

翻訳パズル：クラウドソーシング上における 不完全な翻訳を用いた対訳作成手法

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、単言語話者のみで多言語用例対訳候補の作成を行う、多言語用例対訳作成手法の提案を行う。従来、単言語話者のみを用いる場合は機械翻訳が多く用いられてきたが、正確性が低い翻訳文から正確な対訳文を作成することは困難であった。本提案では、一部の単語のみを多言語辞書を用いて正確に変換し、文の構造をグラフ化した画像を提示することで、正確な多言語対の作成を目指す。また、クラウドソーシング上の労働者に作業委託を行うことで、安価な用例対訳の作成を目指す。本稿では、正確な単語や文の構造を用いた提案手法により、不真面目な Worker の減少の可能性や正確な対訳候補作成の可能性を示す。

キーワード：クラウドソーシング、用例対訳、単言語話者、多言語辞書、グラフ化

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。

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 と呼ぶ。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。単言語話者を対象とした多言語翻訳は、画像や言い換え表現と機械翻訳とを組み合わせたり、複数の機械翻訳器を用いたりする手法が提案されている [15], [16], [17]。しかし、機械翻訳を用いた場合は

¹ 静岡大学大学院工学研究科

² 和歌山大学システム工学部

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

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

機械翻訳の翻訳精度に対訳の精度が依存する。このため、精度が低い機械翻訳文をもとにした場合、正しい対訳作成の可能性が低くなっていた。本稿では、文の構造と正確に翻訳された単語を使用した「翻訳パズル」を提案する。本手法では、多言語辞書を用いて単語を正確に変換し、文の構造をグラフ化することで正確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 [18] や多言語対面環境の討論支援 [19] 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 [20]。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 [21]、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訳の正確性評価を [12]、Negri らはクラウドソーシングを用いた多言語対訳の作成 [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訳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], [16], [17]。これらの手法では、画像と機械翻訳、言い換え表現と機械翻訳、複数の機械翻訳をそれぞれ用いることで、正確な多言語用例対訳の作成を目指している。しかし、これらの手法は機械翻訳の翻訳精度に依存するため、精度の低い機械翻訳文からは正しい対訳文を取得できる可能性が低かった。また、作業タスクが比較的単純であるため、不適切に対価を得ようとする Worker が比較的多く存在していた。

本稿では、正確に翻訳が可能で多言語辞書を用いることで正確性の向上を目指す。また、クラウドソーシングのタスクにおいてゲーム性を持たせた場合、不適切に対価を得ようとする Worker が減少する可能性が示されている [22]。そこで本稿では、グラフ化した文の構造を Worker に提示し、Worker にパズルを解くような感覚を与えることで、不適切に対価を得ようとする Worker 数の減少を目指す。

3. 翻訳パズル

本手法のコンセプトを図 1 に示す。本手法は、形態素解析器と係り受け解析器、多言語辞書を用いたグラフ化、クラウドソーシングでの対訳作成、フィルタリング、正確性評価で構成されている。以降の各節で言語 A の用例を言語 B へ翻訳する流れの詳細について述べる。

3.1 形態素解析器と係り受け解析器を用いたグラフ化

本節では、形態素解析器と係り受け解析器を用いたグラフ化について述べる。本節は図 1(1) の内容にあたる。グラフ化は以下の 3 ステップで構成されている。

Step 1: 形態素解析器を用いて形態素に分割

言語 A 話者が作成した用例を形態素解析器を用いて単語に分割する。本稿では言語 A 話者が作成した用例を「原文」とする。本手法では、以下の 4 品詞のみ Step 2 以降で用いることとした。

- 名詞 (直前が名詞でない接尾語, 助動詞語幹を除く)
- 動詞 (非自立を除く)
- 形容詞

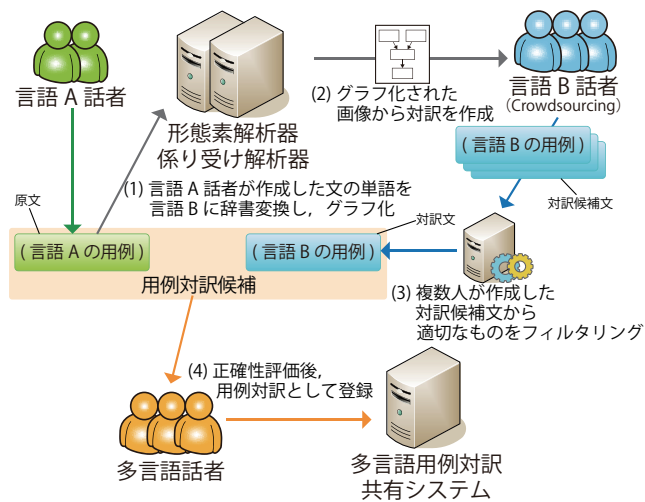


図 1 本手法のコンセプト

● 副詞

また、否定形や疑問形を表現するために下記の処理を行った。

- 終助詞「か」の場合，“?”を付与
- 動詞が未然形かつ、対象動詞の次の単語が助動詞の場合，“NOT”を付与
- 助動詞が未然形かつ、対象助動詞の前の単語が動詞の場合，“NOT”を付与

Step 2 : 多言語辞書を用いた多言語変換

Step 1 で取得した 4 品詞を、下記の手順で多言語変換を行った。

- (1) 多言語辞書を用いて言語 A の単語を言語 B の単語へ辞書変換を行った。単語検索は完全一致検索で行った。なお、動詞と形容詞は原形を用いた。また、名詞が連続しているときは複合語として別途検索した。
- (2) (1) で変換失敗した名詞（複合語を除く）は機械翻訳で翻訳したものを使用した。
- (3) (1) で変換失敗した動詞は以下の手順を繰り返して辞書に存在する単語（動詞）となるまで変換し、使用した。
 - (a) 単語の最後の 1 文字を削る
 - (b) 語尾を「ウ列」にして前方一致検索
 - (c) 語尾を元に戻して前方一致検索
 例として「押さえつける」の場合、(a) で「押さえつけ」とし、(b) で「け」をウ列にして「押さえつく」で前方一致検索を行う。単語が存在しない場合は (c) で語尾を戻して「押さえつけ」で前方一致検索を行う。それでも存在しない場合はさらに (a) に戻って「押さえつ」とする。このように、単語（動詞）が見つかるまで検索を行う形を取った。

Step 3 : 係り受け解析器を用いたグラフ化

係り受け解析器を用いて原文の文の構造を得る。その後、以下の手順でグラフ化を行った。

- (1) 原文を係り受け解析し、係り受け解析結果中の単語を Step 2 の辞書変換結果と置き換える。なお、名詞の複合語に関しては Step 2 で複合語の単語が存在した場合に優先的に使用した。
- (2) Step 1 の “?” および “NOT” を配置する。なお，“NOT” は動詞の前に表示する形を取った。
- (3) 4 品詞以外の単語を除去し、空白となったチャンクも除去する。
- (4) “NOT” が 2 つ連続する場合、強い肯定と捉えて両方の “NOT” を除去する。
- (5) 係り受け関係をもとにチャンク間に矢印を引き、グラフ化する。

3.2 クラウドソーシングでの対訳作成

本節では、3.1 節で作成したグラフをクラウドソーシング上の複数の Worker（言語 B 話者）に提示し、対訳作成を行う方法について述べる。本節は図 1(2) の内容にあたる。

本節では、グラフを Worker に提示し、以下に示すタスクを依頼する。本節の内容により、言語 A を翻訳した言語 B の文（対訳）が取得できる。以降、Worker が作成した文を「対訳候補文」とする。

- (1) グラフをもとに言語 B の正しい文の記入を依頼。
- (2) (1) で入力した文の確信度を 5 段階で依頼 *1。
- (3) タスクに関するコメント（任意項目）。

3.3 対訳候補文のフィルタリング

本節では、3.2 節で作成された対訳候補文のフィルタリングについて述べる。本節は図 1(3) の内容にあたる。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。これは、次節で述べる多言語話者による正確性評価の負担を軽減するために行う。

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) 3 単語以下の対訳候補文で，“not good” や “none”，

*1 評価段階は、1: Poor, 2: Fair, 3: Average, 4: Good, 5: Excellent, である。

表 1 実験データ

	従来手法 1	従来手法 2	提案手法
提示内容	1 つの MT 文	4 つの MT 文	グラフ
原文数	50 文		
1 文あたりの Worker 数	20 人		
Worker の報酬 (5 文につき)	10 セント	25 セント	
翻訳者	3 人		

・ MT 文は機械翻訳文を示す。

“kjlkj”などの明らかに不適切な文。

次章で述べる実験では、(2)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。

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。

- (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。

前処理の後、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、抽出した。これは、文献 [15] や文献 [17] 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。

3.4 対訳文の正確性評価

前節までの内容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。

4. 対訳作成実験

本稿では、前章で述べた手法を提案手法、文献 [17] で行われた、1 つの機械翻訳器を用いて翻訳文を生成し、Worker に提示する手法を従来手法 1、複数の機械翻訳器を用いて翻訳文を生成し、Worker に提示する手法を従来手法 2 とする。本章では、上記の 3 手法を比較する実験について述べる。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。なお、文献 [17] と同じ原文群を用いている。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。

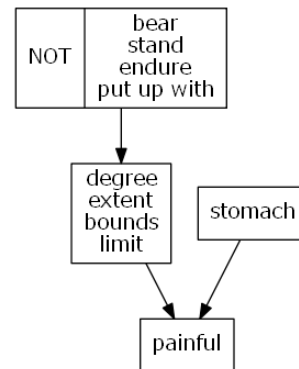


図 2 グラフの例

本稿では形態素解析器として ChaSen^{*2} を、多言語辞書として「デ辞蔵^{*3}」の EDICT 和英辞典を、機械翻訳器として言語グリッド [23] が提供する J-Server を、係り受け解析器として CaboCha^{*4} を、グラフ化には Graphviz^{*5} をそれぞれ使用した。図 2 に作成されたグラフの例を示す。また、従来手法 1 では J-Server を、従来手法 2 では J-Server, WEB-Transter, YakushiteNet, Google Translate を機械翻訳器としてそれぞれ使用している。

なお、一部の単語^{*6}の形態素解析に失敗したが、形態素解析器の辞書への追加により改善が可能であるため、本稿では正しく形態素解析が出来たものとして扱う。また、“デ辞蔵”から取得した動詞が“to V”の形となっていたため、“to”を取り除く処理を加えている。

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、グラフは患者が医療従事者に対して使用する言葉をもとにして作成したことを伝えている。Worker の報酬は 5 文につき 25 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 1.35 ドルの費用がかかった。なお、従来手法は 5 文につき 10 セントであった。これは、従来手法よりも提案手法の方が負荷の高いタスクであると考えたためである。

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^{*7}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} <http://chasen-legacy.sourceforge.jp/>

^{*3} <http://dejizo.jp/dev/>

^{*4} <https://code.google.com/p/cabocha/>

^{*5} <http://www.graphviz.org/>

^{*6} ふくらはぎ, できもの

^{*7} 文献 [24] 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。

表 2 不真面目な Worker と不適切な対訳候補文の数

	Worker 人数			対訳候補文数		
	従来 1	従来 2	提案	従来 1	従来 2	提案
適切	16	17	36	694	686	855
不適切	13	16	8	306	314	145
総数	29	33	44	1000	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。
- ・ 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 3 対訳候補文と対訳文の正確性

	従来 1	従来 2	提案
不正確な対訳候補文除去後 (A)	694	686	855
A のユニーク対訳候補文数	540	476	830
正確な対訳候補文 (B)	96	132	64
正確な対訳候補文を作成できた原文数 (C)	33	41	28

- ・ 単位は文である。
- ・ 原文数は各手法とも 50 文である。

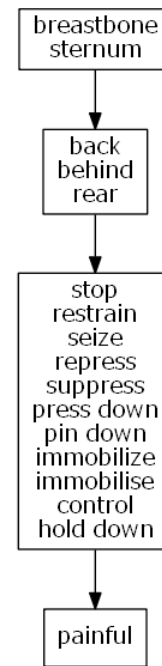


図 3 ID41 の用例から作成されたグラフ

5. 実験結果と考察

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 2 に示す。表 2 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。これは、グラフ化することで Worker にパズルを解くような感覚を与えたこと、報酬の金額が増加したこと（1 文あたり 2 セントから 5 セント）が影響していると考えられる。

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 3 に示す。表 3(B) より、提案手法は従来手法よりも正確な対訳候補文を作ることができていないことが分かる。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 3(C)）を確認した場合も、提案手法の方が作成数が減少していることが分かる。

以降、正確な対訳候補が作成できた原文と作成できなかった原文に分けて考察する。

5.2.1 正確な対訳候補を従来手法よりも作成できた原文

従来手法 2（複数の機械翻訳）よりも多く正しい対訳候補文を作成できた原文は 50 文中 5 文存在していた。例として、ID 41 の「胸骨の後ろが押さえつけられるように痛いです」がある。提示したグラフを図 3 に示す。機械翻訳では、“It is a painful behind the sternum to be pressed” など、翻訳過程で構造が変化し、意図がくみ取りにくい内容となっていた。提案手法では構造が維持されているため

理解が比較的行いやすい形になっていたと考えられる。このことから、提示方法によっては正しい文を作成可能になると考えられる。

また、ID 4 の「頭重感があります」は、従来手法 2 と同じ数の正しい対訳候補文が作成できた。従来手法 2 では、“There is a feeling of sluggishness” や “There is a heaviness of the head” などの機械翻訳が提示されていた。後者は正しい英文（翻訳者評価平均が 4.33）であるが、Worker らの多くは前者を参考に対訳候補文の作成を行っていた。このように、機械翻訳器によって異なる意味に翻訳されていた場合、正しい対訳候補の作成に悪影響を及ぼす場合がある。提案手法では、「頭重感」を “heavy-headed” “haughty” と変換しており、より正しい英文 (“I feel heavy-headed”, 翻訳者評価平均が 4.67) の作成を行うことができた。この点は提案手法の優位な点であると考えられる。

5.2.2 正確な対訳候補が作成できなかった原文

本手法で正確な対訳候補文を作成できなかった理由として、翻訳に必要な情報が欠落していたことが考えられる。例として、ID 8 の「足の指が動きません」がある。この文を用いてグラフ化を行うときは、「足」「指」「NOT」「動く」を使用していたが、“My foot and finger cannot move” など、足と指の関係を誤った内容となっていた。提示したグラフでは “foot” → “finger” → “NOT move” のような形となり、「足の指」なのか「足と指」なのかが判断が行いにくい状況になっていたためであると考えられる。

他の例として、ID 9 の「虫が飛んでいるように見えます」がある。この文を用いてグラフ化を行うときは、「虫」「飛ぶ」「見える」のみを使用していたため、“The insect is

表 4 グラフ提示単語群数と対訳候補文の単語数の差と正解率

利用割合	< 0	< 3	< 6	< 9	< 12	12 ≤
正確	0	18	31	10	5	0
不正確	23	147	219	153	91	133
正確な割合	0.0%	10.9%	12.4%	6.1%	5.2%	0.0%

・利用割合は区間の上限値を示す。
 ・単位は文である。

flying”など、別の意味を示す対訳候補文が作成される場合が多かった。また、同様に ID 29 の「脈が飛びます」、ID 39 の「胸がえぐられるように痛いです」など、比喩的な表現が含まれる文についても正しい対訳候補文の作成が難しい傾向にあった。

また、ID 16 の「目が充血しています」は「目」「充血」「する」を辞書変換している。「する」を辞書変換した場合、“do” “try” “attempt” など 19 単語が提示されるが、この文における「する」はあまり翻訳の必要がない。しかし、これらの単語を Worker が利用しようとしたために不正確になる場合が存在していた。

これらのことから、一部の助詞の提示やグラフ化への反映、比喩的表現への対応、Worker を混乱させる項目の削除などの必要があると考えられる。

5.3 正確な文の抽出指標

本節では対訳候補文から正確な対訳文と不正確な対訳文を分類可能な指標について調査を行う。なお、3.2 節で Worker に確認した確信度と対訳候補文の正確性について相関関係を調べたところ、相関係数は 0.30 であった。弱い相関は見られるものの抽出に使用可能な指標としては使う事が難しいと考えられる。

5.3.1 複数の Worker が同一文を作成

本指標は、文献 [15] や文献 [17] で使用されたものである。本実験では、12 文が複数の Worker から作成され、そのうちの 10 文が正確な対訳であった。特に、3 人以上が作成した同一文 (6 文) はすべて正確なものであった。

従来手法では機械翻訳 (一通り作成された文) を提示していたため、従来よりも同一文が作成される割合が少なくなっている (12/830 = 1.4%。文献 [17] では 84/476 = 17.6%)。しかし、従来手法では (不正確な) 機械翻訳をそのまま使用することが可能であるのに対し、提案手法では完成された文を提示していないため、同一文の作成が困難になっている。このため、適当に作成された文同士が一致する確率が下がるため、本指標を用いた抽出結果は正確な割合が高くなっている (10/12 = 83.3%。文献 [17] では 35/84 = 41.7%)。これらのことから、本手法でも本指標は利用可能であることが分かる。

5.3.2 グラフ提示単語群数と対訳候補文の単語数

表 4 に、対訳候補文の正確、不正確を、Worker が作成した対訳候補文の単語数とグラフで提示した単語群数との

表 5 グラフ化で提示した単語群の利用割合と正解率

利用割合	< 0.33	< 0.5	< 0.67	≤ 1
正確	2	5	8	49
不正確	59	88	102	517
正確な割合	3.3%	5.4%	7.3%	8.7%

・利用割合は区間の上限値を示す。
 ・単位は文である。

差で分類したものを示す。この値は、Worker が作成した対訳候補文の単語数と、“NOT” “?” を除くグラフで提示した単語群の数との差を用いてる。図 2 の場合、“NOT” を除くと 4 単語 (群) のため、Worker が作成した対訳候補文が 7 単語であった場合、使用する値は 3 となる。

表 4 より、差がマイナス (対訳候補文長が提示した単語群数よりも短い) の場合はすべて不正確な文と判定されていることが分かる。また、差が正の値であっても値が大きい場合は不正確と判定されていることが分かる。このことから、対訳候補文の単語数とグラフで提示した単語群数との差は分類時に利用可能な指標の一つになると考えられる。

なお、対訳候補文の単語数とグラフで提示した単語群数との割合についても調査したが、上記で提示した差の方が判定基準として有用な結果が出た。例としては、グラフの単語群数が 5 単語の場合は 300% 以上になると正確な対訳候補が含まれていなかったが、3 単語の場合は 467% のものにも正確な対訳候補が含まれていた。これは、今回のタスクでは 4 品詞以外の単語を追加して正しい英文を作成する必要があるが、追加する単語数は提示した 4 品詞の数には依存しないことを示している。

5.3.3 グラフ化で提示した単語群の利用割合

表 5 に、対訳候補文の正確、不正確を、グラフ化で提示した単語群が Worker 作成の対訳候補文に含まれる割合で分類したものを示す。この割合は “NOT” “?” を除くグラフで提示した単語群のうち、どれだけの単語が対訳候補文で使用されたかを示した値となっている。図 2 の場合、“bear” “stand” “endure” “put up with” から 1 つ以上、“degree” “extent” “bounds” “limit” から 1 つ以上、“stomach” および “painful” の両単語を使用した場合、100% となる。

表 5 より、利用割合が高いほど正確な対訳が作成できていることが分かる。ただし、利用割合が低いものでも正しい対訳文が作成できている場合がある。これは、グラフを参考にして、品詞を変更したり、言い換え表現を利用しているためであると考えられる。

しかしながら、利用割合が 1/3 以下になった場合はほぼ不正確となるため、グラフで提示した単語群の利用割合は分類時に利用可能な指標の一つになると考えられる。

6. おわりに

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対訳作成手法の提案を行った。本手法では、多言語辞書

で変換した単語をグラフ化し、画像として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) 多言語辞書と係り受け解析器を用いて作成したグラフをクラウドソーシング上の単言語話者に提示することで多言語用例対訳の作成を行う対訳作成手法を提案し、実現した。
- (2) 提案手法により、不真面目な Worker の減少の可能性を示した。
- (3) 正確な単語や文の構造を用いた提案手法により、正確な対訳候補作成の可能性を示した。
- (4) 提案手法により作成された対訳候補文から正確な対訳文の抽出を行う指標を提示した。

今後は、本実験で明らかになった課題の解決を行い、新たなグラフ化手法を用いた文の構造の提示手法を検討する。また、従来手法と組み合わせることにより、さらなる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: 携帯型多言語問診票対話支援システムの開発と評価, 電子情報通信学会技術研究報告, 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, *Communications of the ACM*, 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 (EMNLP 2009)*, 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. 32, 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ベ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] 福島 拓, 吉野 孝: 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, 情報処理学会, マルチメディア, 分散, 協調とモバイル (DICOMO2014) シンポジウム, pp. 1209-1215 (2014).
- [18] 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).
- [19] 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).
- [20] 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).
- [21] 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).
- [22] Eickhoff, C., Harris, C. G., de Vries, A. P. and Srinivasan, P.: Quality through Flow and Immersion: Gamifying Crowdsourced Relevance Assessments, *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 871-880 (2012).
- [23] Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration, *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp. 96-100 (2006).
- [24] Walker, K., Bamba, M., Miller, D., Ma, X., Cieri, C. and Doddington, G.: Multiple-Translation Arabic (MTA) Part 1, *Linguistic Data Consortium, Philadelphia* (2003).