

推薦論文

穴あき用例と単言語話者作成の正確な用例とを活用した 多言語用例対訳作成手法

福島 拓^{1,a)} 吉野 孝^{2,b)}

受付日 2018年4月16日, 採録日 2018年10月2日

概要: 正確な多言語間コミュニケーション支援が求められる分野では用例対訳が多く用いられている。用例対訳は用例を正確に翻訳した多言語コーパスであり、正確な多言語変換が可能である。用例対訳の作成は人手で行われているため、十分な数の用例対訳を確保することが困難であった。そこで本論文では、用例の一部を穴あきにして入れ替え可能とした穴あき用例を活用し、単言語話者のみで様々な種類の用例対訳を作成する手法を提案し、手法の評価を行った。本手法では、単言語話者がそれぞれの言語で作成した正確な用例と、対訳になりうる穴あき用例の対とを用いることで、正確な用例対訳の作成を目指している。本論文の貢献は以下である。(1) 穴あき用例を活用した用例対訳作成手法を提案し、実現した。(2) 穴あき文の対と穴埋め単語の対の正確性担保を行ったうえで提案手法を適用することで、高い精度で用例対訳作成が可能であることを示した。

キーワード: 穴あき用例, 用例対訳, 多言語, コミュニケーション支援

Multilingual Parallel Texts Creation Method Using Accurate Texts Created by Monolingual Users and Perforated Texts

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

Received: April 16, 2018, Accepted: October 2, 2018

Abstract: Parallel text is often used in fields where accurate multilingual communication support is required. It is a multilingual corpus that can perform accurate multilingual conversion. The volume of parallel texts available has been inadequate because they were all created manually. In this paper, we propose and evaluate a multilingual parallel texts creation method using a perforated text whereby words in the texts can be interchanged. This method uses accurate texts created in each language by monolingual users and multilingual pairs of perforated texts. The major contributions of this study are as followings. (1) We proposed and developed a multilingual parallel texts creation method using perforated texts. (2) The proposed method can create accurate parallel texts by using perforated texts and filling words being guaranteed accuracy.

Keywords: perforated text, parallel text, multilingual, communication support

1. はじめに

近年の世界的なグローバル化により、日本国内において

も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、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 [1], [2], [3], 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 大阪工業大学情報科学部
Faculty of Information Science and Technology, Osaka Institute of Technology, Hirakata, Osaka 573-0196, 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) taku.fukushima@oit.ac.jp

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

本論文の内容は 2017 年 6 月のマルチメディア, 分散, 協調とモバイル (DICOMO2017) シンポジウムで報告され, グループウェアとネットワークサービス研究会主査により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である。

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、多言語間対話においては高い正確性が求められる。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 [4], [5]。

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 [6], [7], [8], [9]。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多言語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、正確な多言語間対話が可能となる。なお、本論文では、単言語の文を「用例」、多言語変換可能な用例の組を「用例対訳」とする。

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 の開発を行っている [10]。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。また、日常会話を対象とした多言語用例対訳共有システムも開発されている [11]。これらのシステムでは多言語話者が翻訳を担っているが、多言語話者の人数は相対的に少なく、大きな負担がかかっている。

用例対訳作成の効率化を目的として、用例の一部を穴あきにして単語を入れ替え可能とした「穴あき用例」が存在している [12]。穴あき用例は用例対訳全体で必要となる翻訳数を減少可能であるため、穴あき用例を活用した用例対訳作成の取り組みが行われているが [13], [14]、正確な用例対訳の作成手法は確立されていない。

そこで本論文では、新たな穴あき用例を活用した用例対訳作成手法を提案する。本手法では、単言語話者がそれぞれの言語で作成した正確な用例と、対訳になりうる穴あき用例の対の両者を活用することで、正確な用例対訳の作成を目指す。このことで、翻訳コストを減らしたうえで多様な用例対訳の作成の実現を可能とする。

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 [15] や多言語対面環境の討論支援 [16] 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 [17]。ニューラル機械翻訳の登場により翻訳精度の

急激な向上 [18], [19] も見られるが、文の一部が翻訳されない「訳抜け」の問題が新たに生じている [18]。また、従来型の翻訳システムと比べて翻訳誤りの原因の追及が難しい点 [19] や、翻訳結果の流暢性が高いため、翻訳誤りに気付くにくいという問題 [18] が指摘されている。また、機械翻訳は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 [6] や、ケータイ多言語対話システム [7] 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 [8], [9]。

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 [10]。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う、(iv) システムが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 [10]。今後、医療分野で必要な用例対訳を網羅した場合、現在の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かかえている。

2.2 穴あき用例

効率的な用例対訳作成の一方法として穴あき用例が存在している。穴あき用例は、用例の一部の単語を穴あきにして入れ替え可能とした用例である [12]。穴あき用例の穴あき部分の単語を入れ替えることで、具体的な内容の伝達が可能である。また、穴あき用例は翻訳する文の数が少なくなるため、翻訳コストが低いというメリットが存在している。ただし、穴あき用例は言語間で 1 対 1 に対応しておらず、全自動での用例対訳の作成は困難である [12]。以降、用例の一部を穴あきにしたものを「穴あき文」、穴あき部分に入れる単語を「穴埋め単語」とする。

文献 [13] では、穴あき用例のリンク構造を活用した用例対訳作成を行っている。また、穴あき文と穴埋め単語の間の適合可能性を示す指標や、穴あき文どうしの包含関係を用いることで、用例作成者の負担軽減が行われている。実験より、穴あき文と穴埋め単語の組合せを入れ替えることで、多くの新たな用例作成が可能であることが示されている。

文献 [14] では、穴あき用例と単語辞書、Web 検索機を用いて、人手による翻訳作業なしに用例対訳の作成を行っている。これは、単言語の用例作成にとどまっていた文

献 [13] を多言語化する取り組みである。文献 [14] では、穴あき文と穴埋め単語の対訳をそれぞれ作成したうえで、両者を結合して用例対訳を作成するアプローチをとっており、適切な用例対訳が一定程度、生成可能であることが示された。しかし、単数形や複数形の適切な判断が困難であるなど、不適切な対訳が生成される場合も一定数存在していた。また、翻訳可能な文が少ないという欠点も存在していた。

本論文では、文献 [14] の手法で問題となっていた自動的な用例の生成を行わず、正確に作成された用例どうしを言語間リンクでつなぐことで、正確な用例対訳の生成を目指す。

3. 穴あき用例を活用した用例対訳作成

本章では、提案する穴あき用例を用いた単言語話者による多言語用例対訳作成手法について述べる。本手法では、文献 [12] の穴あき用例の作成手法、および、文献 [13] の穴あき用例を用いた用例作成手法を活用する。以降、3.1 節で穴あき用例の作成について、3.2 節で穴あき用例を活用した用例作成について、それぞれ述べた後、3.3 節で提案手法について述べる。

3.1 穴あき用例の作成

穴あき用例の利用について検討した文献 [12] では、以下の手順で穴あき用例の作成を行っている。

- (1) 形態素解析器を用いて、同一言語の用例群を形態素に分割する。その際、句読点や“?”“!”は除去する。
- (2) (1) で分割した形態素のうち、1つのみが異なる用例の対を抽出する。
- (3) (2) で抽出した、1形態素のみ異なる用例群のうち、異なる単語を「穴埋め単語」、それ以外の部分を「穴あき文」として保存する。

なお、本論文では穴あき文の穴あき部分を「[[N]]」のように表記する。使用するアルファベットは品詞を表しており、名詞は“N”を用いて表現する。例として、「頭が痛いです」の「頭」(名詞)が穴埋め単語の場合、穴あき文を「[[N]]が痛いです」と表記する。

3.2 用例の作成

穴あき用例を活用した用例作成手法 [13] では、以下の手順で用例作成が行われている。

- (1) 3.1 節で述べた穴あき用例作成方法 [12] を用いて、穴あき文、穴埋め単語を動的に作成する。
- (2) 任意の穴あき文に属する穴埋め単語群を抽出する。
- (3) (2) の穴埋め単語が属する穴あき文群を抽出する。
- (4) (3) の穴あき文に属する穴埋め単語群を抽出する。このうち、(2) の穴埋め単語群に含まれないものを、(2) の穴あき文に対応する穴埋め単語候補とする。

- (5) (2) の穴あき文に、(4) の穴埋め単語候補を埋めた文をシステム利用者に提示する。利用者は文が一般的に使用されるかどうかを判定し、使用されると判定されたものは新たな用例となる。

また、文献 [13] では、(5) において、(2) の穴あき文から(4) の穴埋め単語候補までのリンクの経路数を提示している。このことで、穴あき文に含まれる可能性が高い穴埋め単語候補を提示可能となり、適切な穴埋め単語候補の選択を補助している。また、穴あき文 A で使用される穴埋め単語群は、穴あき文 B で使用される穴埋め単語群につねに含まれる、などの包含関係を追加可能としている。このことで、穴あき文 A に穴埋め単語が追加された場合、穴あき文 B にも自動的に穴埋め単語が追加される。このような形で、文献 [13] では作業者の作業量の軽減が図られている。

3.3 用例対訳作成手法

本節では、本論文で提案する、穴あき用例を活用した用例対訳作成手法について述べる。

文献 [14] では正確な穴あき文の対と穴埋め単語の対を作成し、それらを組み合わせて対訳となる用例を生成していた。しかし、単数形と複数形の違いなどにより、不正確な用例が生成される場合が存在していた。本手法ではまず、正しい用例の作成を各言語において人力で行う。次に、穴あき文の対、穴埋め単語の対を用いて同じ意味の用例間にリンクを作成することにより、不正確な用例を作成しない手法とした。これらが本手法と文献 [14] との相違点である。

本手法の流れを以下に、図 1 に適用例をそれぞれ示す。

- (1) 3.1 節で述べた手法を用いて、穴あき文と穴埋め単語を各言語でそれぞれ動的に作成する (図 1(1))。
- (2) 3.2 節で述べた手法を用いて、新規の用例を各言語話者がそれぞれ作成する。これは、図 1(2) の※1および※2にあたる。作成される用例はその言語の話者によって作成されるため、文献 [14] の手法とは異なり、正確な用例が作成される。この時点では、言語間にリンクが存在していないため、用例対訳とはなっていない。
- (3) 穴あき文、穴埋め単語の対訳候補*1をそれぞれ動的に取得する (図 1(3))。本手順では、すでに言語間リンクでつながれている用例から作成された、同じ品詞の穴あき文どうし、穴埋め単語どうしをそれぞれ対訳候補とする。図 1 において、言語間リンクは黒色の実線で、対訳候補は緑色の破線でそれぞれ示されている。
- (4) 任意の穴あき用例から作成された別言語の 2 つの用例が、(3) で作成された穴あき文の対訳候補、かつ、穴埋め単語の対訳候補からなっているとき、両用例は対訳であるとする (図 1(4))。図 1 の場合、(2) で作成された※1と※2の各用例は、(3) の穴埋め単語の対

*1 作成された対は同じ意味であると担保されていないため、本論文では対訳候補としている。

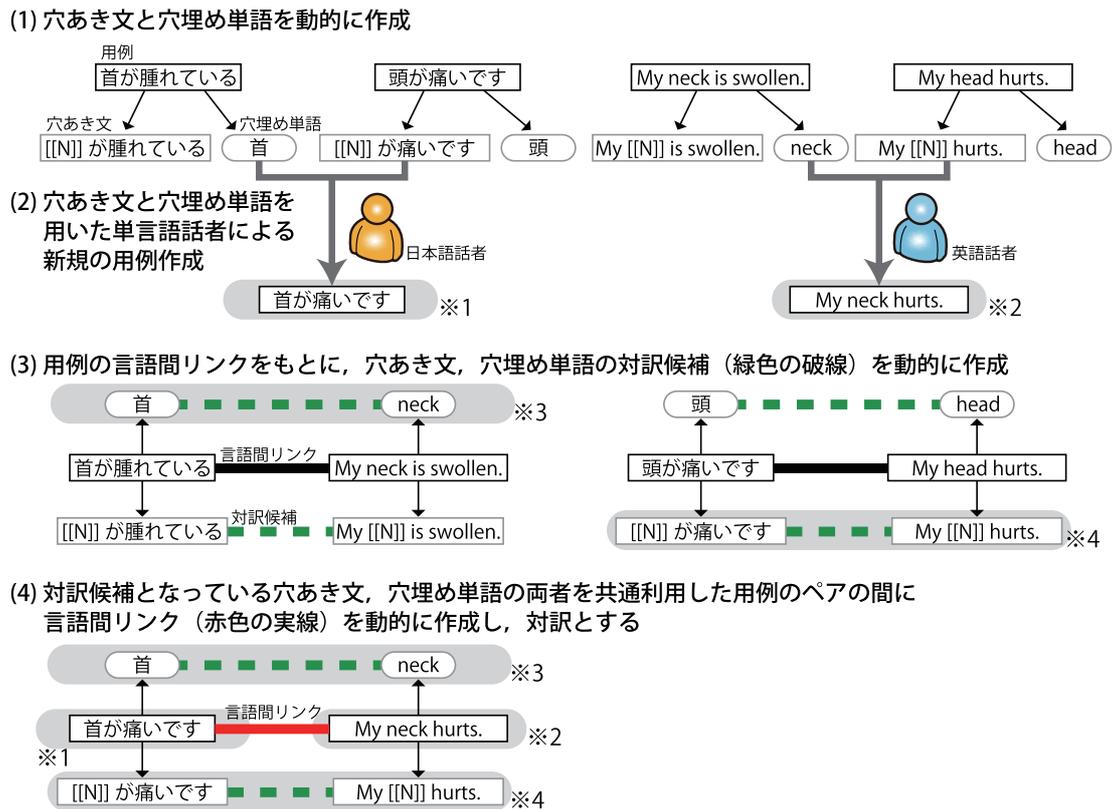


図 1 穴あき用例を用いた用例対訳作成手法

Fig. 1 Parallel texts creation method using perforated texts.

訳候補（※3）と穴あき文の対訳候補（※4）からなっている。この場合、両用例は同じ意味であるとして、言語間リンクを新規に作成し（図1中の赤色の実線）、両者を用例対訳とする。なお、図1では(2)で作成された用例どうしに言語間リンクを作成しているが、既存の用例（図1(1)の用例）どうしや、既存の用例と(2)で作成された用例の間で前述の状態となった場合も、言語間リンクを作成可能とする。

4. 実験

本章では、提案手法を用いた用例対訳作成実験について述べる。本実験の目的は、複数言語の穴あき用例を用いて適切な用例対訳が作成可能であることを検証することである。

表1に本実験で使用した穴あき用例について示す。本実験では、用例対訳コーパスとして、多言語用例対訳共有システム TackPad [10] を用いた。本コーパスは、医療分野で使用される用例で構成されている。また、用例対訳作成対象の言語は日本語と英語とした。

まず、日本語は ChaSen [21]、英語は TreeTagger [22] の形態素解析器を用いて穴あき用例を作成した。これらは、3.3節の手順(1)にあたる。この結果、日本語の穴あき文は195文、穴埋め単語は301単語、英語の穴あき文は110文、穴埋め単語は177単語となった。なお、本実験では対象の品詞を名詞のみとした。日本語用例と英語用例の間に

表 1 実験で使用した穴あき用例の詳細

Table 1 Details of perforated texts.

言語	日本語	英語
全体の用例数	6,011	2,534
形態素解析器	ChaSen	TreeTagger
穴あき文数（名詞）	195	110
穴埋め単語数（名詞）	301	177
穴あき用例の作成元の用例数（#1）	707	364
#1の間に存在する言語間リンク数	214	
新規作成された用例数	692	188

・単位は文および単語である。

・穴あき文数および穴埋め単語数は穴あき部分が名詞であるもののユニーク数である。

存在していた既存の言語間リンク数は214リンクであった。なお、同じ用例対訳コーパスを実験で使用した文献[14]において誤りが見つかった既存の言語間リンクは修正を行っている*2。

次に、言語別に穴あき用例を用いた用例作成を行った。これは、3.3節の手順(2)にあたり、文献[13]のシステムを利用した。日本語の用例作成は日本語を母語とする著者の1人が行った。英語の用例は日本語を母語とする英語教育従事者に作成を依頼した。用例作成の結果、日本語は

*2 文献[14]では、「足がかゆいです - My calves are itchy.」「手首に力が入りません - My ankles feel weak.」など、不正確な用例対訳が発見されている。

表 2 実験結果 (言語間リンク数)

Table 2 Results of the experiment (number of links between languages).

	提案手法			提案手法+穴埋め単語抽出手法 [14]		
	正確	不正確	計	正確	不正確	計
日本語, 英語ともに既存	12	20	32	11	0	11
日本語が既存, 英語が新規作成	38	5	43	38	0	38
日本語が新規作成, 英語が既存	22	12	34	12	1	13
日本語, 英語ともに新規作成	46	7	53	42	0	42
計	118	44	162	103	1	104

- ・新規作成は, 3.3 節の手順 (2) で作成された用例である.
- ・不正確は, 作成された言語間リンク数のうち不正確と判定されたリンク数である.

表 3 作成された用例対訳の例

Table 3 Examples of created parallel text.

	作成された用例対訳		穴埋め単語		穴あき文	
	日本語	英語	日本語	英語	日本語	英語
1	くるぶしが腫れています	My ankle is swollen.	くるぶし	ankle	[[N]] が腫れています	My [[N]] is swollen.
2	肘が腫れています	My elbows are swollen.	肘	elbows	[[N]] が腫れています	My [[N]] are swollen.
3	手が動きません	I am unable to move my hands.	手	hands	[[N]] が動きません	I am unable to move my [[N]].
4	膝が動きません	I can't move my knees.	膝	knees	[[N]] が動きません	I can't move my [[N]].

692 文, 英語は 188 文, それぞれ新しい用例が作成された. なお, 各作成者は他方の言語で作成された用例を閲覧できない状態でそれぞれ用例を作成している. これは, 本手法を実環境に適用する際は, 単言語での作業を想定しているためである.

その後, 3.3 節の手順 (3), 手順 (4) により, 新たな用例対訳の作成を行った. 本実験では, 新規作成された用例 (対訳なし) と, 既存の用例 (対訳ありと対訳なしが混在している) の両者を用いて, 新たな用例対訳の作成を行った. また, 本手法と文献 [14] の穴埋め単語抽出手法を併用した場合の調査も行った. 文献 [14] の穴埋め単語抽出手法を併用した場合, 手順 (3) において抽出された穴埋め単語の対訳候補のうち, 多言語辞書に登録されている対のみを使用するアルゴリズムとなる. なお, 多言語単語辞書は「デ辞蔵^{*3}」の EDICT 和英辞典を用いた.

最後に, 作成された用例対訳の正確性評価を著者の 1 人が行った. ここでは, 穴埋め単語の対の正確性, 穴あき文の対の正確性, 多義語となる穴埋め単語をあらかじめ確認したうえで正確性評価を行っており, 機械的に判断できる作業であった.

なお, 本実験で使用した日本語および英語の用例作成者は 1 人ずつであった. このことにより, 正確ではないが使用される可能性のある用例^{*4}や, 使用頻度の低い用例において, 判断の揺れが発生している可能性がある. しかし, このような用例は限られているため, 実験結果に与える影響は小さいものであると考えられる.

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

^{*4} 歯茎の腫れの意味で「歯が腫れています」など [13].

5. 実験結果と考察

5.1 提案手法の効果

実験結果を表 2 に示す. 表 2 の「提案手法」の列より, 一部に不正確な用例対訳が作成されているものの, 既存の用例間や新規作成された用例間など, すべてのパターンで正確な用例対訳が作成されていることが分かる. また, 新規に正確な言語間リンクを 118 リンク作成できたことが分かる. 使用したコーパスには 214 リンクの言語間リンクが存在していた (表 1 の「#1 の間に存在する言語間リンク数」). このことから, 正確な言語間リンク数を約 1.55 倍 ($= (214 + 118)/214$) にできたことが分かる.

なお, 本手法で作成された不正確な用例対訳の多くは, 文献 [14] の穴埋め単語抽出手法を活用することで動的に発見可能であった (表 2 の「提案手法+穴埋め単語抽出手法 [14]」). ただし, 文献 [14] の手法では一部の正確な用例対訳も除去されたため, 正確な言語間リンク数は 103 リンク, 不正確な言語間リンク数は 1 リンクとなり, 約 99% の精度となった. また, 正確な言語間リンク数は約 1.48 倍 ($= (214 + 103)/214$) となった.

作成された用例対訳の例を表 3 に示す. 表 3-1 および表 3-2 は, 英語の穴埋め単語が単数形であるか複数形であるかの違いが存在している. 従来手法である文献 [14] では, 三人称単数現在を示す “s” が動詞についている穴あき文にもかかわらず, 複数形の名詞が穴埋め単語として入れられた用例が作成されるなど, 不正確となっている例が存在していた. それに対して本手法では, 表 3-1 および表 3-2 に示すとおり, 適切に用例対訳を作成できていることが分

表 5 穴埋め単語の対訳候補が誤りであった例

Table 5 Examples of improper translation candidate of filling words.

穴埋め単語の対	作成元の用例対訳	修正後の穴埋め単語の対
指 - toes	足の指が痛いです - I hurt my toes.	足の指 - toes
指先 - fingers	指先の色が悪いです - The tips of my fingers are discolored.	(私の) 指先 - tips of my fingers
爪 - toenails	足の爪に異常があります - My toenails are abnormal.	足の爪 - toenails

表 4 不正確と判定された用例対訳の分類結果

Table 4 Classification of inaccurate parallel texts.

	理由	件数
類型 1	穴埋め単語の対訳候補が誤り	39
類型 2	穴あき文の対訳候補が誤り	1
類型 3	穴埋め単語と穴あき文の両者の対訳候補が誤り	4

かる。このような例は、男性名詞や女性名詞が存在するラテン系の言語などでも起きうるが、この場合でも本手法では適切な用例対訳が作成できると考えられる。

表 3-3 および表 3-4 は、日本語の穴あき文は共通であるが、英語の穴あき文の表現が異なっている例である。既存の用例対訳には「手が動きません - I can't move my hands.」「膝が動きません - I am unable to move my knees.」が存在していたが、本手法の適用により、別表現の英語が用例対訳として利用可能となったことが分かる。用例対訳はあらかじめ正確性を担保したコーパスであり、完全一致した文のみ翻訳が可能であるため、同じ意味でも様々な表現の翻訳が可能である必要がある。本手法によって異なる表現の用例間にリンクが作成されたことから、本手法の適用により、より良いコーパス作成が可能になると考えられる。

また、表 2 に示すとおり、日本語と英語がともに既存の用例であった場合でも言語間リンクが作成され、12 件 (文献 [14] の穴埋め単語抽出手法を併用した場合は 11 件) の正しい用例対訳が作成されていた。既存の用例どうしをつないだ例として「手がしびれています - My hands are numb.」がある。用例「My hands are numb.」は「手がしびれます」と用例対訳となっていたが、本手法を適用することで、英語の対訳が存在していなかった用例「手がしびれています」とも用例対訳となった。このように、本手法を適用することで不足している言語間リンクを作成することが可能となり、翻訳可能な文数を増加可能であることが分かる。

5.2 不正確と判定された用例対訳

本節では、提案手法で作成された用例対訳のうち、不正確と判定されたものについて考察する。本実験において不正確と判定された用例対訳は、3.3 節の手順 (3) で作成された穴埋め単語や穴あき文の対訳候補の誤りが原因であった。また、不正確と判定された用例対訳は 3 類型に分類可能であった。分類結果とその件数を表 4 に示す。なお、前

節で述べた、文献 [14] の穴埋め単語抽出手法を適用した場合、類型 1 と類型 3 の用例対訳についてはすべて不正確であると正しく判定された。

なお、表 2 より、既存の英語の用例に対して日本語の対訳が作成された場合、不正確な用例対訳となりやすい傾向にあることが分かる。これは、日本語の用例よりも英語の用例の方が既存の対訳が多く存在していた^{*5}が、さらに英語の用例に対しても新たな対訳作成を試みた結果、不正確な言語間リンクが作成されたためであると考えられる。

(1) 穴埋め単語の対訳候補が誤り

本類型は、3.3 節の手順 (3) で作成された穴埋め単語の対訳候補が誤っていたため、作成された用例対訳が不正確となった例である。本類型は 44 件の誤りのうち、最も多い 39 件の誤りが存在していた。

本類型で使用された穴埋め単語の対は、表 5 の 3 種類であった。これらは、表 5 の「修正後の穴埋め単語の対」の列に示したように、穴埋め単語を複合語へ対応させることで、正確な用例対訳の作成が可能になると考えられる。

(2) 穴あき文の対訳候補が誤り

本類型は、3.3 節の手順 (3) で作成された穴あき文の対訳候補が誤っていたため、作成された用例対訳が不正確となった例である。本類型は 1 件の誤りが存在していた。

本類型で使用された穴あき文の対は、「[[N]] を打撲しました - I bruised my lower [[N]].」であり、「前腕を打撲しました - I bruised my lower arm.」の用例対訳から作成されていた。このため、前項と同様に、「[[N]] を打撲しました - I bruised my [[N]].」のように複合語へ対応することにより、正確な用例対訳作成が可能になると考えられる。

なお、本類型の用例対訳は、文献 [14] の穴埋め単語抽出手法を用いても検出することができなかったが、文献 [14] の穴あき文適切性判定手法を活用することで発見可能であると考えられる。ただし、穴あき文適切性判定手法では、作成された用例対訳候補の適切性の順位付けを行う手法であるため、動的な判定を行う場合は、1 位の用例対訳候補のみを正確と判定することとなる。このため、5.1 節で述べた表現の異なる言語間リンクも不正確と判定される可能性が高い。このことから、本論文で述べた提案手法および文献 [14] の穴埋め単語抽出手法を用いて用例対訳の自動作

*5 表 1 の用例数および言語間リンク数より、日本語は 3 割程度の用例に対してのみ対訳が存在していたことに対し、英語は半数以上の用例に対して対訳が存在していたことが分かる。

成を行っただけで、文献 [14] の穴あき文適切性判定手法で検出された、2 位以下の用例対訳候補のみ人手で正確性を判定する必要があると考えられる。

(3) 穴埋め単語と穴あき文の両者の対訳候補が誤り

本類型は、3.3 節の手順 (3) で作成された穴埋め単語、および、穴あき文の対訳候補が誤っていたため、作成された用例対訳が不正確となった例である。本類型は 4 件の誤りが存在していた。

本類型の用例対訳の例として「胸が痛いです - I have a congestion」がある。これは、穴あき文の対訳候補「[[N]]が痛いです - I have a [[N]]」、穴埋め単語の対訳候補「胸 - congestion」から作成された。前述の穴あき文は「頭が痛いです - I have a headache」から、穴埋め単語は「胸が詰まった感じがする - chest congestion」からそれぞれ作成されていた。このように、穴あき文と穴埋め単語ともに誤った対訳候補が作成された場合、不適切な対訳が作成されることが分かる。

本類型の用例対訳は、元となる用例どうしの意味が厳密には異なっていた。このような用例対訳を本手法で利用することは困難であるため、穴あき文の対や穴埋め単語の対の正確性判定を行い、このような用例対訳から作成された穴あき文や穴埋め単語を排除する必要があると考えられる。

5.3 従来の用例対訳作成手法との比較と今後の展開

本手法は、単言語話者のみで作成された各言語の用例を動的に組み合わせて用例対訳を作成している。従来行われている人手による用例対訳作成と比較すると、単言語話者の作業量はほぼ同程度であるものの、多言語話者が行っていた翻訳作業が不要となる。このため、提案手法により、人的リソースが少ない多言語話者の時間的・金銭的なコストが大幅に減少可能になると考えられる。また、5.1 節で述べたように、提案手法と文献 [14] の穴埋め単語抽出手法とを併用することで、約 99% の高い翻訳精度となることを本論文では示した。人手による作業はスペルミスや文法ミスなどが生じる可能性があるため、本論文で提案する自動作成手法は人手と同程度かそれ以上の翻訳精度であると考えられる。

また、従来の人手による用例対訳作成においては、翻訳誤りを検出するために、複数の多言語話者によるすべての用例対訳に対する評価が必要となっていた。提案手法においても用例対訳の評価は必要となるが、人的リソースが比較的多い単言語話者による用例の評価が中心となる。多言語話者は文献 [14] の穴あき文適切性判定手法で検出された用例対訳のみ評価を行うため、評価対象の用例対訳数が少なくなると考えられる。

これらの結果から、今後は各言語で穴あき用例を活用した用例の作成を行っただけで、最低限の翻訳を多言語話者に依頼することを検討する。このことで、より効率的に正

確な用例対訳作成が可能になると考えられる。

また、穴あき用例は、1 つの穴あき文と、穴あき文に対応する穴埋め単語群の 2 種類を翻訳することで、多くの用例対訳を容易に生成可能であるという利点がある。本論文の実験では、人手による翻訳も比較的容易な短文の用例を用いていた。これを、翻訳誤りが生じる可能性が高くなる長文の穴あき文にも適用することで、翻訳の人的コストを下げたうえで、正確性を担保した多様な用例対訳を作成可能になると考えられる。

6. まとめ

本論文では、穴あき用例を活用した用例対訳作成手法について述べた。本手法では、穴あき用例を活用して別々に作成された各言語の正確な用例と、用例対訳コーパスの言語間リンクを用いることで新たな用例対訳の作成を行った。

本論文の貢献は、以下の 2 点である。

- (1) 穴あき用例を活用した用例対訳作成手法を提案し、実現した。
- (2) 穴あき文の対と穴埋め単語の対の正確性担保を行っただけで提案手法を適用することで、高い精度で用例対訳作成が可能であることを示した。

今後は、5.3 節で述べた内容および複合語への対応を検討し、実装を行うことで、多様な用例対訳作成の実現を目指す。その際に、翻訳者の負担がどの程度減少したのかを、定量評価および聞き取り調査から明らかにする。また、本論文では医療分野への適用について述べたが、用例対訳の概念は特定の分野に特化していないため、他の分野においても本手法は適用可能である。このため、今後、正確性が求められる他分野に対しての適用を目指す。

謝辞 本研究の一部は JSPS 科研費 JP26730105, JP18K18096 による。

参考文献

- [1] 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).
- [2] 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).
- [3] Kim, K.J. and Bonk, C.J.: Cross-Cultural Comparisons of Online Collaboration, *Journal of Computer Mediated Communication*, Vol.8, No.1 (2002).
- [4] 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, Vol.2 (2005).
- [5] 伊藤美保, 飯田奈美子, 南谷かおり, 中村安秀: 外国人医療における医療通訳者の現状と課題—医療通訳者に対する質問紙調査より, 国際保健医療, Vol.27, No.4, pp.387-394 (2012).
- [6] 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).

- [7] 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).
- [8] 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).
- [9] 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研究報告, Vol.AI2010-47, pp.19-24 (2011).
- [10] 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.23-33 (2012).
- [11] Bond, F., Nichols, E., Appling, D.S. and Paul, M.: Improving Statistical Machine Translation by Paraphrasing the Training Data, *Proc. IWSLT 2008*, pp.150-157 (2008).
- [12] 福島 拓, 吉野 孝: 多言語用例対訳共有システムにおける穴あき用例の利用可能性, 電子情報通信学会技術研究報告, 人工知能と知識処理, Vol.114, No.461, pp.23-28 (2015).
- [13] 福島 拓, 吉野 孝: 穴あき用例を用いた多言語用例対訳作成手法の効果, 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.2017-GN-100, No.48, pp.1-6 (2017).
- [14] 福島 拓, 吉野 孝: 穴あき用例を活用した対訳作成手法の比較, 電子情報通信学会技術研究報告, 人工知能と知識処理, Vol.116, No.474, pp.7-12 (2017).
- [15] Matsuda, M. and Kitamura, Y.: Development of Machine Translation System for Japanese Children, *Proc. 2009 ACM International Workshop on Intercultural Collaboration (IWIC '09)*, pp.269-271 (2009).
- [16] 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).
- [17] 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).
- [18] 中澤敏明: 機械翻訳の新しいパラダイム: ニューラル機械翻訳の原理, 情報管理, Vol.60, No.5, pp.299-306 (2017).
- [19] 鶴岡慶雅: ニューラル機械翻訳の衝撃, 情報処理, Vol.58, No.2, pp.96-97 (2017).
- [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] 北内 啓, 宇津呂武仁, 松本裕治: 誤り駆動型の素性選択による日本語形態素解析の確率モデル学習, 情報処理学会論文誌, Vol.40, No.5, pp.2325-2337 (1999).
- [22] Schmid, H.: Probabilistic part-of-speech tagging using decision trees, *Proc. International Conference on New Methods in Language Processing*, pp.44-49 (1994).



福島 拓 (正会員)

1986年生。2008年和歌山大学システム工学部中退。2013年同大学大学院システム工学研究科博士後期課程修了。博士(工学)。現在、大阪工業大学情報科学部特任講師。CSCWの研究に従事。



吉野 孝 (正会員)

1969年生。1992年鹿児島大学工学部卒業。1994年同大学大学院工学研究科修士課程修了。博士(情報科学)。現在、和歌山大学システム工学部教授。CSCW, HCIの研究に従事。

推薦文

DICOMO2017の発表論文の中で特に評価が高かったため(問題設定を医療コミュニケーションとしており、想定シーンの制約を設けることによって、実用性・有用性の高いシステムを実現している)。

(グループウェアとネットワークサービス研究会主査
齊藤典明)