (3) で、クラウドソーシングによって、Step 1 で取得した用例の回答に適した文と、さらにそれに続く 1 文の入力を依頼する。本手法では 1 つの会話文の作成で最大 2 組の質問と回答のペアが作成される。

Step 3 類似文の検索

図 2 の (4) で、Step 2 で作成された用例の類似文がすでに TackPad に登録されているかどうかを調べる。本手法では、類似文を類似度の高い順に最大で 3 文抽出する。類似度の判定は N-gram に基づく用例対訳検索手法を利用している [22]。文献 [22] では、用例を 2-gram もしくは 4-gram に分割し、検索文字列との

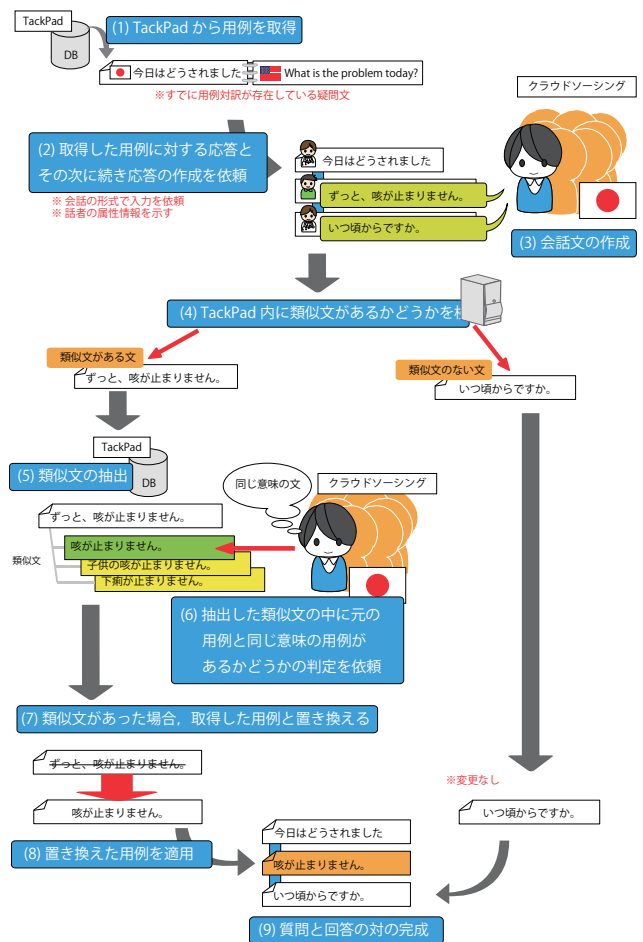


図 2 クラウドソーシングを用いた質問と回答の対作成の流れ
Fig. 2 The flow of creating a combination of question and answer using crowdsourcing.

共起を調べることで、多言語の類似文検索を実現している。また、本実験で作成した用例同士で類似文している文の調査も行った。この類似判定は、作成された用例同士の類似度の判定に PHP の similar text 関数を用い、類似度が 80%以上のものを類似文とした。なお、TackPad より抽出した類似文と、本実験で作成した用例から抽出した類似文を区別するために、前者を類似文 A、後者を類似文 B とする。

Step 4 類似文の判定

図 2 の (6) で、クラウドソーシングによって Step 3 で抽出した類似文の中に、作成した文と同じ意味の用例があるかどうかの判定をクラウドソーシング上で行う。

Step 5 応答用例対作成のための会話文の作成

図 2 の (9) で、Step 4 で類似文が存在しなかった場合には Step 2 で作成された文を使用し、類似文が存在した場合は TackPad 内の用例を作成された文の代わりに使用し、質問と回答の対を含む会話文を作成する。

Step 5 で類似文を取得し、クラウドソーシングを用いて取得した文の代わりに使用することによって、意味が同じ

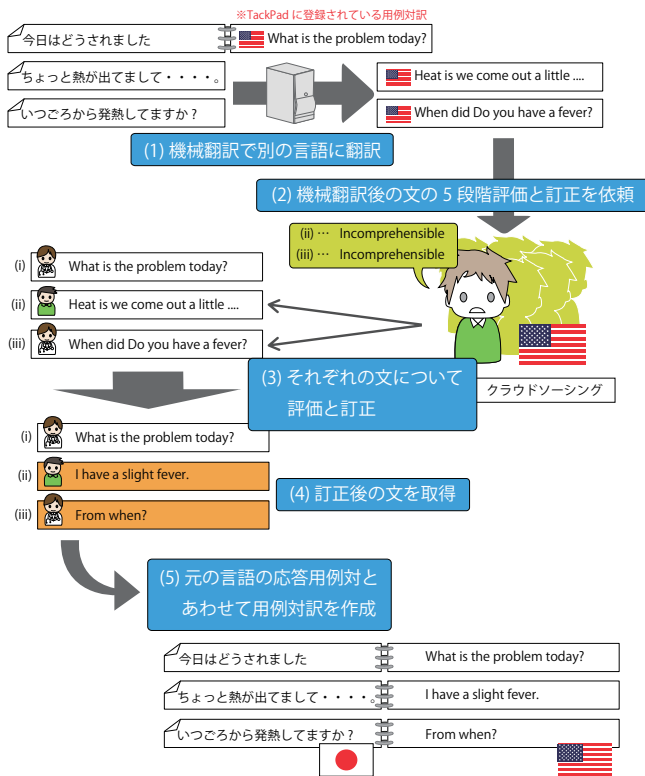


図 3 クラウドソーシングを用いた多言語テキストペア評価と用例対訳作成の流れ

Fig. 3 The flow of evaluation multilingual text pair and creation parallel text.

用例を TackPad に複数登録することを防ぐ。また、取得した類似文に用例対訳が登録されていた場合、作業者に提示する会話文にその用例対訳を使用することで他の多言語テキストペアの評価、修正がより正確に行える可能性がある。

3.2 クラウドソーシングを用いた多言語用例対訳作成手法

本節では、クラウドソーシングを用いた多言語用例対訳作成手法について述べる。本手法は、3.1 節で述べた手法によって取得した質問と回答の対を機械翻訳で翻訳し、クラウドソーシングで評価、訂正文の作成を依頼することによって、より正確な用例対訳の作成を行う。

図 3 に提案手法の流れを示す。本手法は以下の 3 ステップで構成されている。

Step 1 機械翻訳文の取得

図 3 の (1) で、3.1 節で取得した質問と回答の対を含む会話文の中で TackPad 内にまだ用例対訳がない文に対して機械翻訳を用い、別の言語に翻訳する。

Step 2 翻訳された用例の評価

図 3 の (2), (3) で、クラウドソーシングによって多言語テキストペアの評価、訂正文の入力を行うタスクを依頼する。図 3 の (4) で、作業者は提示された用例を 5 段階で評価し、訂正がある場合には訂正文を入力す

る。また、評価の際は文献 [23] の評価基準^{*4}を用い、複数選択を可とする。

Step 3 質問と回答の対を含む会話文の作成

Step 2 で取得した多言語テキストペアに対する評価と訂正文より、3.1 節で取得した文の用例対訳を作成し、それを質問と回答の対を含む会話文とする。

4. クラウドソーシングを用いた会話と質問の対作成実験

本章では、3.1 節において述べた手法の実験について述べる。本実験では、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^{*5}を使用した。Yahoo!クラウドソーシングは Yahoo!JAPAN の運営するクラウドソーシングサービスであり、作業者を Yahoo!クラウドソーシングサービスに登録した Yahoo!JAPAN ユーザとしている。また、作業者に支払われる報酬は T-ポイント^{*6}として支払われる。

4.1 実験用データセット

本節では、実験用のデータセットについて述べる。本実験では、TackPad に登録されている用例 20 文を用いた。使用する用例の条件について以下に示す。

条件 1 正確性の評価された英語の用例対訳がすでに登録されている。

条件 2 英語の用例対訳に “What” または “How” を含む疑問文である。

条件 3 英語の用例対訳が “How much ~” や “How many ~?” などの数値が回答となる疑問文でない。

4.2 質問と回答作成

本節では、3.1 節の Step 2 で述べたクラウドソーシングを用いた質問と回答の対作成実験について述べる。本実験では、4.1 節で述べた用例 20 文についてそれぞれを始点とする会話文を 2 個作成するタスクを 1 つのタスクとして、1 文あたり 10 人の作業者に依頼した。1 タスクあたりの報酬は 5 ポイントとした結果、本タスクに対する請求金額は 1512 円 (税込)^{*7}となった。そのため、1 つの会話文作成あたり 7.56 円のコストがかかったことになる。また、質問と

*4 評価段階は、1: Incomprehensible(理解できない), 2: Disfluent English(流暢でない), 3: Non-native English(非母語言語), 4: Good English(良い英語), 5: Flawless English(完璧な英語)

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

*6 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り、ポイントは商品や現金と交換することが可能。

*7 Yahoo!クラウドソーシングが定める基準により、1 タスクあたり 2 ポイントの報酬が設定されたが、本タスクでは文の作成を依頼するため、作業者の負担を考慮して報酬を 3 ポイント追加した。1 ポイント追加するごとに 1.5 円かかり、3 ポイント追加した場合は、少数点以下を切り捨てた 4 円のコストがかかる。また、Yahoo!クラウドソーシングが定める標準利用料は 1 タスクあたり 10 円のため、報酬ポイント追加分の 4 円とあわせて、1 タスクあたり税抜きで 14 円のコストとなる。

ブラウザ、スマートフォン本体の戻るボタンは押さないようご注意ください。

最初の文に続くような会話文を作成してください。
 会話する場所は、病院などの医療現場を想定しています。

2人の役割はそれぞれ、「医療従事者」と「病院利用者」です。詳しくは下の画像を参考にしてください。

※最初の文に続くように解答欄につき1文を入力してください。
 ※「はい」や「いいえ」、単語や相槌のみの回答は不可です。

医療従事者

- ・医師
- ・看護師
- ・病院事務
- ・作業療法士
- など

病院利用者

- ・受診患者
- ・入院患者
- ・見舞い客
- など

属性情報

医療従事者	体の調子は良いですか。
病院利用者	<input type="text"/>
医療従事者	<input type="text"/>

※「はい」や「いいえ」、単語や相槌のみの回答は不可です

確定して次へ

[前の設問に戻る](#)

・不正防止のため、設問の合間にかんたんチェック設問が入ることがあります。
 ・チェック設問の誤解答が多い場合、続けてタスクを実施できなくなります。

図 4 クラウドソーシングにおける質問と回答の対を含む会話文作成タスク画面

Fig. 4 Screenshot of creating a combination of question and answer using crowdsourcing.

回答の対の作成にあたって、作業には以下のことを禁止した。

- 「はい」や「いいえ」のみの回答
- 単語や相槌のみの回答
- 1つの解答欄に2文以上記入すること

また、作業に対して、提示する用例の使用者の属性情報^{*8}を示した。この属性情報は用例対訳として登録されている用例の主語によるものや、著者の一人が判断して決定したものである。タスク画面の例を図4に示す。

本実験では、作業によって200組の会話文が作成され、400文の用例が取得できた。禁止事項を守らない作業によって作成された会話文のデータなどは、後述の表1に示す。また、用例を区別しデータの整理を行うために、作成された用例には1~400までのIDを付与した。用例ID1~200までは提示された用例に対する回答であり、用例ID201~400は作業者がそれぞれに作成した用例に対応する用例である。

^{*8} 本稿では、用例の使用者の属性情報として「医療従事者」と「病院利用者」に使用した。

◆選択肢の中から、意味が似ている文を選ぶタスクです。◆

問題文：「内臓脂肪が多く見られます。」

上の問題文と同じ意味または似ている意味の文を、以下の選択肢より選んでください。

(※複数選択可)

この文の使用場所は病院などの施設を想定しています。

- *選択の仕方***
- ・意味が似ていると感じた文を選択してください。
 - ・選択する数に制限はありません。
 - ・同じ意味の文がない場合には“同じ文がない”を選択してください。
 - ・選択肢が全て単語の場合も“同じ文がない”を選択してください。

- *入力欄について***
- ・複数選択した場合に入力してください。
 - ・最も意味が似ていると感じる文を入力してください。
 - ・選択肢の文を変更せずに入力してください。(コピーして入力すると確実です)

問題文：「内臓脂肪が多く見られます。」

- 内臓脂肪
- 脂肪
- 脂肪肝
- 同じ文がない

入力欄 (複数選択した場合に入力)

図 5 クラウドソーシングにおける類似文判定タスク画面

Fig. 5 Screenshot of judging similarity using crowdsourcing.

4.3 類似文判定

本節では、3.1節のStep 4で述べたクラウドソーシングを用いた類似文判定実験について述べる。

図5にタスク画面の例を示す。

本実験では、後述する表1の(4)に対して図2の(5)を行った結果、類似文A(TackPadに類似文が存在していた用例)と、類似文B(本実験で作成した用例同士で類似文が存在する用例)をあわせた319文(表1(8))を使用した。なお、この319文は図2の(6)を行うための選択肢を作成するにあたって、表1の(6)と(7)の用例について、まったく同じ書き方の用例の統合や、類似文が一致した用例の連結などを行ったためこの数となった。

用例10文についてそれぞれの類似文判定を行うタスクを1つのタスクをして、1文あたり10人の作業者に依頼した。1文あたりの類似文判定を10回行うため、総タスク数は320件^{*9}となった。また、Yahoo!クラウドソーシングの定める基準により、1文の用例の類似度判定には10.8円(税込み)のコストがかかった。

^{*9} 1タスクあたり10文の用例の類似文判定を行うため、設問数が10の倍数である必要があった。そのため、319番目の用例を複製し2文にした。複製した1文については用例の類似度判定の結果に含めないものとする。

表 1 類似文判定を行った結果

Table 1 Result of the judging similar texts.

データの分類	データの数
(1) 4.2 節で作業者によって作成された会話文	200 組
(2) 4.2 節で作業者によって作成された用例	400 文
(3) 4.2 節で作成された不正な会話文	5 組
(4) 4.2 節で作成された不正な用例	10 文
(5) (2) から (4) を除いた用例	390 文
(6) 類似文 A((5) をもとに TackPad から抽出)	349 文
(7) 類似文 B((5) 同士で類似した文)	72 文
(8) 類似文 A と類似文 B より, 類似文候補を持つ用例 (4.3 節の実験で使用)	319 文
(9) 類似文 A や類似文 B に含まれない用例	42 文
(10) 図 2 の (6) で類似文なしと判定された用例	172 文
(11) 類似文がない用例 ((9)+(10))	214 文
(12) 図 2 の (8) を行った用例	61 文
(13) (12) のうち, TackPad に英語の用例対訳が登録されている用例	6 文
(14) (11), (12) の用例で構成される会話文	38 組
(15) (13) のみで構成される会話文	1 組

(10), (12) の判断基準として, 類似文の判断を行った作業者の 8 割以上が一致する回答を行った場合と定めた。

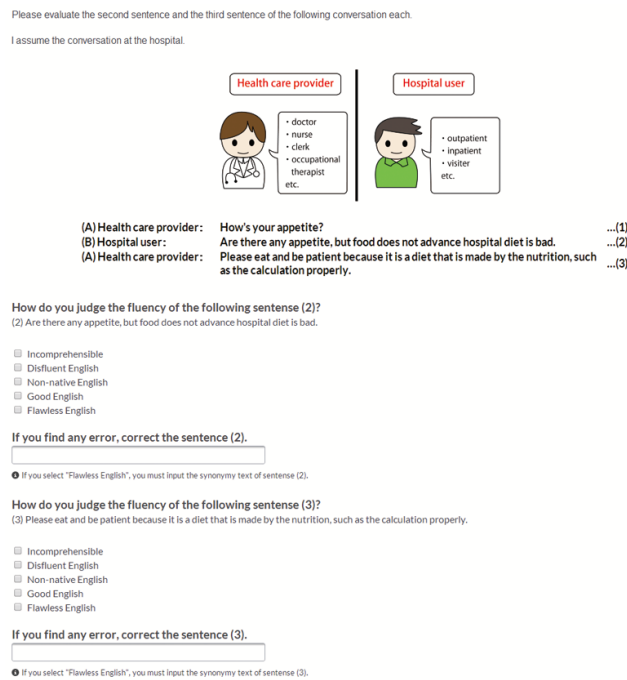


図 6 クラウドソーシングにおける対作成タスク画面

Fig. 6 Screenshot of example-text evaluation using crowd-sourcing.

本章での実験で取得したデータについて, 表 1 にデータの種類とその数を示す。

5. クラウドソーシングを用いた用例対訳作成実験

本章では, 3.2 節において述べた手法の実験について述

べる。本実験では, クラウドソーシングサービスとして CrowdFlower^{*10} を使用した。CrowdFlower は何かを判定することをタスクとするクラウドソーシングサービスを提供している。また, 機械翻訳サービスとして Google 翻訳^{*11} を使用し, 日本語から英語に翻訳した。

5.1 実験用データセット

本節では, 実験用のデータセットについて述べる。本実験で使用するデータは表 1 の (14) の 38 組の会話文とした。この 38 組は, 表 1 の (1) のうち, (11) と (12) に 2 文目と 3 文目の両方が存在している会話文を抽出したものである。表 1 の (15) の 1 組の会話文はすでに応答用例対が完成しているが, クラウドソーシングを用いた多言語テキストペアの訂正精度の評価に用いるため, 表 1 の (14) の 38 組の会話文から除かずに使用した。

5.2 評価データの収集

本節では, クラウドソーシングを用いた多言語テキストペアの評価と訂正を行うことによって作成される評価データの収集について述べる。本実験では, 機械翻訳や TackPad 内の用例対訳を用いて作成した英語の会話文の中の 2 文目と 3 文目の評価, 訂正を行うことを 1 つのタスクとした。また, 1 つの応答用例対あたり 10 回の評価が行われるように CrowdFlower の作業者に評価依頼を行った。1 つのタスクに対する報酬は 10 セントと設定し, 1 つの応答用例対について 10 人からの評価を得る。評価を行う会話文は 1 タスクあたり 3 個とし, 本実験で行う総タスク数は 127 件となった。127 件のタスクの依頼に必要な金額は \$16.85 となった。そのため, 今回の実験では 1 つの応答用例対の評価に対し, \$0.443 (日本円でおおよそ 45.31 円)^{*12} のコストがかかった。なお, 評価は CrowdFlower によって定められたレベル 3^{*13} の, 国を United States とする作業者に依頼した。タスクの画面例を図 6 に示す。

また, クラウドソーシングではタスクの終了期限を設けていない。そのため, 今回の実験では, 1 文につき 5 回以上の評価・訂正データを収集することが出来た多言語テキストペアについて分析を行った。本実験において分析に使用できた応答用例対の数は 20 個である。

6. 考察

6.1 質問と回答の対作成の評価

本節では, 4 章におけるクラウドソーシングを用いた質問と回答の対作成の結果について分析を行う。本実験では 20 文の用例をもとに, それぞれに対応する質問と回答の対

*10 <http://crowdfunder.com/>

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

*12 2014/5/14 現在, 1 ドル = 102.27 円で計算。

*13 CrowdFlower は, 作業者群を 3 段階の能力評価基準で分けている。レベル 3 は最も上のレベルである。

表 2 類似文判定の結果の例

Table 2 Example of result of judging similarity.

ID	判定用の用例	選択肢
118	清潔を保ってください。	足を清潔に保ってください
		頑張ってください。
		同じ文がない

・最も回答数の多かった選択肢を太字で示す。

表 3 正確な用例対訳に対する評価結果

Table 3 Result of evaluation of the parallel texts.

ID	用例対訳	評価値				
		1	2	3	4	5
2	I feel pressure in my head.	0	1	0	1	5
133	Please keep your feet clean.	0	0	0	2	3
202	I will take a CT scan of your brain.	0	0	0	2	3

・表中の数字は、評価した人数を示す。

・各項はそれぞれ 1 : Incomprehensible, 2 : Disfluent English, 3 : Non-native English, 4 : Good English, 5 : Flawless English である。

を含む会話文を 200 個取得した。本実験で得た会話文 200 個のうち、タスクを依頼する際の禁止事項に従わなかった作業員^{*14}がいた。そのため、4.3 節の実験では 195 個(表 1 の (1) から (3) を引いたもの)の会話文について分析を行った。

また、4.3 節の類似文判定では、判定用に提示した用例と選択肢のどちらが会話に適した用例であるかの評価は行わなかった。なお、本実験では TackPad に登録されている用例を優先して使用した。これは、3.1 節に述べた理由によるものである。そのため、元の文と意味の異なる用例が選択される可能性が考えられる。表 2 に例を示す。なお、本実験では水虫に関する会話だったため、後述の表 5 にあるように正しい訂正が行われたが、質問と回答の対を含む会話文が足に関する会話でない場合もある。そのため必ずしもそうなるとは限らないと考えられる。

6.2 多言語テキストペアの正確性評価

作業員によって作成された質問と回答の対を含む会話文の中には、用例が 2 文含まれている。そのため、本実験で評価と訂正を依頼した多言語テキストペアの数は、評価に用いた質問と回答の対を含む会話文 20 組の 2 倍の 40 文となった。40 文のうち 3 文はすでに TackPad に用例対訳が作成されている用例であった。3 文の評価結果を表 3 に示す。

この結果より、クラウドソーシングを用いた多言語テキストペアの評価、訂正が正しく行われる可能性があると考えられる。また、この 3 文を除いた 37 文に対する評価と

*14 「はい」や単語のみの回答を行ったものが 4 個、1 つの解答欄に 2 文の入力が行われた回答が 1 個あった。

訂正結果の例を表 4 に示す。

本実験では 1 つ応答用例対につき 5 回の評価を行った。評価データを取得できた応答用例対は 20 個なので、評価した用例の数は 40 文であり、作業員によって入力された訂正文の総数は 200 文である。ただし、そのうち 58 文は訂正ではなく、感想や評価を行った多言語テキストペアの入力が行われていた^{*15}。

また、多言語テキストペアの評価において “Incomprehensible” が選択された回答は 23 個であったが、そのうちの 5 文は元の用例の意図に合った訂正が行われていた。しかし、訂正文として “NO IDEA WHAT HE MEANS” や空白などといった入力が行われ、訂正が行われていない回答は 6 文あり、応答用例対を用いるだけでは全ての多言語テキストペアを正しく訂正することは出来ないことが分かった。6.2.1 項で正しい訂正が行われた多言語テキストペアについての考察述べ、6.2.2 項で正しい訂正が行われなかった多言語テキストペアについての考察を述べる。

6.2.1 正しい訂正が行われた多言語テキストペア

用例 ID73 の “There is a root of its own way” は、4 名の作業員が “Incomprehensible(理解できない)” と評価しており、1 名が “Disfluent English(流暢でない)” と評価した。評価結果が低い点は作業員全員に共通しているが、訂正結果では差が現れている。5 名のうち 2 名は訂正を行うことができず、“What does it mean?” や “?” と回答しているが、その他の 3 名は、用例 ID73 には書かれていない足または水虫という言葉を使用して訂正文を作成している。

用例 ID73 を含む会話文について、作業員に示した英語と、本実験において作成された応答用例対を表 5 に示す。

表 5 より、用例 ID73 は水虫の感染源についての会話を行うときに使用する用例だと分かる。そのため、訂正を行った 3 名の作業員は前後の文から推測し、訂正を行ったと考えられる。また、用例 ID73 に関して、翻訳前の文の意図とあった訂正が行われたのは、“There are various way to contract athlete’s foot” と “There are many different roots of athlete’s foot.” の 2 文である。この訂正文を回答した作業員はどちらも、多言語テキストペアの評価では “Incomprehensible” としていたが、正しい訂正が行われている。このことより、多言語テキストペアの評価、訂正を行う際、応答用例対を用いて前後の文を示し訂正を依頼することで、応答用例対を用いない場合よりも、正しい用例対訳が作成できる可能性があると考えられる。

6.2.2 正しい訂正が行われなかった多言語テキストペア

“Incomprehensible” という回答があり、正しく訂正が行われなかった用例について、使用した回答と質問を含む会話文の例を表 6 に、提示した機械翻訳文と訂正文の例を表 7 に示す。なお、訂正結果を示すために、各会話文に ID

*15 “none” や “na”, “What does it mean?” といった、訂正文とは異なる英語の入力が行われていた。

表 4 クラウドソーシングを用いた多言語テキストペアの評価、訂正結果の例
 Table 4 Example of the evaluation results and the input correction sentences.

ID	元の用例	評価値					訂正文
		1	2	3	4	5	
7	And throbbing	0	4	1	0	0	Throbbing pain.
							I have a throbbing pain.
							It's a throbbing pain.
							A throbbing pain.
73	There is a root of its own way	4	1	0	0	0	What does it mean?
							?
							There are various way to contract athlete's foot
							There is problem on my foot.
							There are many different roots of athlete's foot.
182	It may bot be 100% successful	0	0	0	2	3	none
							It might not be perfect.
226	Are there any symptions to be worried about by health?	0	1	3	1	0	Are there any symptoms that worry you?
							Are there any symptoms to worry about?
							Do you have any symptoms that worry you?
							Do you have any health issues?
							Are there any symptoms you're worried about?

- ・表中の数字は、評価した人数を示す。
- ・評価値の各項目はそれぞれ 1 : Incomprehensible, 2 : Disfluent English, 3 : Non-native English, 4 : Good English, 5 : Flawless English である。
- ・訂正が行われていない場合と、他の訂正文と重複がある場合は省略した。

表 5 用例 ID73 を含む会話文における訂正結果
 Table 5 Result of the correction of sentences including sentence(73).

	日本語	作業者に提示した英語	作成された応答用例対
(1) 質問	水虫の感染源はどこですか？	What is infectious source of athlete's foot?	What is infectious source of athlete's foot?
(2) 回答	それなりのルートがあります	There is a root of its own way	There are various way to contract athlete's foot
(3) 質問	私の場合は何ですか？	What is in my case?	What is it in my case?

- ・(1) は TackPad に登録された用例対訳を使用しているため、訂正を行わなかった。
- ・(2), (3) は 4 章で作成した用例に機械翻訳を用いた多言語テキストペアのため、訂正を依頼した。

を付与した。

表 6 と表 7 の (130, 330), (71, 271) では、機械翻訳によって作成した英語に文法的な誤りが見られないことがわかる。このことから、機械翻訳で文法的に正しい翻訳が行われた場合に正確な訂正が行えない可能性があると考えられる。

7. おわりに

本稿では、クラウドソーシングを用いた応答用例対の作成手法と、作成した応答用例対を用いた多言語用例対訳作成手法の提案を行った。また、クラウドソーシングを用いた応答用例対作成と、多言語テキストペア評価の実験を行った。

本手法では、クラウドソーシングを用いた多言語テキストペアの評価と訂正を行う際に、評価する多言語テキストペアを含む会話文を提示することで、正確な多言語用例対訳作成を行うことを目指した。また、使用する会話文は応答用例対を含むものとし、クラウドソーシングによって作成した。本稿の貢献は、クラウドソーシングを用いた多言語テキストペアの評価と訂正を行う際に、応答用例対を含む会話文を提示することで、誤った翻訳が行われた機械翻訳文の訂正が行える可能性を示したことである。

今後は、今回の実験でデータを収集しきれなかった 18 文についての分析を行う。また、クラウドソーシングを用いた多言語テキストペアの評価、訂正による多言語用例対訳作成における 5 章で使用しなかった会話文の使用につい

表 6 正しい訂正が行われなかった多言語テキストペアを含む会話文の例

Table 6 Example of dialogs including false corrections.

ID	言語	(A) の発言	(B) の発言	(A) の発言
(130, 330)	日本語	何が原因でしょう。	日頃の不摂生ですね。	不規則な仕事なので。
	英語	What is the cause?	It is the neglect of health everyday.	Because it is the irregular work.
(71, 271)	日本語	水虫の感染源はどこですか？	家の中だと思います	家族の方で水虫の方がいますか
	英語	What is infectious source of athlete's foot?	I think it's in the house	Do you have a person of athlete's foot in the way of family
(13, 213)	日本語	食欲はありますか？	余りありません。	症状はずっと続いていますか。
	英語	How's your appetite?	It is not too much.	Symptoms Do you have followed all the way.

・太字で示した用例に対して行われた訂正の結果を表 7 に示す。

・本稿で使用している ID は、質問と回答の対に対して付与している。本表には質問と回答の対が二つずつ含まれているため、ID も二つ存在している。

表 7 表 6 に示した多言語テキストペアの訂正結果の例

Table 7 The result of corrections in Table 6

ID	機械翻訳文	訂正文
(130, 330)	Because it is the irregular work.	It is the irregular work.
		Because, it is the irregular work.
(71, 271)	I think it's in the house	I think that it's in the house.
		i thinks in hose
		NO IDEA WHAT HE MEANS
(13, 213)	Symptoms Do you have followed all the way.	What are your symptoms?
		Do you have any other symptoms?
		Did you follow it all the way?
		GIBBERISH
		What continuing symptoms have you had?

・“na”や“none”などの入力や他の訂正文と同一のもの、評価対象文と同一のものは省略した。

て検討する。6.2 節において正しく訂正が行われなかった多言語テキストペアの分析を行い、より最適な多言語用例対訳作成手法について検討する。

謝辞

本研究の一部は、和歌山大学平成 25-26 年度独創的研究支援プロジェクトおよび JSPS 科研費 (26730105) の補助を受けた。

参考文献

- [1] 独立行政法人日本学生支援機構:平成 23 年度外国人留学生在籍状況調査結果,独立行政法人日本学生支援機構(オンライン),入手先<<http://www.jasso.go.jp/statistics/intl.student/data11.html>>(参照 2014-5-10).
- [2] 日本政府観光局(JNTO):訪日外客数の動向(オンライン),入手先<http://www.jnto.go.jp/jpn/reference/tourism_data/visitor_trends/>(参照 2014-5-10).
- [3] 首相官邸:第 2 回観光立国推進閣僚会議(2014-6-11),首相官邸(オンライン),入手先<<http://www.kantei.go.jp/jp/singi/kankorikkoku/dai2/siryuu.html>>(参照 2014-5-10).
- [4] 山田 豊通,清水 由美子,関 裕志:異言語間コミュニケーション支援システムの提案,武蔵野工業退学環境情報学部情報メディアセンタージャーナル,第 2 号,pp. 40-45 (2001).
- [5] Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration, IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06), pp. 96-100 (2006).
- [6] 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. 22-33 (2012).
- [7] 福島拓,吉野孝:正確な多言語間対話支援を目的とした応用例対構築モデルの検討,情報処理学会,マルチメディア,分散,協調とモバイル(DICOMO2012)シンポジウム,pp. 551-559 (2012).
- [8] 藤居 薫和,重信 智宏,吉野 孝:異文化コミュニケーションのための機械翻訳を用いたチャットシステム AnnoChat の開発と適用,第 4 回情報科学技術フォーラム(FIT2005),pp. 437-438 (2005).
- [9] Naomi Yamashita, Rieko Inaba, Hideaki Kuzuoka, Toru Ishida: Difficulties in Establishing Common Ground in Multiparty Groups using Machine Translation, Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'09), pp. 679-688 (2009).
- [10] 宮田玲,立見みどり,Anthony Hartley,影浦峽,井佐原均:日英機械翻訳の制度改善と原文の読みやすさ向上の

- ための日本語書き換えルールの作成と評価言語処理学会,
第 19 回年次大会発表論文集, pp. 710-713 (2013).
- [11] 井佐原均:クラウドソーシング後編集, Japio YEAR BOOK
pp. 238-241(2013).
 - [12] 宮部真衣, 吉野孝: 翻訳不適切箇所指摘による翻訳リベ
アコストと翻訳リベア精度の評価, 情報処理学会論文誌,
Vol. 50, No. 4, pp. 1390-1398 (2009).
 - [13] 角田 啓介, 菱山 玲子: コミュニケーション支援エー
ジェントを組み込んだ多言語参加型ゲーミングシステム
の設計, 2010 年度人工知能学会全国大会, 1C1-4 (2010)
 - [14] 宮部 真衣, 吉野 孝, 重野 亜久里: 外国人患者のため
の用例対訳を用いた多言語医療受付支援システムの構築,
電子情報通信学会論文誌, Vol.J92-D, No.6, pp. 708-718
(2009).
 - [15] 尾崎 俊, 松延 拓生, 重野 亜久里, 吉野 孝: 携帯端
末を用いた多言語医療対話支援システムの開発, 情報処
理学会, 第 73 回全国大会講演論文集, No.1, pp. 215-217
(2011).
 - [16] Philip van Oosten, Veronique Hoste: Readability annota-
tion: replacing the expert by the crowd IJNLPBEA '11
Proceedings of the 6th Workshop on Innovative Use of
NLP for Building, Educational Applications, pp. 120-
129 (2011).
 - [17] 木實 新一, 大野 航, 少路 健太: 災害情報の収集を支援
するクラウドソーシングツールキットの開発, 情報処理
学会, 第 75 回全国大会講演論文集 (第 3 分冊), 6D-1,
pp. 1-2 (2013).
 - [18] 小山聡, 馬場雪乃, 櫻井裕子, 鹿島久嗣: クラウドソーシ
ングにおけるワーカーの確信度を用いた高精度なラベル
統合, 第 27 回人工知能学会全国大会, pp. 1-4 (2013).
 - [19] 馬場雪乃, 鹿島久嗣, 木下慶, 山口豪志, 秋好陽介: 機
械学習による不適切なクラウドソーシングタスクの検出,
第 5 回データ工学と情報マネジメントに関するフォーラ
ム, pp.1-8 (2013).
 - [20] Kaxiang Mo, Erthng Zhong and Qiang Yang: Cross-Task
Crowdsourcing, In Proceedings of the ACM SIGKDD
International Conference on Knowledge Discovery and
Data Mining (KDD'2013), pp. 677-685(2013).
 - [21] 山本 里美, 福島 拓, 吉野 孝: クラウドソーシング上
における使用者の属性情報を用いた用例対訳生成手法の提案,
電子情報通信学会研究報告, AI2013-37, pp. 7-12(2014).
 - [22] 田淵裕章, 坂本廣, 北村泰彦: N-gram に基づく用例対訳
検索手法, 信学技報, 人工知能と知識処理研究会, Vol.108,
No.441, pp. 43-48 (2009).
 - [23] Kevin Walker, Moussa Bamba, David Miller, Xisoyi
Ma, Chris Cieri, and George Doddington: Multiple-
Translation Arabic (MTA) Part 1, In Linguistic Data
Consortium, Philadelphia (2003).