

多言語用例対訳共有システム TackPad の用例評価手法の評価

Evaluation of Parallel-text Evaluation Methods
of Multilingual Parallel-text Sharing System TackPad福島 拓[†]
Taku Fukushima吉野 孝[†]
Takashi Yoshino重野 亜久里[‡]
Aguri Shigeno

1. はじめに

現在、在日外国人や訪日外国人数は増加傾向にある [1]。しかし、在日外国人や訪日外国人の中には、日本語を理解できない人が多数存在している [2]。一般に多言語を十分に習得することは難しく、母語以外の言語によるコミュニケーションは困難である [3]。このため、日本語を理解できない外国人と日本人とのコミュニケーションは十分に行うことができない。

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。このような言葉の正確性を求められる場では、正確な用例対訳を用いた支援が行われている [4]。用例対訳とは用例を多言語に翻訳した多言語コーパスのことを指す。

我々は Web 上での多言語用例対訳の収集、共有、提供を目的とする多言語用例対訳共有システム TackPad (タックパッド) の開発を行い、評価機能によって用例対訳の正確性の確保を目指している [5]。

本稿では、3 つの評価手法の評価を行い、用例対訳の評価に適した評価手法について検証する。

2. システム設計

本章では多言語用例対訳共有システム TackPad の設計について述べる。

2.1 システム概要

本システムは、多言語用例対訳を収集するため画面インタフェースを多言語としている。本システムの画面例を図 1 に示す。収集言語は、日本語、英語、中国語、韓国・朝鮮語、ポルトガル語、スペイン語、ベトナム語、タイ語、インドネシア語の 9 言語である。また、PHP と MySQL を使用して Web 上での収集を可能としている。

本システムの主要機能は以下に示す 3 つである。

1. 用例の提案

医療従事者や患者などが他の言語に翻訳してほしい用例を提案する機能である。実際に用例を使用する利用者がそれぞれの立場から提案するため、必要な用例を集めることができる。

2. 対訳の作成

「用例の提案」で提案された用例を翻訳者が翻訳する機能である。医療分野では正確な翻訳が必要なため、本機能は翻訳者のみが利用する。

3. 用例対訳の検索

本システム内の用例対訳を検索する機能である。本機能は、医療従事者や患者、翻訳者などすべての利用者が利用可能である。



図 1: TackPad の画面例

2.2 評価機能

用例を評価する方法として、3 つの評価手法を用意した。各評価手法を図 2 に示す。なお、各評価手法をそれぞれ評価手法 A、評価手法 B、評価手法 C とする。

1. 評価手法 A (5 段階評価) (図 2-(1))

文献 [5] で既実装済みの評価手法である。星をクリックすることで、用例の評価を可能としている。

2. 評価手法 B (チェックボックスによる評価) (図 2-(2))

予め用意した文に対して同意できるかどうかを、チェックボックスで行う評価手法である。評価手法 A では一つの評価軸のみで評価を行っていたが、評価手法 B では複数の評価軸で評価を可能とした。

3. 評価手法 C (複合評価) (図 2-(3))

予め用意した対訳語の組に対して、5 段階評価を行う評価手法である。評価手法 B は評価の度合いを表現できないが、評価手法 C は可能とした。

3. 試用実験

本章では、2.2 節で述べた評価手法の評価実験について述べる。本実験の目的は、3 つの評価手法のうち用例の評価として適している評価手法の確認である。

実験対象の用例として、30 文の日本語の用例を用いた。これは、事前に著者の一人が下記の条件を満たす用例を用意したものである。

1. 被験者が実験対象用例を作成していない

2. 被験者が実験対象用例を既に評価していない

被験者は 30 文の用例それぞれに対し、3 つの評価手法を用いて評価を行った。なお、3 つの評価手法の試用順序は被験者ごとに順序交換を行っている。また、実験後にアンケートへの回答を依頼している。

[†]和歌山大学[‡]多文化共生センターきょうと

(1) 評価手法 A (5 段階評価)



(2) 評価手法 B (チェックボックスによる評価)

おすすめポイント	みんなのおすすめポイント
<input checked="" type="checkbox"/> この文例は使える!	1
<input checked="" type="checkbox"/> わかりやすい	1
<input type="checkbox"/> 対訳がほしい	
<input type="checkbox"/> 表現がぎこちない	

(3) 評価手法 C (複合評価)

文語的	<input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	口語的
あまり使わない	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/>	よく使う
流暢でない	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/>	流暢

図 2: 3 つの評価手法

被験者は、本システムの利用者である和歌山大学の学生 15 名である。今回の被験者は患者の立場で用例の評価を行った。しかし、本システムは翻訳者や医療従事者なども利用者である。このため、今回の実験結果は限定的であると考えられる。

4. 実験結果と考察

本章では 3 章で行った実験の結果とその考察を行う。被験者が各評価手法の評価の行いやすさを 5 段階で評価した結果を表 1 に、3 つの評価手法を評価しやすかった順に順位付けした結果を表 2 にそれぞれ示す。表 1、表 2 から、試用実験では評価手法 A が他の評価よりも評価を行いやすい傾向が見られた。ただし、表 1 の結果を用いてフリーマン検定を行った結果、有意確率は 0.115 となり有意差は見られなかった。

また、評価手法 A に対しては、「評価の尺度が一つで漠然としていてどう評価していいか戸惑った」「どんな点がおすすめなのかは伝えにくいと思う」などの意見をアンケートの自由記述から得ることができた。これは、評価軸が一つであるために曖昧さが発生しているためであると考えられる。このため、評価手法 A は簡易的な評価は可能だが、詳細な評価は難しいと考えられる。

評価手法 B や評価手法 C に対しては、「おすすめポイント (評価手法 B) の内容が分かりにくい (評価手法 B の自由記述)」「文語的、口語的の判断基準が分かりにくかった (評価手法 C の自由記述)」などの意見をアンケートの自由記述から得ることができた。これは、評価手法 B、評価手法 C の評価項目の内容が「誰に対しておすすめするか」が曖昧なためであると考えられる。評価手法 B や評価手法 C に対して使いにくいと回答した被験者の多くは、評価手法ではなく評価項目に問題があったと回答していた。このため、評価項目の内容は誰に対する評価であるかをわかりやすく提示する必要があると考えられる。また、評価項目の中に翻訳者向けの項目も含まれていた。このことが原因で被験者の評価が下がっていたこ

表 1: 評価の行いやすさ (5 段階評価)

	評価手法 A	評価手法 B	評価手法 C
平均	3.93	3.20	3.27
標準偏差	0.70	1.08	0.96

・質問:「この評価のやり方は、用例の評価を行いやすかった」
 ・評価項目は、1:強く同意しない, 2:同意しない, 3:どちらとも言えない, 4:同意する, 5:強く同意する, の 5 段階である。

表 2: 評価の行いやすさ (順位付け)

順位	評価手法 A	評価手法 B	評価手法 C
1	9 人	3 人	3 人
2	3 人	6 人	6 人
3	3 人	6 人	6 人
合計	24	33	33

・3 つの評価手法の順位付けを被験者が行った結果である。
 ・表中の「合計」は評価手法ごとに「順位×人数」を合計した結果である。

とも考えられるため、評価を行う利用者の属性ごとに評価項目を変更する必要があると考えられる。

5. おわりに

本稿では、多言語用例対訳共有システム TackPad の用例評価機能の開発とその評価について述べた。試用実験の結果、評価手法 A (5 段階評価) が評価を行いやすい評価手法である傾向が見られたが、本実験は患者のみが評価を行っているため結果は限定的である。今後、本実験の結果をもとに医療従事者や翻訳者も含めて各評価手法の評価実験を行う必要がある。また、評価手法 B (チェックボックスによる評価) や評価手法 C (複合評価) の評価項目の内容を誰に対する評価であるかをわかりやすく提示したり、評価項目を評価者の属性に合わせて変更したりする必要があると考えられる。

謝辞

本研究の一部は総務省の戦略的情報通信研究開発推進制度 (SCOPE) の平成 20 年度採択課題「多言語共生社会における医療対話支援のための多言語対話用例プラットフォームの構築」による。

参考文献

[1] 法務省: <http://www.moj.go.jp/PRESS/>
 [2] 田村太郎: 多民族共生社会ニッポンとボランティア活動, 明石書店 (2000).
 [3] Takano, Y., et al: A temporary decline of thinking ability during foreign language processing, Journal of Cross-Cultural Psychology, 24, pp.445-462(1993).
 [4] 宮部真衣ほか: 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, 信学論, Vol.J92-D, No.6, pp.708-718(2009).
 [5] 福島拓ほか: 多言語用例対訳共有システム TackPad の評価機能の実現と評価, 情処研報, 2009-GN-70(21), pp.121-126(2009).