応答用例対と穴あき用例を活用した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．正確な多言語間対話支援が求められる場では，用例対訳が多く用いられている．また，用例対訳を質問と回答の対の形で保存した応答用例対を用いることで，より正確な多言語間対話支援が可能となる．しかし，応答用例対は多言語間対話支援システムの入力ログからの生成や，システムの管理者による登録が主なものとなっており，応答用例対の数の増加が課題となっていた．そこで本稿では，用例対訳の一つであり，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の概念を活用して，応答用例対および用例の作成を行う手法を提案する．本稿の貢献は，応答用例対と穴あき用例を活用した用例対訳作成手法を提案し，応答用例対の回答となる用例作成ができることを示した点である．

Proposal of Parallel Texts Creation Method Using Dialogue Parallel Texts and Perforated Texts

Taku Fukushima[1,a)]　Takashi Yoshino[2,b)]

1. まえがき

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．日本国内でも在留外国人数や訪日外国人数は増加傾向にあり [1], [2]，今後，外国人住民のさらなる増加が予想されている [3]．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 [3]，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 [4], [5], [6]，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 [7]．

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 [8], [9], [10], [11]．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 [12]．収集した用例対訳は，正確性評価を行っ

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

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 [13].

用例対訳を質問と回答の対でまとめたものである，応答用例対を用いることで，より正確な多言語間対話支援が可能であることが示されている [14]．しかし，応答用例対は多言語間対話支援システムの利用ログからの生成や，システムの管理者による登録が主なものとなっており，応答用例対数の増加が課題となっていた．そこで本稿では，用例対訳の一つであり，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の概念を活用して，応答用例対および用例の作成を自動で行う手法を提案する．本手法を利用することで，自動的な応答用例対の作成を可能とし，より正確な多言語間対話支援の実現を目指す．

2. 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]．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 [18]，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）[8] や，ケータイ多言語対話システム [9] 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 [10], [11].

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 [12]．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 [12]．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抱えている．また，より正確な多言語間対話支援が可能な応答用例対の提案や，多言語間対話支援システムの入力ログからの応答用例対の自動生成が提案されている [14]．しかし，文献 [14] では利用者がシステムへ入力した内容から応答用例対を作成しているため，十分な応答用例対を得るためには十分なシステム利用者数が必要であるという課題を抱えている．本稿では，応答用例対と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の概念を活用することで，自動的な応答用例対および用例の生成を目指す．

3. 応答用例対と穴あき用例を活用した多言語用例対訳作成手法

3.1 既存技術

3.1.1 応答用例対

本項では，応答用例対について述べる．本項で述べる応答用例対は文献 [14] に基づいており，質問と回答の用例対訳を 1 つずつ含んだものを応答用例対とする．現在，応答用例対はあらかじめシステム管理者が作成する方法と，システム利用者が使用した用例をもとに自動作成する方法が存在している．

応答用例対を用いない場合，文脈が異なる回答候補の提示が行われる場合が存在する．例として，質問「どのような症状がありますか？」に対して，一般的な文脈非依存の検索手法で「気持ちが悪いです」という文を検索した場合，「空気が悪いです」という回答候補をシステムが提示する場合が存在することが挙げられる [14]．また，用例対訳は多言語の対であるため，応答用例対を用いず，単純に単言語での質問と回答の対だけを見た場合に不正確となる場合が存在する．例として，「アレルギーを起こす食品は何ですか？」という質問に対して，「卵」という回答を用いたとする．しかし，この回答は中国語では「虫卵（昆虫の卵）」となる用例対訳を用いていたという事例が存在している [19].

文献 [14] では，多言語の質問と回答の対を適切に管理し，正確性の確保についても考慮されている．このため，用例対訳を用いた多言語間対話支援システムを利用する場合，文献 [14] の応答用例対の使用が必要になると考えられる．

しかし，文献 [14] では医療従事者や患者などの利用者の入力をもとに応答用例対を作成しており，入力数が少ない場合に応答用例対数が増えないという課題が存在している．本稿では，応答用例対を次項で述べる穴あき用例の概念を用いて自動生成する手法の提案を行う．

3.1.2 穴あき用例

本項では，穴あき用例について述べる．穴あき用例は，用例の一部を穴あきにして入れ替え可能としたものである．穴あき用例の穴あきの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となる．以降，用例の一部を穴あきにしたものを「穴あき用例」，穴あき部分に入れる単語を「穴埋め単語」とする．

穴あき用例の利用について検討した文献 [20] では，以下の手順で穴あき用例の抽出を行っている．本稿でも文

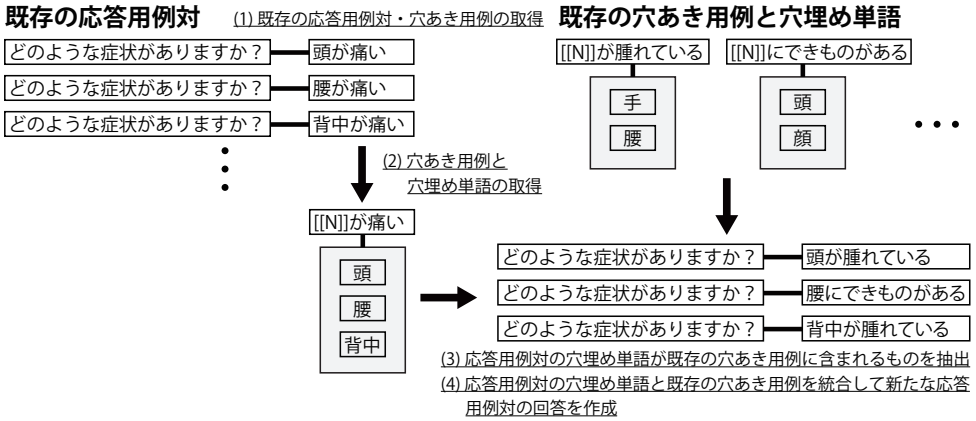

図 1　提案手法の流れ

献 [20] の手法を用いて穴あき用例の作成を行う．

Step 1　形態素解析器を用いて用例を形態素に分割

形態素解析器を用いて用例を単語に分割する．その際，句読点や “?” “!” は除去する．

Step 2　1 つの形態素のみが異なる用例の対を抽出

Step 1 で分割した形態素の 1 つのみが異なる用例の対を抽出する．その際，異なる形態素は同一品詞のもののみ抽出する．

Step 3　穴あき用例と穴埋め単語の抽出

Step 2 で抽出した，1 形態素のみ異なる用例群のうち，異なる単語群を「穴埋め単語」，それ以外の部分を「穴あき用例」として保存する．

なお，本稿では穴あき用例の穴あき部分を “[[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，穴あき用例としては「[[N]] が痛いです」と表記する．

文献 [20] では，多言語用例対訳共有システムで収集された用例対訳をもとに，自動的に穴あき用例対訳が作成可能かどうかを調査している．この結果，穴あき用例は言語間で一対一に対応しておらず，自動的な穴あき用例対訳の作成は困難であることが示されている．また，穴あき用例の概念を使用した不足用例の発見，対訳作成支援，類似文の作成補助の可能性が示されているものの，応答用例対の作成については課題点として残されていた．

前述の通り，多言語間対話支援システムで用例対訳を使用する場合，応答用例対を用いることで，より正確な多言語間対話支援が可能となる．そこで本稿では，文献 [14] の応答用例対と文献 [20] の穴あき用例を活用した多言語用例対訳作成手法について検討を行う．

3.2　提案手法

本節では，応答用例対と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．本手法では以下の手順で用例対訳の作成を行う．また，**図 1** に提案手法の流れを示す．

表 1　本実験で用いた既存の応答用例対の概要

質問文	応答用例対数
（内科）どのような症状がありますか？	164 対
（外科）どのような症状がありますか？	226 対

表 2　本実験で用いた既存の穴あき用例の概要

	文・単語数
穴あき用例数	206 文
穴埋め単語数	357 単語
穴あき用例のもととなった用例	792 文

図 1 中の数字は下記の項目と対応している．なお，本手法では穴埋め単語の品詞は名詞のみとした．

(1) 既存の応答用例対と穴あき用例を取得する．本稿では，(1) で抽出した穴あき用例を「既存の穴あき用例」とする．本稿では，医療分野の応答用例対および穴あき用例を使用することとする．

(2) (1) で取得した応答用例対のうち，回答の用例から穴あき用例と穴埋め単語を取得する．抽出には，前節で述べた穴あき用例抽出手法を用いる．

(3) (1) で取得した穴あき用例のうち，(2) で抽出した穴埋め単語を含むものを抽出する．

(4) (2) の穴埋め単語群を (3) の穴あき用例に入れることで，新たな用例および応答用例対を作成する．

上記の手順は文献 [20] で提案されている不足用例の発見手法と類似している．しかし，本手法では応答用例対内で使用された単語をもとに新たな用例を作成している．このことにより，質問に対応した用例対訳の作成が可能になると考えられる．

4.　試用実験

本節では，本稿で行った試用実験について述べる．本実験では，文献 [14] で作成された応答用例対と，文献 [20] で作成された穴あき用例を用いて新たな用例の作成を行う．

本稿で用いた応答用例対の概要についてを**表 1** に，穴あき用例の概要についてを**表 2** にそれぞれ示す．

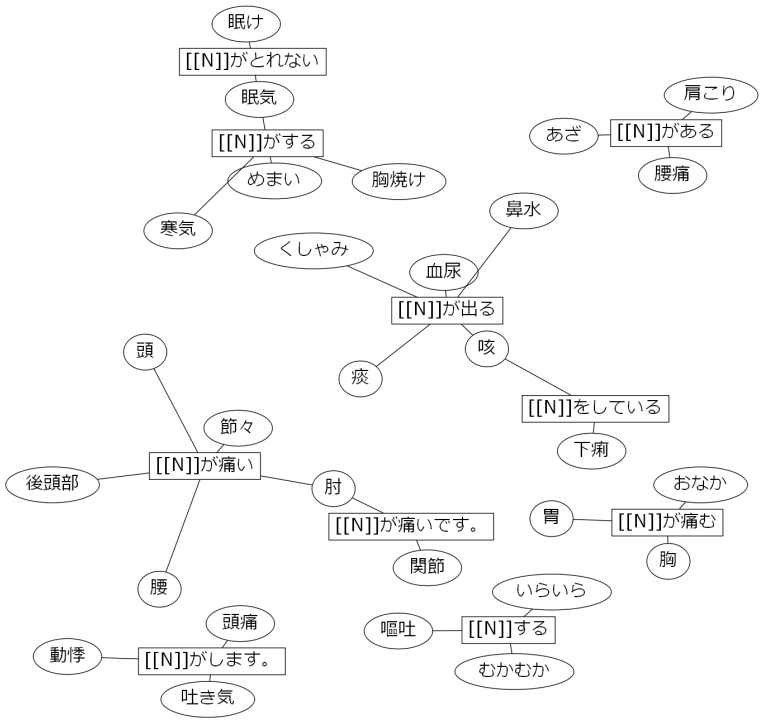

図 2　内科の応答用例対から作成された穴あき用例と穴埋め単語

表 1 の応答用例対は，(1)10 名の日本人被験者が質問文に対してそれぞれ 5〜10 文の回答文を作成，(2)4 人の日本人被験者が (1) の文を多言語用例対訳共有システム TackPad[12] および多言語問診票作成システム [10] の日本語用例 5,122 文から検索し，文献 [14] の定義に従って応答用例対を作成した結果を用いている．なお，文献 [14] では 5 件の質問文を用いていたが，表 1 以外の質問文から得られた回答が単語のものが多かったため，本稿では表 1 に示す 2 件の質問文のみ使用することとした．

表 2 の穴あき用例は，多言語用例対訳共有システム TackPad[12] の日本語用例 6,011 文から作成している [*1]．なお，本稿では穴埋め単語が名詞のもののみを使用している．また，名詞のみの文，複数の連続した名詞が含まれる文，助詞「の」の前後に穴あき部分がある文は，正確な文が作成されない可能性が高いため，使用しないこととした．

5. 実験結果と考察

5.1 応答用例対からの穴あき用例の作成

本節では応答用例対から作成された穴あき用例について考察する．これは，提案手法の（2）の段階にあたる．応答用例対の内科での質問「どのような症状がありますか？」の回答から作成された穴あき用例と穴埋め単語を**図 2** に，応答用例対の外科での質問「どのような症状がありますか？」の回答から作成されたものを**図 3** にそれぞれ示す．図 2 および図 3 内の四角が穴あき用例，丸が穴埋め単語，丸と四角を結んだ直線が穴あき用例と穴埋め単語の対応をそれぞれ示している．

図 2 および図 3 より，応答用例対の回答から穴あき用例を作成できていることが分かる．また，穴あき用例に使用された用例数（応答用例対の回答）は，内科に関しては 30 文，外科に関しては 42 文であった．これは，応答用例対の回答全体（表 1）に占める割合がそれぞれ 18.3%，18.5%となる．言い換えると，この割合は穴あき用例の作成効率であり，応答用例対の回答のうち，約 18%の用例が穴あき用例になったことを指す．既存の穴あき用例の作成効率は 13.2%（=792 文/6,011 文）であるため，応答用例対を用いた穴あき用例の作成は，用いなかった場合に比べて作成効率が良い傾向にあることが分かる．既存の穴あき用例は全ての用例対訳から作成しているため，穴あき用例に向かない用例も多く含まれている．このことが作成効率の差につながっていると考えられる．

また，図 2 および図 3 より，応答用例対の質問によって作成される穴あき用例および穴埋め単語が異なっていることが分かる．「[[N]] が痛いです」「[[N]] がある」など，一部共通している穴あき用例も存在しているものの，穴埋め単語群の傾向が異なっている．また，内科の穴あき用例・穴埋め単語は「[[くしゃみ/痰]] が出る」「[[動悸/吐き気]] がします」が，外科の穴あき用例・穴埋め単語は「[[釘/ガラス]] が刺さった」「[[傷/傷口]] がふさがらない」など，各診療科の特徴が穴あき用例や穴埋め単語に現れている．

応答用例対から作成された穴埋め単語は，図 2 および図 3 より，2 個から 10 個の単語群に分類されていること

[*1] 文献 [14] とは別の時期のデータを使用しているため，用例数が異なっている．

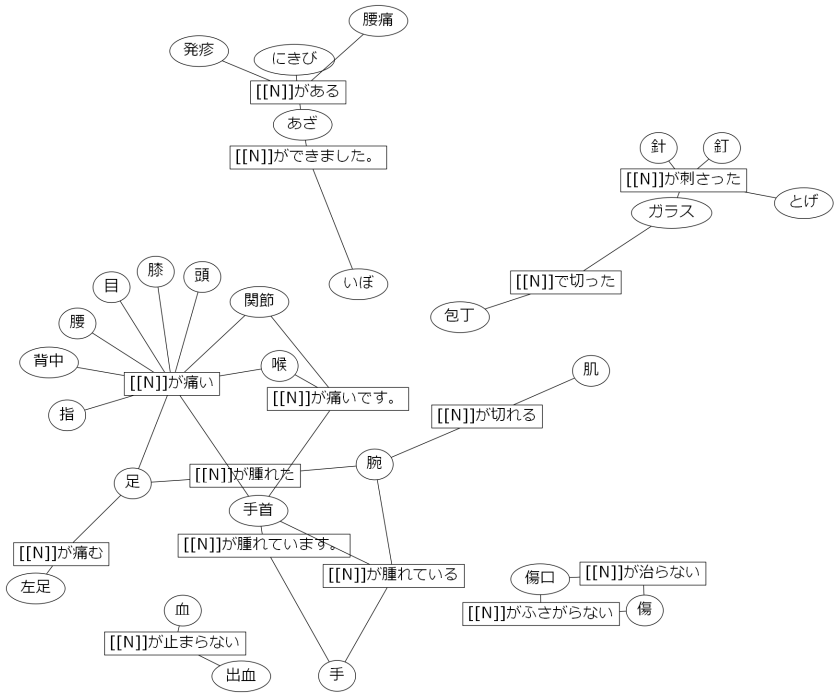

図 3　外科の応答用例対から作成された穴あき用例と穴埋め単語

が分かる．また，1 つの単語が複数のグループに属している場合も存在している．本節で抽出された単語群は，次節で分析する用例の作成で使用する．

5.2　応答用例対と穴あき用例からの用例の作成

本節では本手法で生成された応答用例対の回答（用例）について考察する．本手法では，内科に関する応答用例対が 202 文，外科に関する応答用例対が 727 文，それぞれ生成された．**表 3** に応答用例対と穴あき用例から生成された用例の例を示す．表 3 のうち，応答用例対の穴埋め単語と既存の穴あき用例を合わせたものが新しい応答用例対の回答である．なお，以下の手順は新たな用例作成を表 3-1 を例に挙げて説明したものである．

(1) 応答用例対の回答である「手が腫れている」「手首が腫れている」などから，「[[N]] が腫れている」という穴あき用例と「手」「手首」などの穴埋め単語を得る（表 3 中の「応答用例対」列）．

(2) 既存の穴あき用例から「[[N]] が乾燥します」「 [[N]] が動きません」など，(1) で取得した穴埋め単語が使用されている穴あき用例を取得する（表 3 中の「既存の穴あき用例」列）．

(3) (1) で取得した穴埋め単語と (2) で取得した穴あき用例を組み合わせて，「手首が乾燥します」「手が動きません」など，新たな用例および応答用例対を作成する．

表 3-1 は，正確な応答用例対が作成された例である．この例では，「手」「手首」「腕」という，よく似た部位が穴埋め単語群として使用されている．このことから，似たような部位を穴埋め単語として集めることができた場合，効率的に新たな用例の作成が可能になると考えられる．

表 3-2 は，一部の応答用例対が不正確となった例である．「[[N]] が痛くて眠れません」という穴あき用例は正確な用例として作成が可能である．しかし，「[[N]] がかゆいです」については，「頭がかゆいです」「腰がかゆいです」など，皮膚科での質問に適した文が生成されていることが分かる．これは，既存の穴あき用例は応答用例対ではないことに起因している．また，「頭」「腰」など，必ずしも内科に限定されない部位が穴埋め単語群として抽出されていることも一因であると考えられる．

表 3-3 では，10 個の単語が穴埋め単語群として抽出されている．部位も多岐にわたるため，「喉が熱を持っています」「背中にできものができました」など，適切な用例も多く生成できているものの，「目が熱を持っています」「頭に異物が入りました」など，病気の症状として不適切な文が生成される場合が存在している．特に，「[[N]] に異物が入りました」は，目など特定の部位でしか使用できない言い回しである．このことから，「[[N]] が痛い」など，汎用性の高い穴あき用例に対応する穴埋め単語群は，応答用例対として適切な用例生成が難しくなる可能性が考えられる．ただし，日本語として不適切な文については，Web 検索や，Web 上の単語をまとめた Google N-gram を活用して共起情報を取得し，この情報を元に適切性を判断することで，一定の正確性判定が可能になると考えられる．また，前述の通り，本手法での応答用例対の自動生成では不適切な応答用例対が生成される可能性があるため，多言語対話支援システムへの適用を行う際には，人手による正確性判定が必要になると考えられる．

表 3　応答用例対と穴あき用例から生成された用例の例

	診療科	応答用例対		既存の穴あき用例
		穴あき用例	穴埋め単語	穴あき用例
1	外科	[[N]] が腫れている	手，手首，腕	[[N]] が乾燥します
				[[N]] が動きません
				[[N]] がかゆいです
2	内科	[[N]] が痛い	頭，腰，肘，節々，後頭部	[[N]] が痛くて眠れません
				[[N]] がかゆいです
3	外科	[[N]] が痛い	頭，腰，喉，腕，手首，指，足，背中，目，膝	[[N]] が熱を持っています
				[[N]] にできものができました
				[[N]] に異物が入りました

・生成された用例は，応答用例対の穴埋め単語を既存の穴あき用例に入れたものである．

6. まとめと今後の課題

本稿では，応答用例対と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，既存の応答用例対に穴あき用例の概念を適用することで，新たな応答用例対の回答に当たる用例の作成を行った．

本稿の貢献は，応答用例対と穴あき用例を活用した用例対訳作成手法を提案し，応答用例対の回答となる用例作成ができることを示した点である．

今後の課題としては，効率的な穴埋め単語の抽出手法の検討が挙げられる．本稿では既存の応答用例対のみを利用していたため，穴埋め単語群の数が少なく，その結果，用例対訳作成の効率が悪くなっていると考えられる．応答用例対だけでなく，シソーラスなどを活用することで，適切な穴埋め単語群の抽出を行う．また，文献 [20] では，穴あき用例を活用した対訳作成手法についても述べられている．本稿にも適用可能であると考えられるため，応答用例対となっている用例の対訳作成に取り組む．

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２７年末現在における在留外国人数について（確定値），法務省（オンライン），入手先 ⟨http://www.moj.go.jp/nyuukokukanri/kouhou/nyuukokukanri04_00057.html⟩（参照 2016-05-10）．

[2] 法務省：平成２７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_00056.html⟩（参照 2016-05-10）．

[3] 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 2016-05-10）．

[4] 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).

[5] 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).

[6] Kim, K. J. and Bonk, C. J.: Cross-Cultural Comparisons of Online Collaboration, *Journal of Computer Mediated Communication*, Vol. 8, No. 1 (2002).

[7] 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).

[8] 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).

[9] 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).

[10] 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).

[11] 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).

[12] 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).

[13] 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).

[14] 福島　拓，吉野　孝：正確かつ自由度を高めた多言語間対話支援を目的とした応答用例対構築モデル，情報処理学会論文誌，Vol. 56, No. 1, pp. 219–227 (2015).

[15] Matsuda, M. and Kitamura, Y.: Development of Machine Translation System for Japanese Children, *Proceedings of 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] 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).

[19] 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答用例対作成システムの開発，情報処理学会，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).

[20] 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，Vol. 114, No. 461, pp. 23–28 (2015).