

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

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. 本提案では, 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, 安価に用例対訳の作成を目指す. 本稿の貢献は, 画像を用いることで, 単言語話者のみで機械翻訳よりも高精度の多言語用例対訳候補の作成が可能であることを示した点である.

キーワード: クラウドソーシング, 用例対訳, 多言語翻訳, 単言語話者

Proposal of Image-based Multilingual Parallel-text Creation Method Using Crowdsourcing

TAKU FUKUSHIMA^{1,a)} TAKASHI YOSHINO^{2,b)}

Abstract: Recently, worldwide globalization has helped to increase communication among people with different native languages. However, it is not enough that multilingual accurate information sharing. In general, multilingual support systems for applications that require high accuracy use multilingual parallel texts. Multilingual parallel-text creators require the multilingual knowledge. However, the number of multilingual speakers is few. Multilingual parallel-text creators are imposed a big burden. Therefore, we have proposed multilingual parallel-text creation method that creates by only monolingual speakers. Our method creates accurate multilingual texts pair using images. Moreover, our method commissions workers in the crowdsourcing to create at a low price. The contribution of this paper is the following. Our method can create more accuracy parallel-text candidates than the existing machine translation.

Keywords: crowdsourcing, parallel-text, multilingual translation, monolingual speaker

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. 日本国内でも在日外国人
人数や留学生数, 訪日外国人数は 10 年前のそれぞれ約 1.3
倍, 約 1.4 倍, 約 1.2 倍と増加傾向にあり [1], [2], [3], 今
後, 外国人住民のさらなる増加が予想されている [4]. この
ため, 政府内でも多文化共生の推進に関する研究会が開か
れており [4], 今後, 多文化共生社会になると考えられる.
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 [5], [6], [7], 日本語を理解できない外国人と日本

¹ 和歌山大学大学院システム工学研究科
Graduate School of Systems Engineering, Wakayama Uni-
versity, Wakayama 640-8510, 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@yoslab.net

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

人との間で正確な情報共有を十分に行うことはできない。

日本語を理解できないことの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。特に、日本語が通じない外国人と日本人の医療従事者間でのやり取りは、意思の疎通を十分に行うことができない。現在、日本語を理解できない外国人の支援は医療通訳者が行っているが、医療通訳者は慢性的な人員不足となっている。また、通訳者の身分保障や通訳者自身のメンタルケアなどの問題が存在している [8]。

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 [9], [10], [11], [12]。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 [13]。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対訳の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、単言語話者のみで多言語用例対訳候補の作成を可能とする手法を提案する。その際、画像を媒体とすることで、正確な対訳作成を目指す。また、クラウドソーシング [14], [15] 上で単言語話者へ作業委託を行う。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。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 [16], [17], [18]。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。

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 [19] や多言語対面環境の討論支援 [20] 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 [21]。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 [22]、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) [9] や、ケータイ多言語対話システム [10] がある。 M^3 はタッチパネルで操作可能としたシステムで、対話機能、外国人患者の受診支援機能 (問診機能、受診科選択機能など) を有している。また、ケータイ多言語対話システムは多言語問診を携帯電話上で実現している。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 [11], [12]。

このような用例対訳の収集・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 [13]。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 14,000 文が存在しているが、用例対訳の数は不十分であることが分かっている [13]。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。

Hu らは単言語話者間で機械翻訳を介して適切な多言語対訳作成の試みを行っている [23]。Hu らは複数回、単言語話者間で文の修正作業を行っている。また、画像は単語のアノテーションとして用いている。我々のアプローチは、画像をもとに原文を作成する。また、一度のみの機械翻訳で対訳文の作成を行う。これらが Hu らの手法と異なっている点である。

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 [16]、Negri らはクラウドソーシングを用いた多言語対訳の作成 [17] 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、不適切に対価を得ようとするクラウドソーシング上の労働者が存在することを考慮した手法がそれぞれ提案されている。また、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 [18]。

しかし、これらの研究では、クラウドソーシング上に少

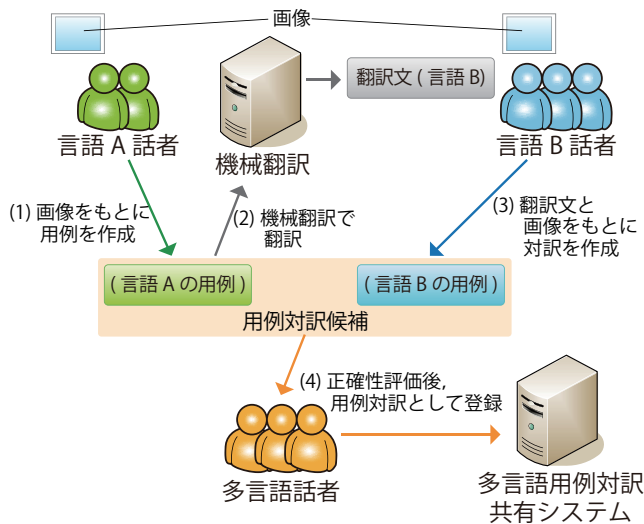


Fig. 1 Concept of our method.

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労働者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労働者が多く作業を行う可能性が考えられる。

このように、多言語が関係する作業委託は、単言語話者をはじめとした不適切な作業を行う労働者が多く存在する問題を抱えている。そこで我々は、多言語の作業タスクを単言語に分割して作業委託を行うことで、対象となる労働者数を増やすこととした。さらに、従来よりも単純な作業とすることで、不適切な作業者を減らすことを目指す。

3. 画像提示型多言語用例対訳作成手法と実験概要

本章では、提案する画像提示型多言語用例対訳作成手法について述べる。本手法は、画像を用いることで、単言語話者のみで多言語対訳の作成を行う。まず、本手法の概要について述べた後、日本語話者による用例の作成、英語話者による対訳の作成、用例対訳候補の正確性判定方法についてそれぞれ述べる。

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 用例の作成

図 1(1) で、言語 A を母語とする利用者が画像をもとに用例の作成を行う。

Step 2 翻訳

図 1(2) で、言語 A の用例を機械翻訳で言語 B へ翻訳を行う。

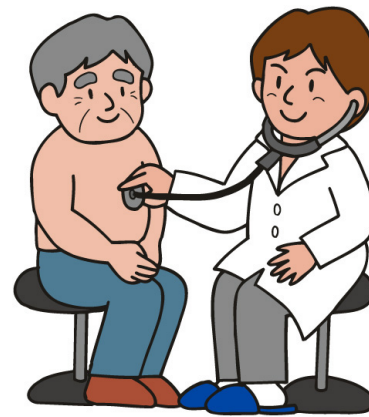


図 2 実験で使った画像の例

Fig. 2 Example of an artwork using in the experiment.
(参照: <http://design.taiho.co.jp/>)

Step 3 対訳の作成

図 1(3) で、言語 B を母語とする利用者が翻訳文の正確性判定と、機械翻訳文をもとに正しい言語 B の用例作成（手法上は対訳作成）を行う。この時、図 1(1) で言語 A 話者へ提示した画像も合わせて表示することで、適切性を高めた言語 B の用例作成を行う。また、言語 B 話者を複数人用いて、言語 B の用例が複数作成されるようにする。このとき、正しい用例の場合は複数人が作成すると考えられる。このため、本手法では複数人が全く同じ用例を作成した場合、その用例は他の用例より正確であると判断することとする。

Step 4 正確性評価

本手法では、単言語話者のみで用例対訳の作成を行っている。このため、実際に作成された用例対訳を使用するときは、図 1(4) のように、両言語を理解する多言語話者が確認する必要がある。

なお、本稿では元言語（言語 A）を日本語、翻訳先言語（言語 B）を英語として、手法の有用性確認を行う。また、英語話者はクラウドソーシング上で作業委託を行った。

3.2 用例の作成と翻訳

本節では、図 1(1) で示した用例の作成について述べる。用例の作成は下記の手順で行った。

(1) 画像を用いた用例作成

日本人用例作成者 2 名に医療に関する画像（イラスト）を提示し、その画像をもとに日本語用例を作成するよう依頼した。実験で使った画像の例を図 2 に示す*1。なお、日本人用例作成者は多言語用例対訳共有システム TackPad[13] で活発に用例登録を行っていた日本人用例作成者から選定した。その際、各作成者に 10 個の画像を提示し、それぞれの画像に対して 5 文字ずつ用例を作成するよう依頼した。この結果、用例作成

*1 <http://design.taiho.co.jp/>の画像を利用した。

者1名につき50文の用例を収集し、計100文の日本語用例を得た。なお、平均文字長は12.0文字、標準偏差は3.26文字であった。

(2) 用例の正確性評価

日本人用例評価者3名に、(1)で用例作成者が作成した日本語用例100文の正確性評価を依頼した。また、不正確なものは正しい文を記入するように依頼した。その後、評価結果をもとに、著者の一人が用例の修正を行った。なお、用例作成の後に用例の正確性評価を行うことは、多言語用例対訳共有システム[13]と同様の流れである。

(3) 翻訳

日本語用例100文を機械翻訳でそれぞれ英語に翻訳を行った(図1(2))。以降、機械翻訳で翻訳した文を「翻訳文」とする。本稿では、機械翻訳エンジンとして言語グリッド[24]のJ-Serverを利用した。

以降、機械翻訳による対訳作成を「既存手法1」とし、日本語用例と翻訳した英文の100対を提案手法と比較を行う基準とする。

3.3 対訳の作成

本節では、図1(3)で示した対訳の作成について述べる。本稿は、英語の対訳作成の場として、クラウドソーシングを用いた。なお、クラウドソーシングサービスとしてCrowdFlower*2を介してAmazon Mechanical Turk*3を利用した。また、用例の作成(図1(1))で使用した画像を提示するグループと提示しないグループの2つに分けて作業依頼を行っている。以降、画像を提示したグループの結果を「提案手法」、画像を提示しなかったグループの結果を「既存手法2」とする。また、作成された文を「英文」と表記する。

以下に、クラウドソーシングで行ったタスクを示す。また、タスクの画面例を図3に示す。

(1) 前節で作成した機械翻訳の英文を提示し、5段階で英文の流暢性評価*4を依頼した。

(2) 正しい英文の記入を依頼した。ただし、提示した英文が正しいと判断した場合は類似文の記入を依頼した。

なお、1対の評価と対訳作成につき5セント*5支払った*6。また、タスクは両グループとも英文100文の評価を10人ずつ行うようにした。

なお、質問の前にQualificationと呼ばれる資格審査を行っている。これは、CrowdFlowerの機能の一つであり、

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

*3 <https://www.mturk.com/>

*4 文献[25]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。

*5 約4.2円(2012/12/12現在。1ドル=83.25円で計算.)。

*6 クラウドソーシングを用いた研究では、コストについても議論を行う場合が多いため、従来研究にならって金額を記載している。



Base Sentence: I don't have enough calcium a little.

How do you judge the fluency of "Base Sentence"? (required)

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

Please input Flawless English sentence to be based on "Base Sentence". (required)

図3 タスクの例

Fig. 3 Screenshot of a task.

不適切な作業者を取り除くために作業者の資格審査を可能とした機能である。本実験では、日本の首都とアメリカの前大統領の名前を問う資格審査を行い、一定程度の知識を持つ作業者を選定した。

この結果、提案手法は29名から、既存手法2は33名から、それぞれ1000文の評価を取得した。この中で、明らかに不適切な英文を記入しているデータ*7を除き、提案手法(画像提示あり)は885文、既存手法2(画像提示なし)は973文の結果を得た。なお、既存手法2(画像提示なし)の71文は英文の取得に失敗したため、次章の分析で英文を用いるものに関しては、902文のみを用いる。

3.4 用例対訳候補の正確性判定

本節では、図1(4)で示した用例対訳候補の正確性判定について述べる。本稿では、本節の評価結果を各手法の評価で用いる。

本稿では、翻訳者3名に5段階で日英対の適合性評価*8を依頼した。評価依頼を行った日英対は下記のものである。

- 既存手法1(機械翻訳)で作成された英文と、もとの日本語文の対(100対)。
- 流暢性評価で3以下と判断された機械翻訳文に対して、提案手法(画像提示あり)で作成された英文と、もとの日本語文の対(重複を除いた444対)。
- 流暢性評価で3以下と判断された機械翻訳文に対して、既存手法2(画像提示なし)で作成された英文と、もとの日本語文の対(重複を除いた352対)。

*7 "Base Sentence", "This is OK", "This is good", "understandable", など

*8 文献[25]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 1 作成された翻訳文と英文の例

Table 1 Examples of translated text and created English texts.

用例作成者	機械翻訳			提案手法 (画像あり)			
	翻訳文	評価	正確性	英文	人数	評価	正確性
日本語用例							
こちらの影が見えますか?	Is this shadow seen?	3.33	×	Do you see this shadow?	3 人	4.67	○
				Do you see a shadow in the x-ray?	1 人	3.67	×
				Can you see this shadow?	1 人	5.00	○

- ・日本語用例は、医療従事者が患者に対してレントゲン写真を示しているイラストをもとに作成された。
- ・表中の評価は、翻訳者 3 名が行った日本語用例と翻訳文もしくは英文の適合性評価の平均である。
- ・表中の正確性は、適合性評価をもとに判定し、○が正確、×が不正確を示す。
- ・この翻訳文に対しては 8 名の労働者が評価した。そのうち、3 名が流暢性評価で 4 と評価した。表中の提案手法には流暢性評価で 3 以下をつけた労働者の結果のみを表示している。

表 2 既存手法 1 (機械翻訳) の翻訳精度

Table 2 Translation accuracy of the existing method 1 (machine translation).

	不正確	正確	合計
評価数	86	14	100

- ・翻訳者 3 名の 5 段階評価の平均が、4 より大きいものを正確、4 以下を不正確と判定している。

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。

4. 分析と考察

表 1 に実験で作成された翻訳文と英文の例を示す。なお、この文は、医療従事者が患者に対してレントゲン写真を示しているイラストをもとに作成された。表 1 は機械翻訳で不正確な日英対となっていたものが、提案手法によって正確な日英対となった例を示している。以降の各節で、作成された英文の正確性について詳しく議論を行う。また、最後にクラウドソーシングを使用することによるコストについて議論を行う。

4.1 各手法の正確性

本節では、各手法の正確性について考察する。

表 2 に既存手法 1 (機械翻訳) の翻訳精度を示す。表 2 より、既存手法で作成された英文は、8 割以上がそのままでは使用できないことが分かる。このため、用例対訳作成においては、機械翻訳をそのまま使用することは適切ではないことが分かる。

次に、表 3 に提案手法 (画像提示あり) と既存手法 2 (画像提示なし) の判定精度を示す。なお、労働者の判定は、労働者の流暢性評価の平均が 4 より大きい場合は正確、4 以下の場合には不正確と判定している。表 3 より、提案手法、既存手法 2 とともに、比較的正確に判定できていることが分かる。しかし、翻訳者が不正確と判定したものを、労働者が正確と判定した例が、提案手法では 7 文、既存手法 2 では 14 文存在していた。これは、労働者と翻訳者はそれぞれ流暢性評価と適合性評価を行っており、評価軸が異なっ

表 3 提案手法と従来手法 2 の判定精度

Table 3 Evaluated accuracy of the proposal method and the existing method 2.

労働者の判定	正確と判定		不正確と判定		正解率
	正確	不正確	正確	不正確	
翻訳者の判定					
提案手法 (画像あり)	10	7	4	79	89%
既存手法 2 (画像なし)	12	14	2	72	84%

- ・労働者の正解率は、労働者の判定と翻訳者の判定が同一となった割合を示す。
- ・合計は 100 文である。

いたことが原因であると考えられる。流暢性評価 (労働者) で正確、適合性評価 (翻訳者) で不正確と判定された文として、「It's a cold.」がある。この文は、原文の「風邪です」とはあまり適合していないが、英語の流暢性は高い。このため、労働者は提示された画像に適した正しい対訳を作成できなかったと考えられる。このため、画像の提示を行っている提案手法では、流暢性評価の他に、英文が画像に適しているかどうかを評価する必要があると考えられる。

4.2 画像の提示有無による正確性

本節では、画像の提示の有無による用例対訳の正確性について考察する。なお、本稿では、労働者の流暢性評価で不正確であると判定された機械翻訳文 (提案手法: 83 文、既存手法 2: 74 文) から作成された英文を、提案手法と既存手法 2 の比較で用いる。その際、流暢性評価で 3 以下と評価された、提案手法: 524 文、既存手法 2: 408 文をそれぞれ用いた。なお、これらの文数は重複を含んだ数である。

4.2.1 正確性判定基準

表 1 に示した労働者が作成した英文には、複数人が全く同じ文を記入する場合が存在しており、3.1 節で述べた事象が発生していることが分かる。複数人が同一の文を作成した場合、正しい文となっている可能性が高いと考えられる。このことから本稿では、英文を何人の労働者が作成したかを基準に正確性判定を行う。

4.2.2 提案手法と既存手法 2 の正確性

表 4 に各手法の正確割合を示す。なお、正確、不正確の

表 4 各手法の正確割合 (流暢性評価 3 以下)

Table 4 The rate of accuracy of each method (3 or less of the fluency evaluation method).

	評価利用基準		
	1 名	2 名以上	3 名以上
提案手法 (画像あり)	35.6% (83)	53.4% (58)	65.2% (23)
既存手法 2 (画像なし)	26.6% (74)	42.9% (35)	62.5% (16)

- ・評価利用基準の人数は、同じ文の記入を行った労働者の閾値である。
- ・表中の割合は、各手法で作成された英文の正確割合である。分析対象の英文数は、提案手法が 524 文、既存手法 2 が 408 文である。
- ・表中の括弧内は、使用している翻訳文の数である。分析対象の翻訳文数は、提案手法が 83 文、既存手法 2 が 74 文である。

表 5 各手法の正確割合 (全データ使用)

Table 5 The rate of accuracy of each method (All data is used).

	評価利用基準		
	1 名	2 名以上	3 名以上
提案手法 (画像あり)	36.9% (83)	50.0% (66)	69.7% (33)
既存手法 2 (画像なし)	27.6% (74)	37.3% (51)	52.2% (23)

- ・評価利用基準の人数は、同じ文の記入を行った労働者の閾値である。
- ・表中の割合は、各手法で作成された英文の正確割合である。分析対象の英文数は、提案手法が 885 文、既存手法 2 が 902 文である。
- ・表中の括弧内は、使用している翻訳文の数である。分析対象の翻訳文数は、提案手法が 83 文、既存手法 2 が 74 文である。

判定は、翻訳者による適合性評価を用いている。表 4 中の評価利用基準の人数は、同じ文の記入を行った労働者の閾値である。

表 4 より、画像の提示を行っている提案手法は、画像の提示を行っていない既存手法 2 より 9 ポイント (表 4 の評価利用基準が 1 名の差)、元の日本語用例と正しい対となる英語用例の作成ができていくことが分かる。また、評価利用基準を 2 名以上や 3 名以上と制限することで、正確な日英対が作られる割合が大きくなっていることが分かる。このことから、複数人が作成した英文は、正しい多言語対の作成に貢献していると考えられる。

4.2.3 提案手法の精度向上

4.2.2 項では、表 4 は 3 以下と評価した労働者が作成した英文のみを対象としていた。しかし、英文記入は流暢性評価で 4 以上と評価した労働者も行っている。このため、作成されたすべての英文 (提案手法: 885 文、既存手法 2: 902 文) をもとに再調査を行った。その結果を表 5 に示す。

表 5 より、提案手法は評価利用基準が 3 名以上のデータを用いた場合、流暢性評価で 3 以下のデータのみを利用した場合 (表 4) と比較して、4.5 ポイント正確性が向上したことが分かる。これは、流暢性評価で 4 以上をつけた労働者も、画像をもとにさらに正しい英文記入を行っていることが要因であると考えられる。この結果より、提案手法は流暢性評価の評価値を考慮に入れずに英文の選択を行う方が、より正しい日英対を選択できる可能性が考えられる。

しかし、既存手法 2 は流暢性評価で 3 以下のデータを利

表 6 翻訳文の適合性評価と日英対の正確性

Table 6 The adequacy evaluation of translated texts and the accuracy of Japanese-English pairs.

翻訳文の 適合性評価平均	提案手法の英文	
	正確率	対象文数
1.00	0.0%	8
1.33	6.1%	33
1.67	8.7%	23
2.00	5.7%	53
2.33	16.3%	43
2.67	30.0%	90
3.00	39.1%	87
3.33	42.7%	110
3.67	78.8%	52
4.00	56.0%	25

用した場合よりも、10.3 ポイント減少 (表 4 と表 5 の評価利用基準が 3 名以上の差) したことが分かる。これは、画像の閲覧を行っていないため、推測のみで英文作成を行った結果、元の日本語からみると不適切な英文が多く作成されたためであると考えられる。

これらのことから、画像の提示を行っていない既存手法 2 より画像の提示を行った提案手法の方が、適切な日英対を作成できていることが分かる。

4.3 機械翻訳の精度と日英対

本節では、機械翻訳の精度と、それをもとに作成される英文の関係について考察する。表 6 に機械翻訳で翻訳した文の適合性評価と、提案手法で作成された日英対の正確性の関係を示す。表 6 より、機械翻訳の適合性評価の平均が高い翻訳文の方が、適切な英文作成の手がかりとなっていることが分かる。

なお、適合性評価平均が 4.00 のときに正解率が下がっているのは、翻訳文が「Medicine for 3 days.」(原文は「3 日分のお薬です」) をもとに作成された英文が、適合性評価ですべて不正確であったことが要因となっていた。

4.4 クラウドソーシングのコスト

本節では、クラウドソーシングを用いることによるコストについて議論する。

翻訳者による翻訳業務を行っているエキサイト翻訳依頼プロ^{*9}の場合、医療用例の翻訳には 1 文字あたり 25 円が必要である^{*10}。今回利用した用例は平均 12.0 文字であった。このため、1 文の翻訳に約 300 円かかることが分かる。

提案手法では、1 文あたり 50 セント (=5 セント×10 人) で対訳の作成を行う。また、文献 [18] では、1 文あたり 50 セント (=5 セント×10 人) で多言語対の評価を行っている。正確性を高めるために、両手法とも倍の人数を使用し

^{*9} <https://orderpro.excite.co.jp/>

^{*10} 2012/12/12 現在。

た場合でも、約 166.5 円*¹¹で翻訳が可能となる。

さらに、翻訳者による翻訳は対訳が 1 文のみ生成されるが、本手法では複数の対訳が生成される。多言語用例対訳は多様性を持つことで自由度を高めた多言語変換を行うことが可能である [26]。これらのことから、安価で多様性を持つ多言語対の作成が可能な本手法は有用であると考えられる。

5. おわりに

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手法の提案とその実験を行った。本手法では、画像を単言語話者に提示することで、正確な多言語対の作成を目指した。

本稿の貢献は、画像を用いることで、単言語話者のみで機械翻訳よりも高精度の多言語用例対訳候補の作成が可能であることを示した点である。

今後は、他のデータセットを用いて同様の結果が得られるか追加実験を行う。その際、画像と英文の適切性について確認を行う。

謝辞 本研究の一部は、科研費基盤研究 (B)(22300044) の助成を受けたものである。

参考文献

- [1] 法務省：平成 23 年末現在における外国人登録者数について (確定値)，法務省 (オンライン)，入手先 (<http://www.moj.go.jp/nyuukokukanri/kouhou/nyuukokukanri04-00021.html>) (参照 2012-12-12)。
- [2] 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 2012-12-12)。
- [3] 法務省：平成 23 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/nyuukokukanri04-00017.html>) (参照 2012-12-12)。
- [4] 総務省：多文化共生の推進に関する研究会報告書，総務省 (オンライン)，入手先 (http://www.soumu.go.jp/kokusai/pdf/sonota_b5.pdf) (参照 2012-12-12)。
- [5] 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).
- [6] 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).
- [7] Kim, K. J. and Bonk, C. J.: Cross-Cultural Comparisons of Online Collaboration, *Journal of Computer Mediated Communication*, Vol. 8, No. 1 (2002).
- [8] 高嶋愛里：在日外国人支援活動：京都における「医療通訳システムモデル事業」，国際保健支援会 2 (2005)。
- [9] 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)。
- [10] 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)。
- [11] 福島 拓，吉野 孝，重野亜久里：用例対訳を用いた多言語問診票作成システムの開発と評価，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. 2011-GN-78, No. 14, pp. 1-7 (2011)。
- [12] 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)。
- [13] 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 (2012)。
- [14] Howe, J.: Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business, *Crown Business* (2008)。
- [15] Doan, A., Ramakrishnan, R. and Halevy, A. Y.: Crowdsourcing systems on the World-Wide Web, Vol. 54, No. 4, pp. 86-96 (2011)。
- [16] 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)。
- [17] 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)。
- [18] 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 (2012)。
- [19] 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)。
- [20] 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)。
- [21] 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)。
- [22] 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)。
- [23] Hu, C., Bederson, B. B. and Resnik, P.: Translation by Iterative Collaboration between Monolingual Users, *HCOMP '10 Proceedings of the ACM SIGKDD Workshop on Human Computation*, pp. 54-55 (2010)。
- [24] Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration, *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp. 96-100 (2006)。
- [25] Walker, K., Bamba, M., Miller, D., Ma, X., Cieri, C. and Doddington, G.: Multiple-Translation Arabic (MTA) Part 1, *Linguistic Data Consortium, Philadelphia* (2003)。
- [26] 福島 拓，吉野 孝：正確な多言語間対話支援を目的とした応答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)。

*¹¹ 2012/12/12 現在。1 ドル = 83.25 円で計算。