

マイクロブログとクラウドソーシングを用いた 用例評価手法および多言語用例対訳作成手法の提案

山本 里美^{1,a)} 福島 拓^{2,b)} 吉野 孝^{1,c)}

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非常に困難である．医療分野など正確な情報の共有が求められる場では，正確性の確保が行われた多言語の対である用例対訳が多く用いられている．用例対訳の提供には正確性評価が重要となるが，評価すべき用例対訳は多く，十分な評価を集めることは困難だと考えられる．そこで本稿では，短文の投稿ができるマイクロブログと，Web上で不特定多数の人へ作業委託を行うことが可能なクラウドソーシングを用いた，用例対訳作成および評価手法を提案する．本手法では，マイクロブログを用いて用例評価を行い，その結果をクラウドソーシングを用いて多言語用例対訳を作成する．本稿の貢献は以下の2点である．(1) マイクロブログによって用例評価を行う手法を提案し，用例の評価数の増加の可能性があることを示した．(2) マイクロブログを用いた用例評価の結果をもとに，クラウドソーシングを用いて多言語用例対訳の作成を行う手法を提案し，一部の多言語テキストペアにおいて正確な用例対訳の作成が行える可能性を示した．

1. はじめに

現在，世界のインターネット利用者数が27億人を突破し^{*1}，世界規模のインターネットの普及や，日本における外国人登録者数，留学生の増加などから，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，一般に多言語を十分に習得することは非常に難しく，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通じない外国人への支援は医療通訳者が行っているが，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，24時間の対応が必要であり，緊急の場合に通訳者が対応でき

るとは限らない．このような問題は，外国人が多くない地域でも対応する必要性が出てきている．2007年度の外国人登録者数が全国22位の宮城県において行われた調査 [1] では，79%の医療機関が日本語の不自由な外国人の対応経験があるが，外国人対応体制が整っているのは36%であった．また，内訳を見ると，来院外国人の母語は中国語が51%，韓国朝鮮語が27%，英語が23%であるのに対して，対応体制が出来ているのは中国語7%，韓国朝鮮語4%，英語34%となっており，患者の母語と対応可能言語が一致していない．このため，通訳者がいない状況での多言語対訳支援が必要である．

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)^{*2}による用例対訳の収集を行っている [2]．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ないものが多い [3]．

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もあるが，医療分野で必要な用例の数に見合うだけの評価者を用意することは非常に難しい．他の方法として，より多くの人からの評価を得る手法がある．

そこで本稿では，この TackPad に収められている用例をもとに，マイクロブログにより用例評価を依頼を行

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

² 静岡大学大学院工学研究科
Graduate School Engineering, Shizuoka University,
Hamamatsu 432-8561, Japan

a) s165058@center.wakayama-u.ac.jp

b) fukushima@sys.eng.shizuoka.ac.jp

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

^{*1} ITU “Global numbers of individuals using the Internet, total and per 100 inhabitants, 2001-2013”,
<http://www.itu.int/en/pages/default.aspx>

^{*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対の用例群を「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．

うことで、より多くの用例に対する評価を取得し、評価の行われた用例をクラウドソーシングに対訳作成を行うことで、より正確な多言語用例対訳の作成が可能となる手法を提案する。

2. 関連研究

1つの記事の文字数が200字前後^{*4}と制限されるマイクロブログは、その性質から即時性のある投稿が行われている。マイクロブログユーザの投稿記事内容や語句の情報収集、分析を行うことで、ユーザ情報の分析 [4] や、感情分析 [5] などの、この即時性のある投稿に着目した研究が多く行われている。また、SNS性に着目し、販売業務支援を行うシステムの開発もされている [6]。

計算機では判断の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。見た目の印象や言葉の微妙な差異などの判断をしたり、情報の収集を行ったり出来るため、計算機によって判断、収集を行うよりも正確な情報を得ることが出来る。このため、情報の信憑性を判断するための情報源検索 [7] や災害情報の収集を支援するツールキットの開発 [8] など、様々な分野で利用されている。しかし、クラウドソーシングでは不特定多数の人物に評価や情報収集を依頼するため、不適切に評価を得ようとする作業者が評価を行う可能性を考えなくてはならない。そのため、より精度の高い評価、情報を得るための研究も行われている [9], [10]。

3. 提案手法

本章では、マイクロブログを用いた用例評価手法と、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、作成された用例をマイクロブログを用いて評価する手法について述べる。その後、3.2節で評価済みの用例を機械翻訳で翻訳し、クラウドソーシングを用いて対訳文を作成する手法について述べる。

3.1 マイクロブログを用いた用例評価

本節では、マイクロブログを用いた用例評価手法について述べる。本手法はマイクロブログへ用例の評価を依頼する記事を投稿することで、多くの閲覧者から用例の評価を取得する。また、本手法では、評価者の母語と評価する用例の言語が一致するため、正確な評価が行われる可能性がある。

図1に提案手法の流れを示す。本手法は下記の4ステップで構成されている。

Step1 用例の取得

図1(1)でTackPadより評価を行う用例を取得する。

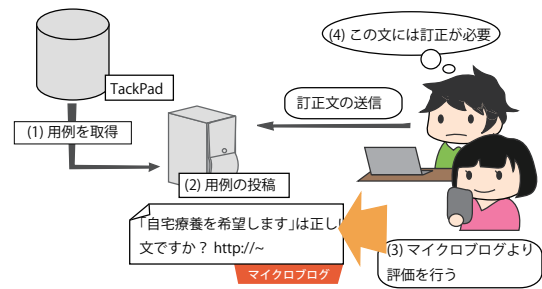


図1 マイクロブログによる用例評価手法の流れ

Step2 評価の依頼

図1(2)で、Step1で取得した用例を1文ずつマイクロブログへ定期的に投稿する。評価を依頼する用例はランダムに抽出する。投稿する記事には、評価される用例と評価を促す文、その用例の評価を行うWebページへ移動するためのURLを含む。

Step3 用例の評価

図1(3)で評価者はマイクロブログに投稿されたURLから評価を行うためのWebページへ移動する。評価ページでは用例の正しさを判断する。評価者はまずその用例が正しいか正しくないかの判断を行い、正しいと判断した場合は「正しい」という評価をデータベースへ送信する。正しくないと判断した場合は、更にその間違いが用例の綴り、文法、言語選択、入力文字、アクセント記号のいずれに該当するかを選択する。また、図1(4)で、評価者が用例訂正の必要性を感じた場合は、訂正文を入力し、データベースへ送信する。なお、本手法では評価する用例の言語を母語とする人に評価を依頼する。

Step4 訂正文の評価

Step3で入力された訂正文についての評価を行う。訂正文をStep2と同様にマイクロブログへ投稿し、評価を依頼する。

なお、本稿で評価に用いる用例の言語は日本語とした。また、評価者のデータを取得するためにログイン機能を実装した。このことにより、Twitterアカウントによって本システムへログインを可能とし、Twitterとの連携認証後はTwitterをログアウトしない限り、本システムのログイン状態が継続されるようになる。なお、Step4については、本稿の実験では行っていない。

3.2 クラウドソーシングを用いた用例対訳作成

本節では、クラウドソーシングを用いた用例対訳作成手法について述べる。本手法は、マイクロブログによって一定以上の評価を得た用例を機械翻訳によって翻訳し、クラウドソーシングによって評価、訂正文の作成を依頼することで、より正確な用例対訳の作成を行う。

図2に提案手法の流れを示す。本手法は下記の4ステップで構成されている。

Step1 用例の取得

^{*4} Twitter(<https://twitter.com/>): 140字
PIYO(<http://piyo.fc2.com/>): 255字
Haru.fm(<http://haru.fm/>): 150字

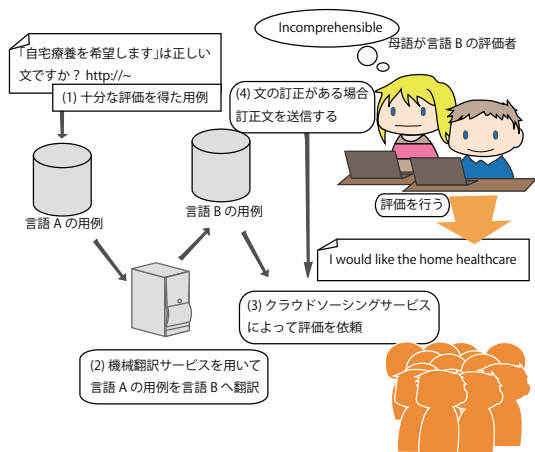


図 2 クラウドソーシングを用いた多言語テキストペア評価手法の流れ



図 4 用例評価を依頼を行う bot の画面例

Please evaluate the following sentence.

Base Sentence : **I is metabolic syndrome?**

Flawless English

- Incomprehensible
- Disfluent English
- Non-native English
- Good English
- Flawless English

If you find any error, correct the sentence.

図 3 作業者に提示されるタスクの例

図 2(1) で、3.1 節で述べた手法により取得した評価データをもとに、十分に正しいと判断された用例を取得する。

Step2 用例の翻訳

Step1 によって取得した用例の言語を言語 A とする。図 2(2) で、言語 A の用例を機械翻訳サービスを用いて言語 B に翻訳し、言語 B の用例を取得する。

Step3 翻訳された用例の評価

図 2(3) で、言語 B の用例のデータを、クラウドソーシングサービスを用いて言語 B を母語とする人に評価を依頼する。評価者は機械翻訳によって言語 B に翻訳された多言語テキストペアを、5 段階で評価し、訂正がある場合は訂正文を入力する。

Step4 訂正文の評価

図 2(4) で、訂正文として得た多言語テキストペアを Step3 と同様にクラウドソーシングサービスから評価を依頼する。

本手法では、評価者の母語と評価する多言語テキストペアの言語が一致するため、評価および訂正が正確に行われる可能性がある。なお、本稿では言語 A を日本語、言語 B を英語とし、評価者の国を United States とした。タスクの画面例を図 3 に示す。文献 [11] の評価基準^{*5}を

^{*5} 評価段階は、1: Incomprehensible, 2: Disfluent English, 3:

用いた。なお、Step4 については、本稿の実験では行っていない。

4. マイクロブログを用いた用例評価実験

本章では、3.1 節で述べた、マイクロブログを用いた用例評価手法の実験について述べる。

4.1 評価用データセット

本節では、評価用のデータセットについて述べる。本実験では、TackPad の日本語の用例のうち、英語へ翻訳されておらず、評価回数が 10 回以上の用例の中からランダムに抽出した 404 文を用いた。この文の内訳は、実験期間中にマイクロブログに投稿された 66 文、評価用の Web ページより表示された 338 文となっている。

4.2 評価データの収集

本節では、マイクロブログを用いた用例の評価データの収集について述べる。本実験では、マイクロブログとして Twitter^{*6}を利用した。Twitter は 140 字以内の文を投稿することのできるマイクロブログサービスである。

評価を依頼する用例の投稿は、自動で投稿を行う bot により 2 時間に 1 度の間隔で行った。この間隔は、作業者が評価を依頼する投稿の数に対し、煩わしく思わない程度の頻度と考えて設定した。本稿では Twitter 上の作業者を Follower と呼ぶ。Follower へは以下の 2 ステップのタスクを依頼した。図 4 に用例評価の依頼を行う bot の画面例を、図 5 に用例評価の画面例をそれぞれ示す。

Step1 図 5(1) で本システムにログインし、図 5(2) で、bot より投稿された用例の評価を行う。図 5(2) で Follower が正しくないと判断した場合には、図 5(3) で、その間違いが用例の綴り、文法、言語選択、入力文字、アクセント記号のいずれに該当するのかが選択する。用例が正しくないと判断し、評価した用

Non-native English, 4: Good English, 5: Flawless English
^{*6} <https://twitter.com/>

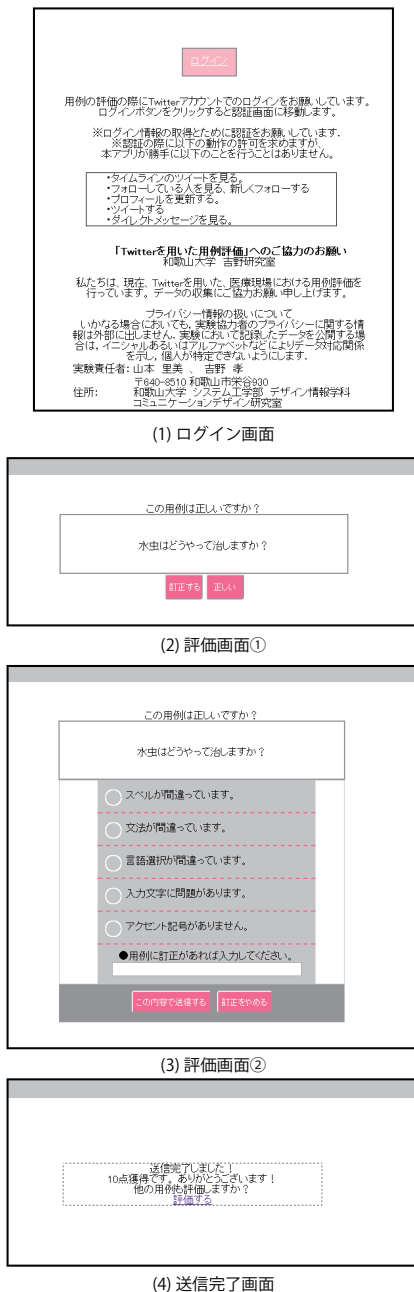


図 5 用例評価の画面例

例に訂正がある場合はより正しい文を入力する。

Step2 評価を終了する。または図 5(4)で、評価結果送信後の画面に表示されるリンクより、評価を連続して行う。なお、連続して評価を行う場合に提示される用例が、4.1 節で述べた評価用の Web ページより表示された文に当たる。また、連続して評価する用例の数に制限は設けていない。

4.3 アンケート調査

本実験では評価データの収集の他に、Follower へのアンケート調査を行った。

4.3.1 アンケート調査の目的

本アンケートを用いて、マイクロブログによる用例評価の有用性について調査を行う。マイクロブログを用い

ることが用例を評価するきっかけになるかどうか、提案手法による用例の評価のタスクがユーザの負担になるかどうかを調べた。なお、アンケートの回答者は Follower のうち男女 10 名 (男: 2 名, 女: 8 名) である。アンケートはシステムの使用開始から 3 日後に行った。

4.3.2 仮説

アンケートの作成にあたり、以下の仮説を立てた。

仮説 1 マイクロブログの利用により、場所や時間に依存しない評価が行えるため、評価数が増加する。マイクロブログは 200 字前後の記事の投稿という特徴から頻繁な投稿や閲覧が行われるが、閲覧に関してはユーザの空いた時間に行われていることが多い。よって、マイクロブログへ定期的に用例評価の依頼文を投稿することで、ユーザの空き時間を利用した用例の評価が行うことが出来ると考えた。スマートフォンなどの携帯端末から評価を行うことが可能なため、時間、場所を制限しない評価を行うことが出来る可能性がある。

仮説 2 評価後、続けて評価を行うことを促すことにより評価数が増加する。

マイクロブログによって用例の評価を促す記事を投稿するが、記事の URL から評価できるのはその記事の 1 文のみである。しかし、評価終了後に次の用例の評価を促す仕組みがあれば、マイクロブログによって与えられたきっかけが継続する。これにより、評価回数の増加につながると考えた。

仮説 3 評価手順を最小限に抑えることが出来れば、ユーザはより多くの評価を行う。

評価の手順が多い場合、ユーザは途中で評価をやめてしまう可能性がある。

5. クラウドソーシングを用いた対訳作成実験

本章では、3.2 節で述べた、クラウドソーシングを用いた対訳作成手法の実験について述べる。

図 6 にクラウドソーシングを用いた多言語テキストペア評価の実験手順を示す。本実験ではまず、評価用のデータセットを機械翻訳により英語に翻訳し、クラウドソーシングによって評価を行う。

5.1 評価用データセット

本節では、評価用のデータセットについて述べる。本実験では、機械翻訳サービス*7により英語に翻訳した多言語テキストペアを用いた。翻訳する用例は、4.2 節により評価された用例の中で、5 回以上の評価がなされ、かつ全ての評価が「正しい」とされた用例とした。

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用例対訳評

*7 本実験では、Google 翻訳を用いた

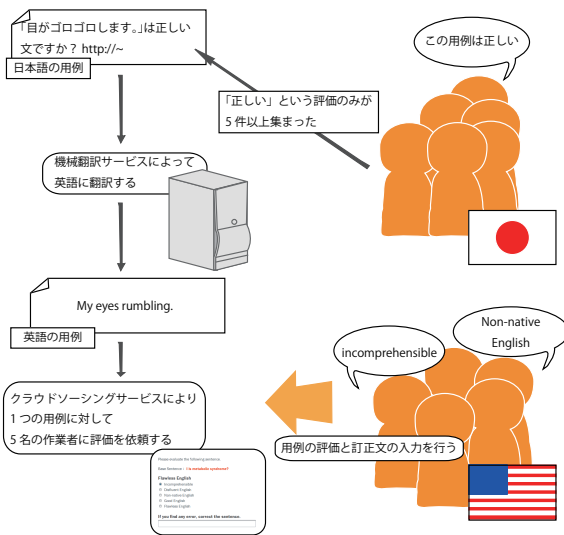


図 6 クラウドソーシングを用いた多言語テキストペア評価の実験手順

価のデータ収集について述べる。本実験では、クラウドソーシングサービスとして CrowdFlower^{*8}を用いた。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、本実験では1つのタスクに対する報酬を10セントと設定し、1つの用例について5人からの評価を得る。なお、評価は CrowdFlower によって定められたレベル1^{*9}の国を United States とする作業者に依頼した。

6. 考察

本章では、6.1 節で4章の用例評価実験の考察を、6.2 節でマイクロブログを用いた評価作業の正確性についての分析と考察を、6.3 節で5章のクラウドソーシングを用いた対訳作成実験の考察をそれぞれ述べる。

6.1 マイクロブログを用いた用例評価手法の有用性

本節では、マイクロブログによって用例評価の依頼を行う有用性についての分析を行う。4.3 節で行ったアンケートの結果、Follower の評価時間、評価回数を用い、4.3.2 項で立てた仮説が正しいかどうかを考察した。

6.1.1 アンケート結果

本項では、Follower に対して行ったアンケートの結果について述べる。

まず、Follower の Twitter の使用頻度についての結果を述べる。アンケートより、普段から1日2回以上 Twitter を利用する人が9名、1日に1回程度利用する人が1名という結果になった。どんな時に Twitter を利用するかという質問については、「気分が上がっていて、何か残したいと思ったとき」という回答があったが、それ以外の回答は「暇なとき」「休憩するとき」「電車での

*8 <http://crowdflower.com/>

*9 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1は最も下のレベルである。

表 1 Twitter を使用する端末について

	PC	スマートフォン	フィーチャーフォン	その他
普段使う端末	8	8	1	0
最も多く使う端末	2	7	1	0
評価に用いた端末	6	7	0	0
最も多く評価に用いた端末	4	6	0	0

移動中」といった、空いた時間に Twitter の閲覧を行うという意見が多かった。また、表1のアンケート回答者の Twitter や用例評価時に使用する端末調査結果からも、携帯端末が活用されていることが分かる。

次に、用例の評価のタスクが Follower の負担になっていたかどうかを調査するための質問の結果を述べる。Follower に対し、(1) 投稿の頻度は適切だったか、(2) 用例の投稿は評価のきっかけになったか、(3) Twitter から評価の Web ページに移動する事を煩わしく感じたか、(4) 評価の手順は多いと感じたか、(5) 評価を連続して行ったかどうかについて質問した^{*10}。質問(5)以外の結果を表2に示す。

表2(1)より、用例の投稿が評価のきっかけになることが分かった。なお、表2(3)に対して「どちらでもない」、「同意する」と回答した人からは、その理由として「スマートフォンで見るのが煩わしい」、「スマートフォンの動きが面倒」、「アカウントとパスワードの入力が面倒」、「Twitter 内で評価が出来ると良いと思った。移動が少し煩わしい」などの意見があった。しかし、評価の手順については多いと感じる人は少なく、「ボタンを押すだけだから」、「手順自体は手間ではない」といった意見があった。このことから、Twitter は用例評価のきっかけになるが、Twitter から評価ページへの移動することが Follower に対して負担になっていることが分かった。また、表2(5)の回答では「連続して評価を行った」と答えたのは9人、「連続して評価を行わなかった」と答えたのは1人であった。連続して評価を行った理由として「1回評価ページを開いたので、他のも評価してみようと思った」、「1度だけ行った。次のページがどうなのかなにか気になったから」、「暇だから」、「評価が簡単で、時間がかからなかったから」、「評価自体はボタンを押すだけだから」という意見があった。これより、本手法で用いたシステムの評価のステップ数は適切だったと言える。

6.1.2 分析結果と考察

本項では、アンケート結果と Follower の評価時間、回数について分析した結果を述べる。

6.1.1 項より、ユーザは空いた時間に Twitter を利用することが多いことが分かった。また、表1より Twitter や本システムの利用はスマートフォンから行われること

*10 質問(1)~(4)の回答は、1:強く同意しない, 2:同意しない, 3:どちらでもない, 4:同意する, 5:強く同意するの5段階とした。

表 2 アンケート結果

質問	評価の分布					中央値	最頻値
	1	2	3	4	5		
(1) 用例の投稿は評価のきっかけになった	0	0	2	6	2	4	4
(2) 用例が投稿される頻度は適切だった	0	1	4	3	2	3.5	3
(3) Twitter から評価の Web ページに移動することを煩わしく感じた	0	3	4	5	0	3	4
(4) 評価の手順は多いと感じた	1	6	1	1	0	2	2

・評価の分布の数字は、評価した人数を示す。

表 3 用例と評価結果

文 ID	用例	評価 (1)		評価 (2)		
		正しい	正しくない	正しい	正しくない	
1	子どもは独立して遠くに住んでいます。	3	1	言語選択 ()	3	0
2	おなかが止まりません。	5	0		3	0
3	留置カテーテル	5	0		2	1
4	私はメタボリックシンドロームですか？	6	0		2	1
5	臓器移植を受ける為の条件はありますか？	5	0		3	0
6	よく聞いてください。	5	0		3	0
7	アレルギー疾患	5	0		3	0
8	落ち着いて聞いてください。	8	0		3	0
9	布団に入っても寝つけません。	6	0		3	0
10	血尿があります。	3	2	文法 ()	2	1
11	爪が剥がれました	6	0		3	0
12	目がゴロゴロします。	6	0		3	0
13	人間ドック・健康診断の検査結果は、検査後どのくらいでわかるのでしょうか	4	0		2	1
14	歩くときは松葉杖を使用してください。	6	0		3	0
15	オクラは風邪の予防に役に立ちます。	3	1	言語選択 ()	3	0
16	目がしばしばする	4	0		2	1
17	息を大きく吸い込んでください	5	0		3	0

・表中の数字は、評価した人数を示す。

・評価 (1) : Follower による評価, 評価 (2) : マイクログを利用しない日本人評価者による評価

・表中の () は、訂正文の入力があったことを示す。

が多い。これより、4.3.2 項の、用例評価の依頼にマイクログを用いることで、時間や場所を制限しない用例評価を行うことが出来るという仮説 1 は成立する。また、表 2 の結果より、用例の投稿は評価のきっかけになるが、Twitter から評価ページへの移動は Follower の負担になっていることが分かった。しかし、用例の評価については Follower への負担は少なく、続けて評価を行ったと回答する人が多かった。また、評価が行われた日時と評価者のデータを見ると、連続して評価を行う Follower が多く、6.1.1 項の質問 (5) に対する回答より、連続した評価を促すことや、評価手順を最小限に抑えることにより、用例を評価してもらえる可能性が高く、仮説 2 および仮説 3 は成立する可能性があるが、今後、長期的な調査および従来手法との比較が必要であると考えている。

6.2 用例評価における正確性

本節では、用例評価における正確性について述べる。

6.2.1 正確性評価の分析方法

本項では、4.2 節で収集した評価結果の正確性を分析する方法について述べる。本分析では、Follower の評価結果と、マイクログを利用しない日本人評価者 3 名による評価結果の比較を行う。

本分析は次の手順で行った。

- (1) Follower によって評価された用例のうち、一定以上の評価回数を得たデータを取得する。
- (2) (1) で取得した用例について、マイクログを利用しない評価者からの評価結果を得る。
- (3) (1) と (2) で取得したデータを比較する。

なお、本実験では用例の評価回数の基準を 4 回以上とした。本分析では (1) で取得した用例の数は 17 文だった。

6.2.2 評価結果の比較

本項では 6.2.1 項で示した評価手法より得た結果について述べる。

表 3 に用例の評価の分析結果を示す。Follower によって評価された結果を評価 (1)、マイクログを利用し

表 4 クラウドソーシングによって評価を行う用例

文 ID	日本語	英語
1	私はメタボリックシンドロームですか？	I is metabolic syndrome?
2	息を大きく吸い込んでください	Please inhale big breath.
3	留置カテーテル	Indwelling catheter
4	布団に入っても寝つけません。	I do not get to sleep into futon.
5	臓器移植を受ける為の条件はありますか？	Is there a condition for receiving an organ transplant?
6	目がゴロゴロします。	My eyes rumbling.
7	おならが止まりません	Fart does not stop.
8	爪が剥がれました	Nail was peeling.
9	落ち着いて聞いてください。	Please listen calmly.
10	歩くときは松葉杖を使用してください。	Please use crutches when walking.

表 5 5段階による評価結果

文 ID	Incomprehensible (理解できない)	Disfluent English (流暢ではない)	Non-native English (非母語英語)	Good English (良い英語)	Flawless English (完璧な英語)
1	0	5	0	0	0
2	0	2	1	2	0
3	2	2	0	0	1
4	0	2	2	1	0
5	0	0	0	1	4
6	2	1	2	0	0
7	0	3	0	2	0
8	1	0	1	3	0
9	1	0	2	0	2
10	1	0	0	0	4

・表中の数字は、人数を示す。

ない日本人評価者による評価結果を評価 (2) とする。数値は評価を行った人数である。また、用例が正しくないと評価された場合は、3.1 節の項目のいずれに該当すると回答されたのかを示した。

表 3 より、本手法を用いた場合の評価と、用いなかった場合の評価では、評価結果に差がほとんど無いことが分かる。また、評価 (1) と評価 (2) の両方で入力された訂正文 (表 3, 文 ID 10) は一致していた^{*13}。これより、マイクロブログを用いて用例の評価の依頼を行った場合も正確な評価が行われていると言える。

6.3 多言語テキストペアの正確性評価

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の正確性評価の分析について述べる。表 4 に使用した多言語テキストペアと機械翻訳の結果を示す。また、評価結果を示すために各文に ID を付与した。各多言語テキストペアへの評価の結果は、表 5 に示す。表 5 の結果より、正確な用例対訳の作成を行うためには、機械翻訳のみの使用では不十分である可能性がある。

作業員から得た訂正文には共通するものがあつた。例えば「布団に入っても寝付けません」に対して機械翻訳を行った “I do not get to sleep well into the futon.” (表 4, 文 ID 4) では、5 名中 4 名が前置詞の “into” を

“on” に変更していた。また、「おならが止まりません」を翻訳した “Fart does not stop.” (表 4, 文 ID 7) では、5 名中 4 名が “fart” に所有代名詞または冠詞を追加していた。このことから、クラウドソーシングに訂正文の入力を依頼することで、これらの文に関しては正しい文が作成できたと考えられる。ただし、「目がゴロゴロします。」といった、擬態語などを含む用例を機械翻訳した場合、作業員に意味が通じていない傾向にあつた。

また、本実験で得られた結果より、訂正文には共通する特徴があり、それによって分類できると考えた。本実験で得た多言語テキストペアの評価と訂正文より、3 つに分類した。表 6 に分類とその例を示す。

分類 (1) は、どの作業員も多言語テキストペアに「流暢ではない」と回答したが、それぞれの訂正文箇所が異なる。分類 (2) は、どの作業員の訂正文も意味に大きな違いはないが、表現が少し異なる。分類 (3) は、訂正文箇所は同じだが、主語が異なる。これらは、作業員が想定する多言語テキストペアが使用される状況がそれぞれ異なっていることが原因だと考えられる。単に機械翻訳された文を提示するだけでは、使用される状況が適切に伝わらない可能性があることがわかつた。文の提示だけでなく、使用される状況やだれが使用するのか (医療従事者、患者) を提示する必要があると考えられる。

*13 「正しくないと評価した 3 人の評価者全員が「血尿が出ます。」と入力していた。

表 6 訂正文による分類

分類	文 ID	入力された訂正文
(1) 訂正箇所が異なる	1	What
		What is metabolic syndrome?
		Do I have metabolic syndrome?
		Am I a metabolic syndrome?
		Do I have a metabolic syndrome?
(2) 訂正文がほぼ等しい	4	I did not sleep well on the futon.
		I do not sleep well in the futon.
		I do not get to sleep well on the futon.
		I do not sleep well on the futon.
		I do not get to sleep well on the futon.
(3) 文の主語が異なる	6	My nail was peeling
		My nail was peeling.
		A nail was peeling.
		The nail was peeling
		Her nail was peeling

7. おわりに

本稿では、マイクロブログとクラウドソーシングを用いた用例評価手法および多言語用例対訳作成手法の提案とその実験を行った。本手法では、マイクロブログによって用例評価の依頼を行うことで、より多くの評価データを取得し、より正確な用例の収集を目指した。また、取得した用例のデータを機械翻訳したのち、クラウドソーシングによって用例対訳の作成を行うことで、より正確な用例対訳の作成を目指した。

本稿の貢献は、以下の2点である。

- (1) マイクロブログを用いた用例評価により評価数の増加を促すことで用例の評価数が増加する可能性を示した。
- (2) クラウドソーシングを用いて多言語用例対訳作成を行う手法を提案し、一部の多言語テキストペアにおいて正確な多言語用例対訳作成ができることを示した。

なお、クラウドソーシングによる多言語テキストペア評価手法では機械翻訳の精度の違いによる影響があると考えられる。よって、今後は翻訳精度の高い機械翻訳サービスを利用することを考えている。また、評価や作成を行う用例が使用される状況を作業者に提示することも必要である。

今後は、Follower を多言語用例対訳共有システム Tack-Pad の利用者限定した場合にも同様の結果が得られるか追加実験を行う。また、擬態語や擬音語などを含む用例に対する評価手法や、訂正文の特徴による分類についての検討を行う。

謝辞 本研究の一部は、独立行政法人科学技術振興機構研究成果 最適展開支援事業 (A-STEP) 探索タイプ「多段クラウドソーシングを活用した多言語用例対訳プラットフォームの構築」の補助を受けた。

参考文献

- [1] 犬飼章：第2回多文化共生の推進に関する意見交換会(宮城県の取り組み事例)(2009年11月6日開催)。
- [2] 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)。
- [3] 福島 拓, 吉野 孝：クラウドソーシング労働者の作業特徴に着目した多言語テキストペアの正確性評価手法, Webとデータベースに関するフォーラム (WebDB Forum 2012), B4-3, pp. 1-8 (2012)。
- [4] 鎌田 健史, 長谷川 大, 佐久田 博司：Twitterにおける専門家判別手法の性能評価, 情報処理学会, 第75回全国大会講演論文集(第3分冊), 1V-5, pp. 105-106 (2013)。
- [5] 山内 崇資, 林 佑樹, 中野 有紀子：日本語解析によるTwitterの感情分析とシーンインデキシングへの応用, 情報処理学会, 第75回全国大会講演論文集(第2分冊), 6R-6, pp. 315-316 (2013)。
- [6] 難波 輝, 南野 謙一, 後藤 裕介, 渡邊 慶和：ペーカリーショップにおけるtwitter広告と連動した販売業務支援システムの提案, 情報処理学会, 第74回全国大会講演論文集(第4分冊), 6ZE-8, pp. 671-672 (2012)。
- [7] 財前 涼, 森嶋 厚行：クラウドソーシングを用いた情報信憑性判断支援のための情報源検索, 情報処理学会, 第74回全国大会講演論文集(第1分冊), 3N-1, pp. 603-604 (2012)。
- [8] 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)。
- [9] 福島 拓, 吉野 孝：クラウドソーシングを用いた画像提示型多言語用例対訳作成手法の提案, 情報処理学会研究報告, 2013-GN-86 (32), pp. 1-7 (2013)。
- [10] 宮嶋 清人, 関 洋平：閲覧者による印象のフィードバックを用いたイラスト投稿者の検索, 情報処理学会研究報告, 2013-DD-89 (10), pp. 1-8 (2013)。
- [11] 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)。