クラウドソーシングを用いた高精度対訳作成のための低品質翻訳の活用

福島 拓[1,a)]　吉野 孝[2,b)]

概要：現在，グローバル化による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情報共有は十分に行われていない．正確な多言語支援が求められる場では，多言語用例対訳が多く用いられている．用例対訳の作成には多言語の知識が必要となるが，多言語話者の人数は少なく，大きな負担がかかっている．そこで本稿では，複数の機械翻訳器と単言語話者を用いて多言語用例対訳候補の作成を行う，高精度対訳作成手法の提案を行う．また，同一文を作成した Worker の数とコサイン類似度を用いた正確性の高い対訳文の抽出を行った．本稿の貢献は以下である．(1) 複数の機械翻訳器とクラウドソーシング上の単言語話者で多言語用例対訳の作成を行う高精度対訳作成手法を提案し，実現した．(2)1 つの機械翻訳器を使用する従来手法よりも複数の機械翻訳器を使用する提案手法の方が，正確な対訳候補文の作成割合が高くなることを示した．(3) 最多対訳候補よりもコサイン類似度を用いた対訳文の抽出の方が，正確な対訳文を多く抽出できることを示した．(4) 翻訳文の重複がある場合は，翻訳文の正確性は高いが対訳候補文の正確性は下がることを示した．

Desterilizing Low-quality Machine Translations for Creating High-quality Parallel Texts Using Crowdsourcing

Taku Fukushima[1,a)]　Takashi Yoshino[2,b)]

1. はじめに

近年の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．訪日外国人数も年々増加傾向にあり，2013 年は過去最高の約 1,126 万人にのぼっている [1]．しかし，すべての訪日外国人が日本語を理解しているとは言い難い．また，一般に多言語を十分に習得することは非常に難しく，母語以外の言語によるコミュニケーションは困難なこともあり [2], [3], [4]，非母語によるコミュニケーションは十分に行うことができない．

非母語によるコミュニケーションで影響が顕著に現れる分野の 1 つに医療がある．医療分野では，わずかなコミュニケーション不足で医療ミスが発生する恐れがあるため，適切な支援が求められている．

そこで，多言語対応の医療支援システムの開発が多く行われている [5], [6], [7], [8]．これらのシステムでは，正確な多言語変換が可能な用例対訳が用いられている．用例対訳とは，用例を多言語に正確に翻訳したコーパスのことを指し，「保険証はお持ちですか？」「はい」「いいえ」などの利用現場で使用される言葉を多言語で提供することができる．この用例対訳を用いて，利用者が適切な質問やその回答を使用することで，正確な多言語対話が可能となる．

また，我々は用例対訳の収集，共有を目的とした多言語用例対訳共有システム TackPad（タックパッド）の開発を行っている [9]．収集した用例対訳は，正確性評価を行った後，多言語対応医療支援システムへの提供を目指している．しかし，本システムでは用例対訳の収集数が十分でないという問題を抱えている．本システムでは多言語の対訳作成は翻訳者が担っているが，翻訳者の人数は少なく，大きな負担がかかっている．

[1] 静岡大学大学院工学研究科
Graduate School of Engineering, Shizuoka University, Hamamatsu 432–8561, Japan
[2] 和歌山大学システム工学部
Faculty of Systems Engineering, Wakayama University, Wakayama 640–8510, Japan
[a)] fukushima@sys.eng.shizuoka.ac.jp
[b)] yoshino@sys.wakayama-u.ac.jp

そこで本稿では，クラウドソーシング [10], [11] を活用する．クラウドソーシングとは，人々（群衆）への作業や業務の委託を指す．本稿では，クラウドソーシング上の労働者を Worker とする．クラウドソーシングでは大量の用例に対して安価で評価依頼を行うことができる利点がある．しかし，クラウドソーシング上で多言語が関係する作業委託を行った場合，特に不正確なものが多く含まれることが分かっている [12], [13], [14]．このため，本稿では多言語による悪影響を減らすために，クラウドソーシングへの作業委託を単言語で行うこととする．その際，機械翻訳器の活用を行う．なお，同様の手法は画像の活用や言い換え表現を活用して従来から行われている [15], [16]．従来手法では，翻訳の際に最も性能の良い機械翻訳のみを使用しているが，高い性能の機械翻訳器が常に良い翻訳結果を返す訳ではない．また，機械翻訳器には得意な文や不得意な文が存在しているが，それらを判定することは難しい．本稿では，翻訳精度に関係なく機械翻訳器を複数使用し，クラウドソーシングへの作業委託を行うことで，高品質な多言語用例対訳候補の取得を目指す．

2. 関連研究

2.1 多言語間コミュニケーション支援

多言語間コミュニケーション支援を目的として，用例対訳を用いた支援技術の研究や，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をすべて多言語に翻訳が可能であるため，子供向けの機械翻訳 [17] や多言語対面環境の討論支援 [18] など，様々な分野で利用されている．しかし，機械翻訳の精度は年々向上しているものの，正確性が求められる医療分野でそのまま利用可能な精度には達していない [19]．また，機械翻訳はルールや統計データに基づいて動的な翻訳を行うため [20]，すべての対訳の正確性を確保することはできない．

そこで現在，正確性が求められる分野においては用例対訳による支援が多く行われている．用例対訳を利用したシステムとして，多言語医療受付支援システム M^3（エムキューブ）[5] や，ケータイ多言語対話システム [6] がある．また，自由文に対応するために，用例対訳と機械翻訳を併用したシステムも提案されている [7], [8]．

このように使用される用例対訳の収集・共有を目的として，我々は多言語用例対訳共有システム TackPad の開発を行っている [9]．TackPad では，(i) 医療従事者や患者などが必要な用例をシステムに登録，(ii) 翻訳者が (i) で登録された用例を各言語に翻訳，(iii) システム利用者が作成された用例対訳の正確性評価を行い，一定の閾値を超えた用例対訳を多言語対応医療システムへ提供する，の手順で，医療現場で求められている用例対訳の収集・共有を Web 上で行っている．現在，本システム上には用例数は全言語合わせて約 15,000 文が存在しているが，用例対訳の数は十分でないことが分かっている [9]．今後，医療分野で必要な用例対訳を網羅した場合，現在の約数十倍の用例が必要であると考えられるが，対訳作成を行う翻訳者への負担が非常に大きくなるという課題を抱えている．

2.2 クラウドソーシングを用いた多言語データ収集

クラウドソーシングを用いた多言語データの収集は多く行われている．Callison-Burch はクラウドソーシングを用いた多言語対の正確性評価を [12]，Negri らはクラウドソーシングを用いた多言語対の作成 [13] をそれぞれ行っている．これらの研究では，多言語話者を対象としており，両言語の文を見せた正確性評価や，一方の言語を提示してもう一方の言語への翻訳の依頼をそれぞれ行っている．また，クラウドソーシング上の不適切に対価を得ようとする Worker を考慮した手法がそれぞれ提案されている．我々も翻訳者の作業特徴を考慮し，作業時間をもとにした多言語対の正確性評価手法を提案した [14]．

しかし，これらの研究では，クラウドソーシング上に少ないと考えられる多言語話者を対象とした作業委託を行っている．2 言語が関係するこれらの作業は比較的難解な作業となる．このため，単純な作業で対価を得ようとする Worker が多いと考えられるクラウドソーシング上では，不適切に対価を得ようとする Worker が存在しており，その対策が求められている．

このため，我々は単言語話者のみを用いて多言語用例対訳の作成を可能とする手法の提案を行った [15]．この手法では，画像と機械翻訳を用いることで，正確な多言語用例対訳の作成を目指している．実験より，クラウドソーシングの複数の Worker が同一文の作成を行った場合に，正確な対訳作成が行われている傾向があることを示した．本稿でも，文献 [15] と同様に単言語話者のみで多言語用例対訳の作成を行う．その際，機械翻訳器を複数使用して提示する翻訳文の多様性を増やすことで，正確な対訳を得ることを目指す．

3. 高品質対訳作成手法と実験概要

3.1 本手法の概要

本手法のコンセプトを**図 1** に示す．本手法は下記の 4 ステップで構成されている．

Step 1：複数の機械翻訳器を用いて翻訳

言語 A の用例を言語 B へ機械翻訳を用いて翻訳する．従来は 1 つの機械翻訳器を用いていたものを複数用いることで，翻訳文の多様性を高めることを目指す．本稿では，言語 A の用例を「原文」，翻訳された文を「翻訳文」とする．

Step 2：クラウドソーシングでの対訳作成

クラウドソーシング上の複数の Worker（言語 B 話者）に，Step 1 で作成した複数の翻訳文（言語 B）を提示

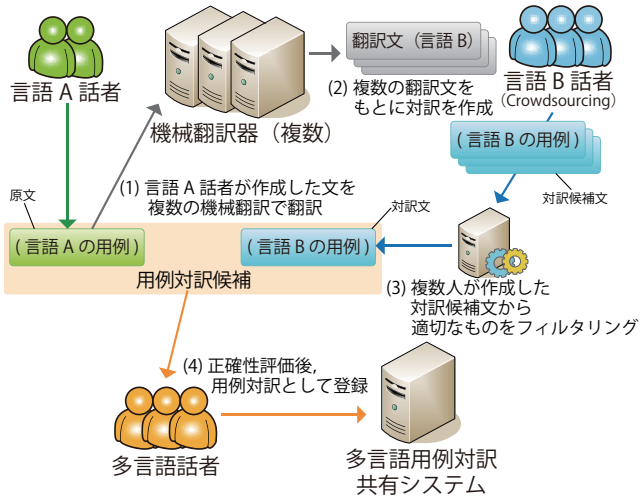

図 1　本手法のコンセプト

Fig. 1　Concept of our method.

する．Worker には提示した翻訳文をもとに，正しい言語 B の文の作成を依頼する．このことで，言語 A を翻訳した言語 B の文（対訳）が取得できる．対訳作成の詳細については 3.2 節で述べる．本稿では，Worker が作成した文を「対訳候補文」とする．

Step 3：対訳候補文のフィルタリング

対訳候補文は，Step 2 の Worker の人数分だけ生成される．このまま Step 4 へ進むと，多言語話者の負担が大きくなるためフィルタリングを行う．なお，フィルタリングの詳細については 3.3 節で述べる．本稿では，フィルタリング後の対訳候補文を「対訳文」とする．

Step 4：正確性評価

Step 3 までは単言語話者のみが関わっているため，正確性を期すために言語 A と言語 B を理解できる多言語話者が最終的な評価を行う．評価対象は，原文（言語 A）と対訳文（言語 B）の正確性である．

3.2　クラウドソーシングでの対訳作成

本節では，3.1 節の Step 2 で述べた，クラウドソーシングでの対訳作成について詳しく述べる．

以下に，クラウドソーシングで行ったタスクを示す．

(1) 前節の Step 2 で作成した複数の翻訳文を提示し，最も流暢性の高い翻訳文の選択を依頼．

(2) (1) で選択した翻訳文の流暢性評価を 5 段階で依頼 [*1]．

(3) 正しい言語 B の文の記入を依頼．ただし，提示した翻訳文が正しいと判断した場合は翻訳文と同一の文の記入を許可した．

3.3　対訳候補文のフィルタリング

本節では，3.1 節の Step 3 で述べた，Worker によって作成された対訳候補文のフィルタリングについて詳しく述べる．本フィルタリングでは，不真面目な Worker により作成された対訳候補文を除去した後，正確性の高い対訳候補文の抽出を行う．

3.3.1　不適切な対訳候補文の除去

関連研究でも述べたとおり，クラウドソーシング上には不適切に対価を得ようとする Worker が存在している．本手法では単言語話者が作業可能なタスクへ分解することで作業の難易度を下げ，不真面目な Worker の減少を目指している．しかし，不真面目な Worker を完全になくすことは難しいと考えられる．このため，本項では不真面目な Worker が作成した対訳候補文の除去を行う．

本稿で不適切な対訳候補文と判定したものを以下に示す．

(1) 同一 Worker が別の翻訳文に対して同一の対訳候補文を記入した場合．

(2) 2 単語以下の対訳候補文で，“not good” や “none”，“jkljlkj” などの明らかに不適切な文．

(3) 翻訳文の意味を推測して説明している文．

(4) 流暢性評価で 3 以下と評価したにもかかわらず，翻訳文をそのまま使用している文．

次章で述べる実験では，上記の内容を著者の一人が調査して除去した．今後は自動化する必要があると考えられる．なお，2 単語以下の対訳候補文はほぼ不正確であったため，(3) 以外についての自動化は可能であると考えられる．

3.3.2　正確性の高い対訳候補文の抽出

本項では前項で除去されなかった対訳候補文の中から，正確性の高いものの抽出を行う．本手法ではまず，下記に示す前処理を行う．なお，次章で述べる実験では (5) のみ著者の一人が作業し，その他の項目は自動で行っている．

(1) 全角のテキストを半角に変換．

(2) ピリオドが存在していない文にピリオドを付与．

(3) 文頭文字および “I” を大文字へ変換．

(4) 文頭，文末にある空白文字の除去．

(5) “or” や “/” によって複数の単語や文が記入されていた場合，1 番目の文または単語を利用．

前処理の後，以下の基準のいずれかに適合する対訳候補文を対訳文として抽出した．

抽出基準 1：最多対訳候補文の抽出

ある原文から作成された対訳候補文のうち，最も多くの Worker が作成した対訳候補文を正確であると判定した．これは，文献 [15] で使用されている手法である．ただし，最多作成人数が 1 人の場合は抽出を行わないこととする．このため，原文 1 文に対して 0 文以上が抽出される．

抽出基準 2：類似度の高い対訳候補文の抽出

対訳候補文同士で単語単位のコサイン類似度を求める．作成した Worker の人数は考慮せずに，コサイン類似度が最も高い対訳候補文を正確であると判定す

[*1] 文献 [21] の評価基準を用いた．評価段階は，1: Incomprehensible, 2: Disfluent English, 3: Non-native English, 4: Good English, 5: Flawless English，である．

表 1　実験データ
Table 1　Data of the experiment.

原文数	50 文
機械翻訳器	J-Server, WEB-Transer, YakushiteNet, Google Translate
1 文あたりの Worker 数	20 人
Worker の報酬	5 文につき 10 セント
翻訳者	3 名

・従来手法の機械翻訳器は，J-Server のみ使用している．

る．これは，抽出基準 1 と同様に，別の Worker 同士が作成した文が類似している場合は正確であると想定して行っている．ただし，もとの翻訳文をそのまま使用している対訳候補文は類似度判定には用いない．これは，機械翻訳器の翻訳精度は十分でないことと，仮に正確な翻訳文であった場合は抽出基準 1 で抽出されると想定して設定した．本項目の抽出では対訳候補文の間の類似度を用いるため，原文 1 文に対して最低 2 文の対訳文が抽出される．

4.　対訳作成実験

本稿では，1 つの機械翻訳器を用いて翻訳文を生成する手法を従来手法，複数の機械翻訳器を用いて翻訳文を生成する手法を提案手法とし，両手法の比較を行う実験を行った．本実験のデータを**表 1** に示す．

本稿では，言語 A（翻訳元言語）を日本語，言語 B（翻訳先言語）を英語とした．使用する原文は，多言語用例対訳共有システム TackPad[9] から，医療従事者に対して患者が使用する原文（日本語）を 50 文抽出した．平均文字数は 14.0 文字，標準偏差は 6.1 文字であった．なお，50 文中 10 文は 20 文字以上の文を選択している．

これらの原文を翻訳する機械翻訳器として，言語グリッド [22] が提供する J-Server, WEB-Transer, YakushiteNet, Google Translate を使用した．なお，従来手法では文献 [23] で最も精度が良いと判定された J-Server を用いた．

また，クラウドソーシングサービスとして CrowdFlower を利用し，アメリカ合衆国在住の Worker を対象として，原文 1 文につき 20 名分を作業依頼した．その際，翻訳文は患者が医療従事者に対して使用する言葉であることを伝えている．タスクの画面例を**図 2** に示す．なお，CrowdFlower 上で従来手法と提案手法を行ったが，同一 Worker が両者のタスクを実行した場合は，一方の作業のみ採用した．また，Worker の報酬として 5 文につき 10 セントを支払った．このため，原文 1 文に対する対訳候補文の取得には手数料を含めて 53.2 セントの費用がかかった．

その後，従来手法と提案手法の正確性を判定するために，日本語と英語を理解できる翻訳者 3 名に 5 段階で日英対（原文と翻訳文および対訳候補文）の適合性評価 [*2] を依頼した．なお，適合性評価の平均が 4 より大きいものを正確，4 以下のものを不正確と判定し，次章の分析で用いた．

[*2] 文献 [21] の評価基準を用いた．評価段階は，1:全く違う意味，2:雰囲気は残っているが元の意味は分からない，3:意味はだいたいつかめる，4:文法などに多少問題があるがだいたい同じ意味，5:同じ意味，である．

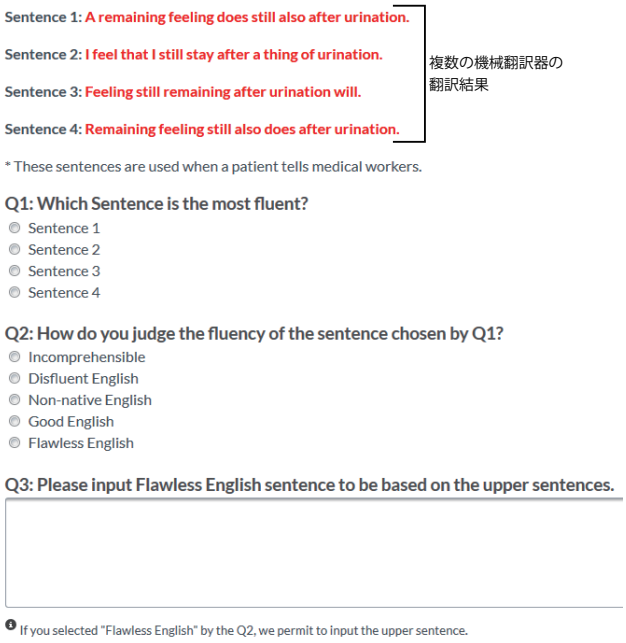

図 2　タスクの例
Fig. 2　Screenshot of a task.

表 2　機械翻訳の翻訳精度
Table 2　Quality of machine translations in the experiment.

機械翻訳器	正確	不正確	正確な翻訳文の割合
J-Server	8	42	16%
WEB-Transer	6	44	12%
YakushiteNet	8	42	16%
Google Translate	8	42	16%

・単位は文である．

5.　実験結果と考察

機械翻訳の翻訳精度を**表 2** に示す．翻訳文そのものの翻訳精度は十分でないことが分かる．以降の各節で本手法の有用性について検証する．

5.1　不適切な対訳候補文のフィルタリング

3.3.1 項のフィルタリングの結果である，不真面目な Worker と不適切な対訳候補文の数を**表 3** に示す．表中の (1)〜(4) は，3.3.1 項に対応している．表 3 の (1)〜(3) より，提案手法を用いることで同一 Worker が別の類似文に対して同一の対訳候補文を作成したり，明らかに不適切な文を入力したりする割合は減少している（41.4% vs 24.2%）．このため，複数の翻訳文を提示することで，対訳候補文の入力がやや容易になったと考えられる．しかし，表 3 の (4) より，提案手法では複数提示した翻訳文に引き

表 3　不真面目な Worker と不適切な対訳候補文の数
Table 3　The number of inadequate workers and translation candidate sentences.

		Worker 人数		対訳候補文数	
		従来	提案	従来	提案
不適切	(1)〜(3)	12 (41.4%)	8 (24.2%)	305	281
	(4)	1 (3.4%)	8 (24.2%)	1	33
	小計	13 (44.8%)	16 (48.5%)	306	314
適切		16 (55.2%)	17 (51.5%)	694	686
総数		29	33	1000	1000

・Worker 人数の列の単位は人，対訳候補文数の列の単位は文である．
・(1)〜(4) は，3.3.1 項の各内容に対応している．
・不適切の行の Worker は，1 文でも不適切な対訳候補文を作成した人数である．
・適切の行の Worker は，すべての対訳候補文が適切であった人数である．

表 4　対訳候補文と対訳文の正確性
Table 4　Accuracy of translation candidate sentences and translation sentences.

	従来	提案
不正確な対訳候補文除去後 (A)	694	686
A のユニーク対訳候補文数	540	476
正確	96	132
正確率 (B)	17.8%	27.7%
正確な対訳候補文を作成できた原文数 (C)	33	41
C の原文全体に占める割合 (D)	66.0%	82.0%
対訳文数	149	142
正確	45	46
正確率 (E)	30.2%	32.4%
対訳文に正解を含む原文数	28	35

・単位は文である．
・原文数は両手法とも 50 文である．

ずられる形で，翻訳文をそのまま使用している傾向が見られる．結果的に，従来手法と提案手法では，種類が異なるもののほぼ同じ数の不適切な対訳候補文が抽出された．本手法設計時は，複数の翻訳文によって対訳候補文を考えやすくなるため，3.3.1 項の (1)〜(3) のような不適切な対訳候補文は減少すると考えていた．しかし，(1)〜(3) は減少したものの，新たに (4)（不適切な翻訳文の使用）の問題が発生するという結果となった．このため，不真面目な Worker を減らすための手法についての検討が今後必要であると考えられる．

5.2　抽出された対訳文の正確性

対訳候補文と対訳文の正確性について**表 4** に示す．表 4(B) より，不正確な対訳候補文除去後の，ユニーク対訳候補文に含まれる正確な対訳候補文の割合から，従来手法よりも提案手法の方が正しい対訳候補文を作成できていることが分かる．また，原文単位で正確な対訳候補文を作成できたかどうか（表 4(C)）を確認した場合も，提案手法の方が多く作成できており，原文全体に占める割合（表 4(D)）も 66.0%（=33/50）から 82.0%（=41/50）まで上昇していることが分かる．これらのことから，最も性能の良い機械翻訳器のみを使用した従来手法よりも，複数の機械翻訳器を使用した提案手法の方が，正しい対訳候補文の作成効果が高いことが分かる．

なお，3.3.1 項の不正確な対訳候補文を除去してから，3.3.2 項で正確な対訳候補文の抽出を行った結果である表 4(E) は，従来手法と提案手法はほぼ同じ性能を示している．これは，3.3.2 項の操作は各手法に依存しておらず，本稿で扱う単言語話者による多言語文の作成において広く適用できる可能性を示唆していると考えられる．

表 5　各抽出基準を用いた場合の正確な対訳文を含む原文数
Table 5　The number of original sentences including accurate translation sentences using each extraction standard.

提案手法		抽出基準 1（最多対訳候補）		計
		正確含む	正確なし	
抽出基準 2（類似度）	正確含む	16	16	32
	正確なし	3	15 (6)	18 (9)
計		19	31 (22)	50 (41)

・表中の各数字は原文数であり，単位は文である．
・括弧内は，対訳候補文に正確なものを含まない原文を除いた原文数である．

次に，3.3.2 項の抽出の詳細について述べる．**表 5** に，提案手法において抽出基準 1（最多対訳候補）を用いた場合と，抽出基準 2（類似度の高い対訳候補）を用いた場合の，正確な対訳文を含む原文数について示す．なお，本実験では原文 50 文中 9 文から正確な対訳候補文は生成されなかった．このため，正確な対訳候補文を生成されなかった原文を除いた値を，表 5 の括弧内に示している．表 5 より，抽出基準 2（類似度の高い対訳候補）の方が多くの正確な対訳文を抽出できていることが分かる（19 文 vs 32 文）．

なお，抽出基準 1 および抽出基準 2 の両者を使用した場合でも，正確な対訳候補文の抽出に失敗しているものが，原文単位で 6 文存在していた．これらの原文からは，平均 11.0 文の対訳候補文が作成され，そのうち正確な対訳候補文は平均 1.3 文であった．正確な対訳候補文が非常に少ないため，抽出に失敗したと考えられる．この問題を解決するために，より多くの正確な対訳候補文を作成する手法の検討が必要であると考えられる．

5.3　正確な対訳候補文を作成できなかった原文

表 6 に，提示した 4 文の翻訳文の重複有無による翻訳精度と対訳候補文の正確性を示す．表 6 では，原文ごとに最も大きな適合性評価であった翻訳文を抽出し，さらにそれを原文単位で正確な対訳候補文を含むか含まないかごとに分けて平均を取ったものである．表 6 より，正確な対訳候補文を作成できなかった原文は，最も正確な翻訳文でも平

表 6　翻訳文の重複有無による対訳候補文の正確性

Table 6　Accuracy of translation candidate sentences by the duplication existence of translated sentences.

	正確対訳候補文		計
	あり	なし	
翻訳文の重複あり	4.33 (5)	3.78 (3)	4.13 (8)
翻訳文の重複なし	3.70 (36)	2.39 (6)	3.52 (42)
計	3.78 (41)	2.85 (9)	3.61 (50)

・表中の数字は，翻訳文の適合性評価平均のうち，最も高い値となったものを原文単位で抽出し，これらを平均したものである．
・適合性評価は 1〜5 の値を取り，4 より大きい場合を正確と判定している．
・括弧内は原文数を示す．

均 2.85 と低いことが分かる．言い換えると，翻訳文の正確性が低いものは正確な対訳候補文を作成することが難しいことを示している．

また，表 6 より，4 つの翻訳文の中に重複があった場合は翻訳文の正確性は高いが，正確な対訳候補文が作成される割合は低くなっていることが分かる（5/8 = 62.5% vs 36/42 = 85.7%）．本手法では同一の翻訳文があっても 4 文全てを Worker に提示していたが，多様性が減少したことによる対訳候補文の作成精度減少の可能性や，Worker が複数の同じ翻訳文に引きずられている可能性が考えられる．

なお，本実験では 10 文の 20 文字以上の原文を使用していたが，他の文との特徴的な差異は見られなかった．

6. まとめ

本稿では，クラウドソーシング上の単言語話者を活用した高精度対訳作成手法の提案を行った．本手法では，複数の機械翻訳器の翻訳結果を提示することで正確性の向上を目指した．

本稿の貢献は以下にまとめられる．

(1) 複数の機械翻訳器とクラウドソーシング上の単言語話者で多言語用例対訳の作成を行う高精度対訳作成手法を提案し，実現した．
(2) 1 つの機械翻訳器を使用する従来手法よりも複数の機械翻訳器を使用する提案手法の方が，正確な対訳候補文の作成割合が高くなることを示した．
(3) 最多対訳候補よりもコサイン類似度を用いた対訳文の抽出の方が，正確な対訳文を多く抽出できることを示した．
(4) 翻訳文の重複がある場合は，翻訳文の正確性は高いが対訳候補文の正確性は下がることを示した．

今後は，本手法で正確な対訳候補文を作成できなかった原文の翻訳手法を検討する．また，正確な対訳候補文の増加を目指す．

謝辞　本研究の一部は JSPS 科研費 (26730105) による．

参考文献

[1] 法務省：平成２５年における外国人入国者数及び日本人出国者数について（確定値），法務省（オンライン），入手先 ⟨http://www.moj.go.jp/nyuukokukanri/kouhou/nyuukokukanri04_00039.html⟩（参照 2014-04-01）．
[2] Takano, Y. and Noda, A.: A temporary decline of thinking ability during foreign language processing, *Journal of Cross-Cultural Psychology*, Vol. 24, pp. 445–462 (1993).
[3] Aiken, M., Hwang, C., Paolillo, J. and Lu, L.: A group decision support system for the Asian Pacific rim, *Journal of International Information Management*, Vol. 3, No. 2, pp. 1–13 (1994).
[4] Kim, K. J. and Bonk, C. J.: Cross-Cultural Comparisons of Online Collaboration, *Journal of Computer Mediated Communication*, Vol. 8, No. 1 (2002).
[5] 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).
[6] 杉田奈未穂，丸田洋輔，長谷川旭，長谷川聡，宮尾　克：ケータイ多言語対話システムとその応用，シンポジウム「モバイル'09」，pp. 63–66 (2009).
[7] 福島　拓，吉野　孝，重野亜久里：用例対訳と機械翻訳を併用した多言語問診票入力手法の提案と評価，情報処理学会論文誌，Vol. 54, No. 1, pp. 256–265 (2013).
[8] 尾崎　俊，松延拓生，吉野　孝，重野亜久里：携帯型多言語間医療対話支援システムの開発と評価，電子情報通信学会技術報告，人工知能と知識処理研究会，Vol. AI2010-47, pp. 19–24 (2011).
[9] 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).
[10] Howe, J.: Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business, Crown Business (2008).
[11] Doan, A., Ramakrishnan, R. and Halevy, A. Y.: Crowdsourcing systems on the World-Wide Web, Vol. 54, No. 4, pp. 86–96 (2011).
[12] Callison-Burch, C.: Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk, *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing(EMNLP2009)*, pp. 286–295 (2009).
[13] Negri, M. and Mehdad, Y.: Creating a Bi-lingual Entailment Corpus through Translations with Mechanical Turk: $100 for a 10-day Rush, *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pp. 212–216 (2010).
[14] 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)，No. B4-3, pp. 1–8 (2012).
[15] 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，グループウェアとネットワークサービス研究会，Vol. 2013-GN-86, No. 34, pp. 1–7 (2013).
[16] 福島　拓，吉野　孝：クラウドソーシングを用いた日本語オノマトペの対訳作成手法の提案，情報処理学会，マルチメディア，分散，協調とモバイル（DICOMO2013）シンポジウム，pp. 989–995 (2013).
[17] Matsuda, M. and Kitamura, Y.: Development of Machine Translation System for Japanese Children, *Proceedings of the 2009 ACM International Workshop*

on Intercultural Collaboration (IWIC'09), pp. 269–271 (2009).

[18] 福島　拓，吉野　孝，喜多千草：共通言語を用いた対面型会議における非母語話者支援システム PaneLive の構築，電子情報通信学会論文誌，Vol. J92-D, No. 6, pp. 719–728 (2009).

[19] 林田尚子，石田　亨：翻訳エージェントによる自己主導型リペア支援の性能予測，電子情報通信学会論文誌，Vol. J88-D1, No. 9, pp. 1459–1466 (2005).

[20] 塚田　元，渡辺太郎，鈴木　潤，永田昌明，磯崎秀樹：統計的機械翻訳，NTT 技術ジャーナル，Vol. 19, No. 6, pp. 23–25 (2007).

[21] Walker, K., Bamba, M., Miller, D., Ma, X., Cieri, C. and Doddington, G.: Multiple-Translation Arabic (MTA) Part 1, *Linguistic Data Consortium, Philadelphia* (2003).

[22] Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration, *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp. 96–100 (2006).

[23] Miyabe, M. and Yoshino, T.: Features of Accuracy Mismatch between Back-translated Sentences and Target-translated senteces, *IEEE 2011 Second International Conference on Culture and Computing*, pp. 65–68 (2011).