

正確な情報共有のための多言語用例対訳共有システム

福島 拓^{1,a)} 吉野 孝^{2,b)} 重野 亜久里^{3,c)}

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. この問題は, 正確性が求められる分野や緊急時において顕著に現れ, 解決が求められている. このため, 正確な情報共有を可能にする一技術である用例対訳を用いた支援が行われており, 用例対訳の作成も多く行われている. しかし, 従来の用例対訳作成においては, (1) 新たな用例対訳の追加が難しい, (2) 翻訳者一人あたりの負担が大きい, (3) 用例対訳の利用現場で求められている用例を収集することが難しい, という問題が存在していた. そこで本稿では, 用例対訳の作成の基盤となる場を用例対訳作成環境に提供し, 円滑に正確な用例対訳の収集を行うことを目的とした, 多言語用例対訳共有システムを提案し, 実装を行った. また, 実際のシステム構築後に顕在化した問題点とその解決策について述べる.

キーワード: 情報共有支援, 多言語, 用例対訳

Multilingual Parallel-text Sharing System for Accurate Information Sharing

TAKU FUKUSHIMA^{1,a)} TAKASHI YOSHINO^{2,b)} AGURI SHIGENO^{3,c)}

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. This problem appears conspicuously in medical field, emergency, and so on. To solve this problem, the scene in necessary accurate communication has been using parallel-text. Parallel-text is able to support accurate information sharing. There are parallel-text creating projects. However, existing parallel-text creating projects have following problems. (1)It is difficult to obtain a new parallel-text. (2)To translate many texts is a heavy burden for translators. (3)It is difficult to collect example sentences in the using field of parallel-text. Therefore, we proposed and implemented multilingual parallel-text sharing system to create accurate parallel-text. This system provided a base field of creating parallel-text to groups of existing creating parallel-text. Moreover, this paper shows discovered problems in operation of the system and solves these problems.

Keywords: information-sharing support, multilingual, parallel-text

¹ 和歌山大学大学院システム工学研究科
Graduate School of Systems Engineering, Wakayama University

930 Sakaedani, Wakayama 640-8510, Japan

² 和歌山大学システム工学部
Faculty of Systems Engineering, Wakayama University
930 Sakaedani, Wakayama 640-8510, Japan

³ 特定非営利活動法人 多文化共生センターきょうと
Center for Multicultural Society Kyoto
21 Sakai-machi, Shimogyo-ku, Kyoto 600-8191, Japan

a) fukushima@yoslab.net

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

c) aguri@tabunka-kyoto.org

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]。

医療分野における多言語間コミュニケーション支援では正確性の確保が可能な用例対訳が多く用いられている。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。なお、用例対訳は正確性が必要な分野で利用できるため、医療分野 [9] のみではなく、防災の分野 [10] などでも利用されている。医療分野の用例対訳は、多言語間コミュニケーションの機会の増加を背景に、様々な医療機関や医療 NPO が作成を行っている。しかし、従来の用例対訳作成には新規用例対訳の追加や翻訳者の負担などに関する問題が生じていた。

そこで我々は、これらの問題を解決するために、ICT を利用した用例対訳の収集システムを提案、実装を行った。本研究では、用例対訳の作成の基盤となる場を用例対訳作成環境に提供し、円滑に正確な用例対訳の収集を行うことを目的とする。なお、本稿では収集の観点から検証を行い、用例対訳の提供に関しては議論を行わない。

本稿では、従来の用例対訳作成の問題点と関連研究について述べた後、問題の解決を目指した多言語用例対訳共有システムとその機能について説明を行う。その後、有用性確認実験と実運用からの考察を行い、最後にまとめを行う。

2. 従来の用例対訳作成の問題点

本章では、従来の用例対訳作成の場で生じていた問題点について述べる。用例対訳は用例対訳の利用者や翻訳者が協力して作成を行っている。一般的な用例対訳の作成手順(以降、従来手順とする)は次のようになっている。

従来手順 (1) 用例対訳の利用者(医療分野の場合、医療従事者や医療 NPO 関係者)が中心となり、用例対訳の元となる用例を選定する。

従来手順 (2) 選定された用例群を翻訳者(通常、各言語につき1名)に依頼して多言語翻訳を行う。

従来手順 (3) 用例対訳の利用者や翻訳者による用例対訳の正確性確認後、用例対訳を利用可能な形に加工する。

このようにして作成された用例対訳は冊子形式や Web 上での公開、ICT を用いた多言語システムなどで用いられ

ている。また、(1) 用例対訳の作成は常に行われているわけではなく、一度に作られることが多い、(2) 翻訳者は翻訳能力の高い人が担当している、という特徴がある。このため、正確な用例対訳を一度に作成できるという利点がある。

しかし、従来手順には次のような問題がある。

課題 1 新たな用例対訳の追加が難しい。

元となる用例の選定後に新たに用例を追加することは、常に作業を行っているわけではない従来の用例対訳作成の仕組みでは難しい。特に、従来知られていなかった病気が流行したときなど、従来の用例対訳では対応できない場合に、この問題が顕著に現われる。

課題 2 翻訳者一人あたりの負担が大きい。

従来の仕組みでは、翻訳能力が高い翻訳者が翻訳作業を担っている。このことは正確な文の作成に寄与しているが、簡単な文も難しい文も翻訳能力の高い翻訳者が担うため、翻訳者一人あたりの負担が大きくなる。前述の通り、特に医療分野の翻訳者は不足しており、翻訳者の負担軽減が求められている。

課題 3 用例対訳の利用現場で求められている用例を収集することが難しい。

従来の手順では、用例の選定は網羅性を高めるために用例対訳を実際に利用している専門家が行っている。しかし、専門家は用例の選定のみを専門業務として行うことは少なく、他の通常業務の合間に用例の選定作業を行っている場合が多い。医療分野の場合、問診中、診察中など、実際に用例対訳を使用する場である、患者と対応時 以外 で用例の選定を行っているため、必要な用例が収集できていない可能性がある。

本稿では、これらの問題解決を行う用例対訳収集システムについて述べる。

3. 関連研究

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 [11] や多言語対面環境の討論支援 [12] 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 [13]。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 [14]、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 [15] 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

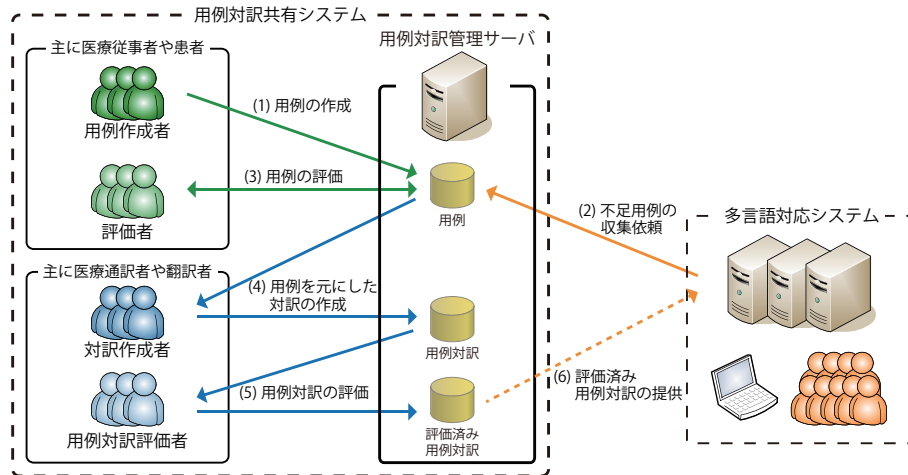


図 1 用例対訳作成の流れ

Fig. 1 Flow of creating parallel texts.

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。

用例対訳の作成は、2章で述べた従来手順で作成されることが多いが、Web上での収集する取り組みも行われつつある。ChenらはWeb上にある用例対訳を自動的に収集する試みを行っている[16]。この研究は、翻訳に関するコストが少ないというメリットがある。しかし、用例対訳の提供先で求められている用例対訳を収集することが難しい。このため、本研究では従来手順と同様に、用例対訳を使用する利用者から用例の提案を受け、その内容を用例対訳にするという形式を取る。

Web上での言葉の収集として、みんなの翻訳[17]や訳してねっと[18]がある。みんなの翻訳はWeb上で文章の翻訳を目的としており、訳してねっとは機械翻訳で使用する単語をWeb上で収集することを目的としている。また、我々と同じ着想でWeb上での用例対訳の収集も行われている。BondらはTanaka Corpus[19]を基に、用例対訳の収集プロジェクトを行っている[20]。このプロジェクトはTATOEBAPROJECTという名前で活動が行われており、日常的に使用する用例の収集を、日本語、英語、フランス語、中国語、ドイツ語など様々な言語で行っている。このようなWeb上での知識の収集は、様々な人から広く情報を集める利点があり、用例対訳の収集にも向いていると考えられる。この研究では、用例対訳の提供先で求められている用例対訳を収集している点で、我々と着想は同じである。しかし、正確性評価が行われていない用例対訳には不正確なものが含まれている[21],[22]が、文献[20]では従来手順(3)のような用例対訳の提供前に正確性評価を行っていない。このため、用例対訳の十分な精度が保てているとは言えない。本研究は言葉の正確性が求められる分野の用例対訳の収集を目的としている。これらの分野では、正確性が十分確保されている用例対訳を使用する必要

があるため、この問題の解決は重要なものとなる。また、文献[20]ではWeb上でのみ用例対訳の収集を行っている。このため、従来手順と同様に、使用される場面を考慮しながらの用例対訳の作成を行う必要があるため、実際に必要な用例対訳を網羅することが難しい。本研究ではこれらの問題を考慮して用例対訳の収集を目指す。

4. 用例対訳収集の方針

本章では、本研究における用例対訳収集の方針について述べる。なお、提案システムは医療分野の用例対訳収集を対象としているが、医療分野に特化していないため、正確性が求められる分野の用例対訳収集に応用が可能である。

本システムでの用例対訳の作成の流れを図1に示す。本システムでは、図1-(1)や図1-(2)で用例対訳化する用例を取得する。取得した用例は図1-(3)の評価機能で評価を行い、図1-(4)の対訳作成機能で用例対訳化を行う。その後、図1-(5)で用例対訳の正確性評価を行い、図1-(6)で用例対訳の提供を行う。図1-(1)~(5)の各項目の説明については5章で行う。なお、現時点では図1-(6)の用例対訳の提供は行っていない。

本システムでは、従来手法の課題1(新たな用例対訳の追加が難しい)、課題2(翻訳者一人あたりの負担が大きい)に対応するために、常に用例対訳を収集可能なWebシステムとした。Web上で常に用例対訳の作成を行うため、新たに用例対訳が必要になった場合も迅速な対応が可能である。また、従来手法のように一度に翻訳を行わず、翻訳者の時間が空いた時に翻訳可能な文を翻訳する仕組みとすることで、翻訳能力が低い翻訳者も参加可能としている。これは、本システムは翻訳をボランティアベースで行っていることも理由の一つである。また、従来手法の課題3(用例対訳の利用現場で求められている用例を収集することが難しい)に対応するために、図1-(2)の「他の多言語対応シ

表 1 従来システムと提案システムとの差異

Table 1 Differences of existing systems and proposal system.

	従来手法	文献 [20]	提案システム
主な作成者	用例対訳の利用者	Web 利用者	Web 利用者
関係者数	少ない	多い	多い
収集対象分野	正確性が必要な分野	日常用語	正確性が必要な分野
正確性の評価	あり (1~2名)	なし	あり (複数人)
用例対訳の収集場所	利用現場外のみ	利用現場外のみ	用例対訳の利用現場と利用現場外の両方

システムからの不足用例の収集機能」を有している。なお、この機能については 5.2 節で詳しく述べる。

本システムでは、従来手順で関わっていた医療従事者や翻訳者、医療 NPO 関係者の他に、医療機関を受診した患者も利用者としている。これは、利用者の限定による用例対訳の多様性が失われる可能性があるためである。このため、本システムは Web 上で広く登録を可能とした。また、本システムはユーザ登録を行った利用者のみ用例の作成、対訳の作成を可能としている。なお、本システムのユーザ権限は、一般利用者と管理者のみとしている。従来手順では、各段階ごとに関わる人が決められていた。しかし、本システムは自由なユーザ登録を可能としていることもあり、各利用者の属性を正確に把握することは難しい。また、課題 2(翻訳者一人あたりの負担が大きい)の解決には、多くの人が本システムの用例対訳作成プロセスに関わる必要がある。このため、本システムでは一般利用者の権限はすべて同じとしている。このように、用例の作成、対訳の作成、用例や用例対訳の評価それぞれに、様々な属性の利用者が関わり正確な用例対訳の作成を目指している。

本手法と従来手法、文献 [20] との差異を表 1 に示す。本手法は Web 上での用例対訳収集を行っているため、従来手法より用例対訳の登録が可能な関係者数が多い特徴がある。多くの人から様々な用例対訳の収集が可能であるため、従来手法より優位であると考えられる。また、収集対象が正確性が必要な分野であるため、正確性評価を行っている特徴がある。この点で、文献 [20] より優位であると考えられる。用例対訳の網羅性においては、実際に用例対訳の利用現場での収集機能があるため、従来手法や文献 [20] より優位であると考えられる。

5. システム設計

本章では、4 章をもとに実装を行った、医療分野を対象とした多言語用例対訳共有システム TackPad の設計について述べる。開発は PHP と JavaScript を用いて行い、Web 上での用例対訳の収集を可能とした。本システムの収集言語は、日本語、英語、中国語、韓国・朝鮮語、ポルトガル語、スペイン語、タイ語、ベトナム語、インドネシア語の 9 言語である。以降、各機能についてそれぞれ述べる。

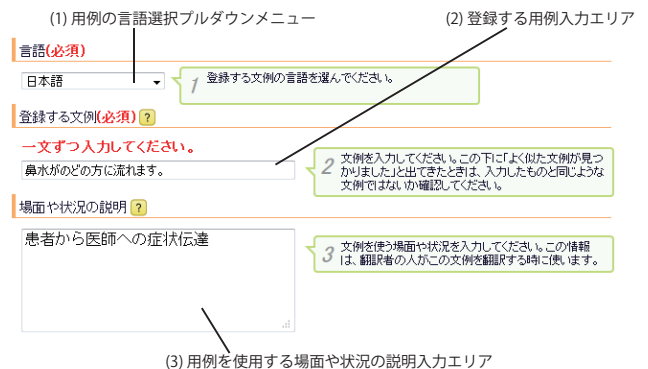


図 2 用例作成機能の画面例

Fig. 2 Screenshot of function of creating example sentences.

5.1 用例作成機能

本節では、用例作成機能について述べる。本機能は、図 1-(1) にあたる。

本機能では、主に医療従事者や患者が、医療現場で使用される用例の提案を行う。本機能の画面例を図 2 に示す。入力必須項目は図 2-(1) の「用例の言語」と、図 2-(2) の「登録する用例*1」のみとし、利用者の負荷軽減を行っている。また、任意入力項目として、図 2-(3) の「用例の使用場面や状況の説明」を留意している。本システムで収集している用例は一文としており、比較的短い言葉となる。短い言葉である用例は、前後の文脈がないと意味が分かりにくいものが発生しやすい [23]。このような状況を防ぐために、「用例の使用場面や状況の説明」の項目により用例の補足を可能としている。

5.2 不足用例の収集機能

本節では、不足用例の収集機能について述べる。本機能は、課題 3(用例対訳の利用現場で求められている用例を収集することが難しい)の解決を目指したものであり、図 1-(2) にあたる。

本機能は、既に用例対訳を利用しているシステムから、不足している用例対訳の情報を受け、その情報から新たな用例対訳の作成を行うことを実現する。現在、多言語問診票作成システム [24] と携帯型多言語間医療対話支援システ

*1 システム内では、より一般的な「文例」という言葉を使用している。

ム「ぷち通」[25]、多言語医療受付支援システム M^3 [9] の各システムから不足用例作成依頼を受けている。各システムの利用場面は異なるが、医療従事者と患者の多言語間コミュニケーション支援を目的としたシステムである。

多言語問診票作成システムと「ぷち通」は、用例対訳と機械翻訳を併用しており、自由文の入力が可能である。このため、これらのシステムに入力された文は利用者が実際に使用したい文であり、用例対訳化が潜在的に求められている文である場合が多い。両システムとも、用例対訳が存在する場合は用例対訳を、存在しない場合は機械翻訳の翻訳結果をそれぞれ提示する。本機能は、正確性が不十分な可能性のある機械翻訳を利用した文の提供を受けている。これにより、医療現場で実際に使用される文を、利用者の負担無く収集できる仕組みを実現している。また、 M^3 は不足している用例を医療従事者がシステム管理者に連絡する機能があり、本機能はその機能と連携を行っている。ただし、本機能は試用段階であるため、今後問題点の抽出を行う必要があると考えられる。

5.3 用例評価機能

本機能は、作成された用例が正確であるかどうか評価する機能であり、図 1-(3) にあたる。本機能は、作成者への修正依頼と、5段階評価の2つの項目から構成している。なお、本機能で用例の評価が行われていない用例に関しても、次節で述べる対訳作成は可能とした。これは、本機能は不正確な用例の抽出を目的としており、正確な用例に対しては評価や修正依頼がつかない可能性があるためである。

5.3.1 作成者への修正依頼

本システムでは、不必要な編集合戦^{*2}を防ぐために他の利用者が作成した用例の編集機能を与えていない^{*3}。このため、用例作成者以外が用例の修正が必要な場合は、用例作成者へ修正依頼を行うという形を取っている。修正依頼の内容は、あらかじめ用意した「スペルが間違っています」「文法が間違っています」「言語選択が間違っています」「入力文字に問題があります」「アクセント記号がありません」の5つの内容から選択する形式を取っている。これは、自由文による利用者同士の不必要な争いを避けるためである。

5.3.2 評価軸による用例の分類

正確な評価には、評価軸を明確にし、複数段階が選択可能な評価項目を用いた評価手法が必要であることが分かっている [26]。そこで本機能では、評価段階を5段階とし、軸の両端に反対の意味の評価を置く形式を取り、「病院であまり使わない - 病院でよく使う」という評価軸を用意している。この軸は、医療分野の用例対訳でない用例を取り除く目的で行っている。

^{*2} 自分の主張を押し通すために、2人以上が互いに内容の修正を行う合うことを指す。

^{*3} ただし、管理者には修正権限を与えている。

5.4 対訳作成機能

本節では、対訳作成機能について述べる。本機能は、図 1-(4) にあたる。本機能では、主に医療通訳者や翻訳者が 5.1 節で作成された用例を他の言語に翻訳を行う。

本機能は次の手順で利用を行う。

- (1) 利用者は、翻訳元となる用例の言語と翻訳先の言語を指定する。これは、翻訳すべき用例は多く存在する上、利用者によって翻訳が可能な言語も異なるため、翻訳対象の用例の絞り込みを行うためである。
- (2) システムは、前項で入力された情報を元に、翻訳先に指定された言語に翻訳が行われていない用例を表示する。日本語と英語が理解可能な利用者が、翻訳元を日本語、翻訳先を英語とした場合を例としてあげる。この場合、システムは英語に翻訳されていない、日本語の用例一覧をシステムは表示する。
- (3) 利用者は一覧表示された用例から翻訳するものを選択し、翻訳を行う。なお、入力項目は 5.1 節の用例作成機能と同様のものが表示される。

5.5 用例対訳評価機能

本機能は、作成された用例対訳が正確であるかどうか評価する機能である。本機能は、図 1-(5) にあたる。本機能で使用する評価基準は、2言語間の意味比較に用いられる Walker らの適合性評価基準 [27] を参考に^{*4}作成し、5：意味は完全に一緒、4：文法などに多少問題あるが、意味はまあまあ一緒、3：意味はだいたいつかめる程度に一緒、2：雰囲気は残っているが、意味は一緒ではない、1：意味は全く違う、の5段階で評価を行っている。

評価数と評価値の平均がそれぞれ一定値を超えた用例対訳は、他の多言語対応システムに提供を行う予定である。

従来手順での評価は、1名～2名程度で行われることが多いが、本システムでは様々な利用者が存在するために、従来手法よりも多い人数で評価を行う必要があると考えられる。現時点では、1つの用例対訳に対して5人以上の評価者が評価をつけ、かつ、評価値の平均が4以上である場合に用例対訳の提供を検討している。しかし、この基準に関しては実験などから得られたものではないため、今後検証が必要であると考えられる。

6. 多言語用例対訳共有システムの実環境への適用

本システムは 2008 年 2 月から、実際に Web 上で運用を行っている。現在、約 200 名がシステムに登録しており、全言語合計で約 14,500 件^{*5}の用例の登録が行われている^{*6}。

^{*4} Walker らの適合性評価の基準は、5:All, 4:Most, 3:Much, 2:Little, 1:None, である。

^{*5} うち約 10,000 件は既存の用例対訳に登録している。

^{*6} 2012 年 4 月現在。

表 2 被験者の性別

Table 2 Sex distribution among subjects.

	男性	女性
人数	8	24

表 3 被験者の年代

Table 3 Age distribution among subjects.

	10 代以下	20 代	30 代	40 代	50 代以上
人数	1	7	13	7	4

表 4 被験者の所属

Table 4 Occupations of subjects.

	通訳	ボランティア	NPO 職員	医療関係者	その他
人数	13	7	6	4	12

・複数選択可能

本章では、運用初期に行った実験について述べた後、運用中に得られた 2 つの問題とその解決策について述べる。

6.1 有用性確認実験

多言語用例対訳共有システム TackPad を使用し、実際に用例作成、対訳作成を行う試用実験を行った。実験の目的は、本システムの有用性評価である。評価対象は、実験時にシステムに実装を行っていた用例作成機能 (図 1-(1)) と対訳作成機能 (図 1-(4)) である。

6.1.1 実験概要

被験者は医療 NPO 法人で募集を行った。アカウント発行人数は 56 人、一度以上のアクセスがあった人数が 52 人、アンケートの回答を得た人数は 32 名である。なお、実験は 21 日間行った。実験では実際の使用状況を想定し、自由にシステムを操作するよう依頼した。なお、参加日時や最低参加時間、用例や対訳の作成数の制限などは行っていない。その後、アンケートへの回答を依頼した。

6.1.2 実験結果と考察

アンケート結果から得られた被験者の性別を表 2 に、年代を表 3 に、所属を表 4 にそれぞれ示す。性別は女性が、年代は 30 代の人が多い傾向となっている。また、通訳やボランティアが多く含まれているが医療関係者も 4 名含まれており、本システムが想定している利用者が被験者になっていることが分かる。

実験期間中に登録された、言語別の用例登録数を表 5 に示す。表 5 より、特に目標設定を行っていないにも関わらず、用例を比較的多く収集できたことが分かる。

また、システムに関するアンケートから、「このようなシステムを待っていた」「ぜひこれから使っていきたい」などが翻訳者から得られた。本システムは、翻訳者に依存している部分が多い。また、翻訳者に対するメリットも小さいと考えていたことから、あまり良い評価は得られないのではないかと危惧していた。しかし、肯定的な意見を多く

表 5 言語別の用例登録数

Table 5 Number of registered parallel texts for each language.

言語名	登録用例数
日本語	154
英語	91
中国語	48
スペイン語	31
ポルトガル語	30
韓国・朝鮮語	29
合計	383

表 6 利用者の属性別登録用例数

Table 6 Number of registered parallel texts in each affiliation.

	人数 (人)	登録用例 (文)	
		合計	平均
翻訳者	10	113	11.3
翻訳者以外	10	207	20.7
不明	33	332	10.1
全体	53	652	12.3

・日本語以外の言語の用例を登録した利用者の人数と登録用例数である。

聞くことができた。これらのことから、本システムの有用性を確認できたと考えられる。

ただし、その後の運用で翻訳者の負担に関する影響が出てきている。これらについては、6.2 節で詳しく述べる。

6.2 翻訳者の負担軽減

用例対訳の収集を行う場合、用例の翻訳が必要なため翻訳者に負担がかかる。このため、本システムは 4 章で述べたとおり、翻訳能力にかかわらず、すべての利用者が翻訳できる仕組みとしている。本節では、本システムが翻訳能力の高い翻訳者の負担軽減につながっているかを議論した後、翻訳者の翻訳支援機能について述べる。

まず、各利用者の用例の翻訳に関して調査を行った。日本語以外の用例を登録した利用者の用例登録数を、属性別に分類したものを表 6 に示す。なお、システムへ属性の登録がない利用者の場合は不明とした。また、利用者の属性は自己申告である。表 6 より、53 人の利用者が日本語以外の用例を 652 用例、登録していることが分かる。また、翻訳者以外の利用者 (翻訳能力が低い) は翻訳者の約 2 倍の用例を作成していることが分かる。このことから、翻訳者以外による翻訳作業が行われており、翻訳能力の高い利用者の負担軽減が行われていると考えられる。

次に、翻訳者の翻訳支援機能について述べる。本システムの利用者に直接聞き取りを行った際に、「何もないところからの翻訳は少し敷居が高い」という意見が得られた。本システムのような用例対訳の収集システムでは、対訳作成による多言語化を行うことが最重要の課題である。このため、対訳作成支援として機械翻訳を利用した。3 章の関



図 3 機械翻訳を利用した対訳作成支援機能画面例

Fig. 3 Screenshot of a function of creating translated text used machine translation.

連研究でも述べたとおり、機械翻訳はそのまま医療分野で利用可能な精度には達していない。しかし、機械翻訳はすべての文を翻訳できるという特徴があるため、対訳作成時の手がかりになると考えられる。また、「少し間違えている言葉を見ると修正したくなる」という利用者からの意見も参考にしている。

機械翻訳を利用した対訳作成支援機能の画面例を図 3 に示す。図 3 は日本語用例の「あなたはどこに住んでいますか?」のページである。図 3-(1) は、既に登録されている、利用者によって作成された対訳である。また、図 3-(2) は、機械翻訳によって翻訳された用例である。機械翻訳は、利用者による対訳作成が行われていない言語に対して行っている。利用者は、図 3-(2) の「修正する」ボタンをクリックし、機械翻訳の文を正しく修正することができる。このようにして、新たな対訳作成を支援している。なお、機械翻訳は言語グリッドが提供するものを利用した。言語グリッドは、機械翻訳や用例対訳を組み合わせて利用可能にする言語基盤構築プロジェクトである [28], [29]。

6.3 類似用例の登録

本システムは、従来手法とは異なり、利用者すべてが用例の作成が可能であるという特徴がある。しかし、このことは他の利用者と協調せずに用例が作成できることも意味する。本システムでは、完全に同じ文の登録はできないが、類似文に関しては登録可能となっている。これは、類似文の判定をシステムが完全に行うことが難しいためである。このため、本システムに登録された用例には、似た意味の言葉が複数存在している場合がある。例えば、頭の痛みに関する用例としては、「頭が痛い」「頭が痛いです」「頭が鈍く痛いです」「頭の片側が痛いです」など、様々な種類が登録されている。特に、「頭が痛い」「頭が痛いです」の 2 文に関してはほぼ同じ意味であり、これらを元にした対訳作

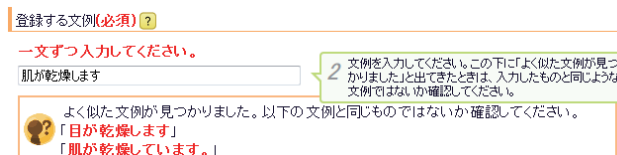


図 4 登録済み類似文提示機能画面例

Fig. 4 Screenshot of function of showing registered similar texts.

成を行った場合、二度手間になることが危惧された。

このため、システムが入力文と登録済みの用例が類似していると判定した場合にその旨を利用者に提示する、登録済み類似文提示機能を作成した。登録済み類似文提示機能の画面例を図 4 に示す。図 4 は用例作成者が「肌が乾燥します」という用例をシステムに登録するために入力した例である。このとき、システムは類似文「目が乾燥します」と「肌が乾燥しています」を提示する。用例作成者は、システムが提示した内容を見て、このまま登録するかどうかを判断することとなる。なお、登録済み類似文提示機能は提示のみをおこなっており、前述の通り類似文が提示された用例でも登録は可能である。このため、類似用例を登録するかどうかは、利用者の判断にゆだねることになり、本機能導入前までの類似用例が多く登録される事態は一定程度防ぐことができると考えられる。

登録済み類似文提示機能は、N-gram に基づく用例対訳検索手法を利用している [30]。文献 [30] では、用例を言語によって 2-gram もしくは 4-gram に分割し、検索文字列との共起を調べることで多言語の類似文検索を実現している。本システムではこの内容を応用し、ある一定以上類似していると判断された文が入力された場合は図 4 のように類似文を提示している。

7. おわりに

本稿では、多言語間での正確な情報共有を可能にする用例対訳に着目し、正確な用例対訳の円滑な作成支援を目的とした多言語用例対訳共有システムを提案し、実装した。本システムは、従来の用例対訳作成コミュニティが抱えていた、(1) 新たな用例対訳の追加が難しい、(2) 翻訳者一人あたりの負担が大きい、(3) 用例対訳の利用現場で求められている用例を収集することが難しい、の各課題の解決を目指している。

本研究の貢献は次の 3 つにまとめられる。

- (1) 従来の用例対訳収集コミュニティにおける問題解決のための用例対訳収集の仕組みを提案し、実現した。
- (2) 翻訳能力が低い翻訳者でも翻訳業務に参加可能な仕組みを提案し、実現した。
- (3) 用例収集時に発生する類似用例に関する問題について述べ、その解決策を提案した。

今後、用例の網羅性を高めるために不足用例収集機能を

実システムで運用し、さらなる用例の収集を目指す。また、収集した用例の提供基準を定め、他の多言語対応システムへの提供を目指す。

謝辞 本研究の一部は、科研費基盤研究(B)(22300044)および、総務省戦略的情報通信研究開発推進制度(SCOPE)の平成22年度採択課題「医療現場における利用者適応型多言語間コミュニケーション支援のための基盤技術の研究開発」による。

参考文献

- [1] 法務省：平成22年末現在における外国人登録者統計について、法務省(オンライン)、入手先〈http://www.moj.go.jp/nyuukokukanri/kouhou/nyuukantourokusya_toukei110603.html〉(参照2012-04-01)。
- [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2012-04-01)。
- [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2012-04-01)。
- [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2012-04-01)。
- [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] Hasegawa, S., Sato, K., Matsunuma, S., Miyao, M. and Okamoto, K.: Multilingual disaster information system: information delivery using graphic text for mobile phones, *AI & Society*, Vol. 19, No. 3, pp. 265–278 (2005)。
- [11] 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)。
- [12] 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)。
- [13] 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)。
- [14] 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)。
- [15] 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)。
- [16] Chen, J., Chau, R. and Yeh, C.-H.: Discovering Parallel Text from the World Wide Web, *ACSW Frontiers'04: Proceedings of the second workshop on Australasian information security, Data Mining and Web Intelligence, and Software Internationalisation*, Vol. 32, pp. 157–161 (2004)。
- [17] Utiyama, M., Abekawa, T., Sumita, E. and Kageura, K.: Minna no Hon'yaku: A website for hosting, archiving, and promoting translations, *Translating and the Computer 31 Conference* (2009)。
- [18] Shimohata, S., Kitamura, M., Sukehiro, T. and Murata, T.: Collaborative Translation Environment on the Web, *Machine Translation in the Information Age* (2001)。
- [19] Tanaka, Y.: Compilation of a multilingual parallel corpus, *Proceedings of PACLING 2001*, pp. 265–268 (2001)。
- [20] Bond, F., Nichols, E., Appling, D. S. and Paul, M.: Improving Statistical Machine Translation by Paraphrasing the Training Data, *Proceedings of IWSLT 2008*, pp. 150–157 (2008)。
- [21] Breen, J. W.: Word Usage Examples in an Electronic Dictionary, *Papillon (Multi-lingual Dictionary) Project Workshop* (2003)。
- [22] 福島 拓, 吉野 孝, 田淵裕章, 北村泰彦：多言語用例対訳を用いたコミュニケーションのための応用例対作成システムの開発, マルチメディア, 分散, 協調とモバイル(DICOMO2009)シンポジウム, pp. 1612–1618 (2009)。
- [23] 上田和子, ジョイデヴェラ, 水野真木子, 角南北斗, 原田マリアフェ：『日本語でケアナビ』と実践的コミュニティー, 国際交流基金関西国際センター日本語教育シンポジウム(2008年3月8日), パネルディスカッション資料, 泉南郡田尻町(2008)。
- [24] 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)。
- [25] 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)。
- [26] 福島 拓, 吉野 孝：用例の正確性評価を目的とした用例評価手法の比較, 情報処理学会論文誌, Vol. 52, No. 1, pp. 131–139 (2011)。
- [27] Walker, K., Bamba, M., Miller, D., Ma, X., Cieri, C. and Doddington, G.: Multiple-Translation Arabic (MTA) Part 1, *Linguistic Data Consortium, Philadelphia* (2003)。
- [28] Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration, *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp. 96–100 (2006)。
- [29] Sakai, S., Gotou, M., Tanaka, M., et al.: Language Grid Association: Action Research on Supporting the Multicultural Society, *International Conference on Informatics Education and Research for Knowledge-Circulating Society (ICKS-08)*, pp. 55–60 (2008)。
- [30] 田淵裕章, 坂本 廣, 北村泰彦：N-gramに基づく用例対訳検索手法, 電子情報通信学会技術研究報告, 人工知能と知識処理研究会, Vol. AI2008-52, pp. 43–48 (2009)。