

## 多言語用例対訳共有システム TackPad における 用例評価手法の比較

福島 拓<sup>†1</sup> 吉野 孝<sup>†2</sup>

現在, 在日外国人数は年々増加しており, 多言語によるコミュニケーションの機会は増加している。コミュニケーションを行う際, 言語の違いは大きな障壁となる。また, 医療の分野では医療従事者と患者間での正確性の高い意思の疎通が重要である。そこで我々は, 多言語用例対訳を共有する TackPad の開発を行っている。用例対訳は高い正確性が必要とされるが, 収集した用例対訳の正確性は用例作成者に依存する。このため, TackPad の利用者相互で用例対訳の評価が可能な評価手法を複数提案し, その評価を行った。本稿では, 比較実験の結果から次の知見を得た。(1) 評価者の意図を汲み取るために, 評価の度合いが選択可能な評価手法を用意する必要がある。(2) 少数の評価で正確性の確保をするために, 評価軸を複数もつ評価手法を用意する必要がある。

## Comparison of Parallel-text Evaluation Methods in Multilingual Parallel-text Sharing System TackPad

TAKU FUKUSHIMA<sup>†1</sup> and TAKASHI YOSHINO<sup>†2</sup>

Recently, the number of foreign residents in Japan is increasing. Consequently, the opportunity of communication among people whose native language are different increases. When people communicate with each other, the difference of the language is a huge barrier. Moreover, it is important to communicate reliably between medical workers and foreign patients in the medical field. Therefore, we have developed a multilingual parallel-text sharing system, called TackPad. Parallel texts require verifying their correctness. The correctness of parallel texts depends on the creator of each parallel-text. Therefore, we proposed parallel-text evaluation methods to evaluate parallel texts among users of TackPad mutually, and evaluated them. We found the following findings from the comparison experiments. (1) It is necessary to provide the evaluation method that can be selected the degree of evaluation by an evaluator. This is because the method is necessary to take an evaluator's intention into consideration. (2) It is necessary to provide the evaluation method that has some evaluation axis. This is because the method is necessary to secure accuracy in small number of evaluation.

### 1. はじめに

現在, 在日外国人数や訪日外国人数は年々増加傾向にあり<sup>1),2)</sup>, 多言語によるコミュニケーションの機会も増加している。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<sup>3)</sup>。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<sup>4)-6)</sup>, 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。このため, 用例対訳や機械翻訳などの言語資源を組み合わせることで利用できる仕組みである言語グリッドの活動が広がるなど<sup>7),8)</sup>, 言語の壁を越える活動が活発化している。

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<sup>9)</sup>。

情報技術を利用した医療分野の支援として, 多言語医療受付支援システム  $M^3$  (エムキューブ)<sup>10)</sup> がある。 $M^3$  は, 正確な用例対訳を使用して医療受付での応対や問診の支援を行っている。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。

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, 試用実験で有用性を確認した<sup>11)</sup>。また, 用例対訳の正確性は用例作成者に依存しているため, 用例対訳作成者以外が用例対訳の評価を行うことによって正確性の確保を目指している<sup>12)</sup>。しかし, 文献 12) では評価手法に曖昧さを残していたため, 正確な用例対訳の評価が行えない可能性が考えられた。

そこで本稿では, TackPad に登録された用例を評価するための評価手法を複数提案し, 評価機能の構築を行う。また, 文献 12) の評価手法も含めた評価手法の比較実験を行い, 用例対訳の評価における評価手法の確立を目指す。

<sup>†1</sup> 和歌山大学大学院システム工学研究科

Graduate School of Systems Engineering, Wakayama University

<sup>†2</sup> 和歌山大学システム工学部

Faculty of Systems Engineering, Wakayama University

## 2. 関連研究

円滑な異文化間コミュニケーション支援を目指して、機械翻訳を用いた支援技術の研究が行われている<sup>13)</sup>。しかし、機械翻訳は正確性が必要な医療分野で利用可能な精度には達していない<sup>14)</sup>。

そこで現在、用例対訳による支援が行われている。用例対訳を利用したシステムとして、“日本語でケアナビ”<sup>\*1</sup>がある。“日本語でケアナビ”は、介護に関する日本語と英語の用例対訳を約 8000 文提供している。しかし、これらの用例対訳の作成には多くの時間がかかっている<sup>15)</sup>。

そこで我々は多くの人々の用例対訳に関する知識を Web 上で収集、共有、提供することを目的とした用例対訳共有システム TackPad の開発を行った<sup>11)</sup>。しかし、用例対訳の正確性の確保に必要な評価手法の確立はできていない。

Web システムの記事の評価機能の例として、Yahoo!ニュース<sup>\*2</sup>や newsing<sup>\*3</sup>などがある。これらのニュースサイトでは、利用者の評価によって注目されている記事をページの上に表示する取り組みが行われている。これらのニュースサイトにおける評価は、「おすすめ」「がっかり」の 2 値評価を提示したり、「びっくりした」「興味深い」などの評価項目を総合した点数を提示したりしている。しかし、評価は「記事の注目度」を閲覧者に提示する目的で行われており、記事そのものの正確性の評価は行われていない。

また、商品の評価機能の例として、Amazon<sup>\*4</sup>や価格.com<sup>\*5</sup>などがある。これらのサイトでは、販売している商品の評価や感想を、その商品を購入、所有している利用者が行っている。実際に商品を購入、所有している評価や感想は、その商品の購入を思案している別の利用者にとって非常に説得力のある情報であり<sup>16)</sup>、評価の有用性を判別する研究も行われている<sup>17)</sup>。

しかし、本システムで評価を行うべき対象は単語や文である。商品の評価は評価者が購入もしくは所有していないと評価することができないが、用例の評価は評価者が理解可能な言語の用例はすべて評価できる。このため、用例の評価は商品の評価と異なり、評価者が評価



図 1 TackPad の画面例  
Fig. 1 Screenshot of TackPad.

できる対象が非常に多くなる。現在、TackPad には全体で 6000 件弱、日本語で 2000 件弱の用例が既に登録されており、今後も増加することが想定されている。このため、用例の評価は商品の評価より評価者の負荷が高くなると考えられる。このため、商品の評価と用例の評価では必要とする要件が異なる可能性が考えられる。

そこで本研究では、用例対訳の正確性の確保が可能な評価手法の確立を目指す。用例の正確性の確保を行う評価手法に必要な要件は下記の項目である。

- (1) 評価者の負荷が少ない
- (2) 用例の正確性の確保に必要な情報が得られる

なお、用例対訳の評価には、用例そのものが適切かどうかを判断する「用例の評価」と、対訳が用例の翻訳として適切かどうかを判断する「用例間の評価」が必要である。本稿では、「用例の評価」に関する手法の確立を目指す。

なお、「用例の評価」では下記の評価を行う必要があると考えられる。

- (1) 登録された用例は実際に用例を使用する現場で使用されるかどうかの評価
- (2) 用例の表現や言葉遣いは適切かどうかの評価

本稿では、これらの項目を判断可能な評価手法の確立を目指す。

## 3. 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。

### 3.1 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。

\*1 <http://nihongodecarenavi.jp/>  
\*2 <http://headlines.yahoo.co.jp/hl>  
\*3 <http://newsing.jp/>  
\*4 <http://www.amazon.co.jp/>  
\*5 <http://kakaku.com/>

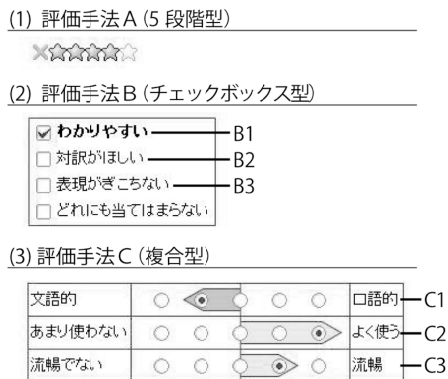


図 2 3つの評価手法の画面例

Fig.2 Screenshots of three evaluation methods.

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上での用例対訳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。また、本機能は翻訳作業が不要なため、理解できる言語が1言語の利用者も本システムが目指している用例対訳の収集、共有に貢献することが可能である。

#### (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。

### 3.2 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。各評価手法の画面例を図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。各評価手法は評価者の負荷を減らすために、1回のクリックで評価を可能としている。

#### (1) 評価手法A(5段階型)(図2-(1))

文献11)で既の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(チェックボックス型)(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。チェックボックスを使用することで、複数の項目にチェックすることも可能とした。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では複数の評価軸による評価を可能とした。今回は、「わかりやすい」「対訳がほしい」「表現がごこちない」の3つの項目を用意した。なお、本稿ではそれぞれB1、B2、B3とする。図2-(2)では4つ目の項目として「どれにも当てはまらない」があるが、これは実験用に用意した項目で、実際のシステムでは使用しない。

#### (3) 評価手法C(複合型)(図2-(3))

予め用意した対義語の組に対して、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。また、評価手法Aと同様の5段階評価だが、対義語による評価軸を提示している点が評価手法Aとは異なる。今回は、「文語的-口語的」「あまり使わない-よく使う」「流暢でない-流暢」の3つの対義語を用意した。なお、本稿ではそれぞれC1、C2、C3とする。

評価手法Aは評価軸が一つのため、上記の評価手法の中では一番評価者の負荷が少ないと考えられる。しかし、評価手法B、評価手法Cのように複数の評価項目を用意していないため、用例対訳の正確性の確保に必要な情報が得られない可能性がある。この点については、次章の実験で検証を行う。

### 4. 比較実験

本章では、3.2節で述べた評価手法の比較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正確性の確保が可能な評価手法についての知見を得ることである。

また、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。

表 1 評価者別の評価手法と用例群の関係 (一例)

Table 1 Example allocated relation among the evaluation methods, parallel texts groups and evaluators.

	評価手法 A	評価手法 B	評価手法 C
評価者 $\alpha$	用例群 1	用例群 3	用例群 2
評価者 $\beta$	用例群 2	用例群 1	用例群 3
評価者 $\gamma$	用例群 3	用例群 1	用例群 2

・表中の用例群 1, 用例群 2, 用例群 3 は, それぞれ用例 10 文を 1 組とした用例群を指す。

- (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
- (3) 被験者 (患者) が使用する用例 (医療従事者が使用する用例は用いない)

このようにして用意した用例は, 10 文を 1 組として 3 つの用例群 (用例群 1, 用例群 2, 用例群 3) に分けた。被験者は 1 つの評価手法に対し, 用例群のうち一つ (用例群 1, 用例群 2, 用例群 3 のいずれか) の 10 文の用例を評価した。被験者別の評価手法と用例群の関係を表 1 に示す。表 1 のように, 実験では同一の用例を複数回評価を行わないようにしている。実験全体では被験者は計 30 文の用例の評価を行った。また, すべての用例に対して必ず評価を付けるように依頼した。評価手法 B については一つの用例に対して一つ以上の項目にチェックを付けるように, 評価手法 C については 3 つの評価軸すべてを評価するように依頼した。評価手法 B の評価項目がどれも当てはまらない場合, 「どれにも当てはまらない」にチェックを付けるように依頼した。なお, 3 つの評価手法の評価順序は被験者ごとに変えている。また, 実験後にアンケートへの回答を依頼した。

被験者は, 本システムの利用者である和歌山大学システム工学部所属の大学生・大学院生 33 名 (男性 17 名, 女性 16 名) である。なお, 今回の被験者は患者の立場で用例の評価を行ってもらった。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ため, 今回の実験結果は限定的であると考えられる。

## 5. 実験結果と考察

### 5.1 被験者による評価

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 2 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 3 にそれぞれ示す。なお, 表 2 中の平均は値が大きいものが, 表 3 中の合計は値が小さいものが, それぞれ支持されていることを示す。

表 2 評価の行いやすさに関するアンケート結果 (5 段階評価)

Table 2 Questionnaire result of easiness of evaluation (5-point Likert scale).

	評価手法 A	評価手法 B	評価手法 C
平均	3.73	3.09	3.52
標準偏差	0.88	1.07	0.94

- ・質問: 「この評価のやり方は, 文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 3 評価の行いやすさに関するアンケート結果 (順位付け)

Table 3 Questionnaire result of easiness of evaluation (ranking).

順位	評価手法 A	評価手法 B	評価手法 C
1	15	7	11
2	9	8	16
3	9	18	6
合計	60	77	61

- ・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。

表 2, 表 3 より, 評価手法 A (5 段階型) と評価手法 C (複合型) が, 評価手法 B (チェックボックス型) より比較的评价を行やすい評価手法であったことが分かる。ただし, 表 2 の結果を用いてフリードマン検定を行った結果, 有意確率は 0.066 となり有意差は見られなかった。

アンケートの自由記述欄から得られた評価手法 A に関する好意的な意見として, 「入力の負荷が少ない」「単純な操作で, 率直な評価ができた」などが得られた。しかし, 「自分の評価基準を決めるためやりにくかった」「基準があいまいな分, 信憑性に欠ける」など評価基準を指定していないことによる曖昧さを問題視する意見も見られた。

アンケートの自由記述欄から得られた評価手法 C に関する好意的な意見として, 「詳細に評価できるので微妙なニュアンスが伝わりやすい」「軸や基準が決まられているので評価しやすかった」などが得られた。しかし, 「話し言葉と書き言葉の判別はむずかしい」, また「流暢かどうかの判定がわかりにくかった」という意見も得られたため, 評価に使用する対義語を再考する必要があると考えられる。

評価手法 B に関しては, 「設問のまま素直に回答できる」という好意的な意見も見られたが, 「当てはまるか当てはまらないかだけで, 微妙な判断ができない」という評価の度合いを選択できないことを問題視する意見が得られた。このため, 評価者の意図を汲み取るため

表 4 各評価手法の評価人数

Table 4 Number of evaluators in each evaluation methods.

	用例群 1	用例群 2	用例群 3
評価手法 A	11 人	11 人	11 人
評価手法 B	10 人	12 人	11 人
評価手法 C	12 人	10 人	11 人

・表中の用例群 1, 用例群 2, 用例群 3 は, それぞれ用例 10 文を 1 組とした用例群を指す。

表 5 各評価手法間の相関係数

Table 5 Correlation coefficient between evaluation methods.

	A	B1	B2	B3	C1	C2	C3
A	1.00						
B1	-0.11	1.00					
B2	-0.21	-0.56	1.00				
B3	-0.05	-0.86	0.40	1.00			
C1	-0.21	<b>0.52</b>	-0.19	-0.61	1.00		
C2	-0.32	<b>0.69</b>	-0.30	-0.57	<b>0.61</b>	1.00	
C3	-0.19	<b>0.69</b>	-0.38	-0.69	<b>0.64</b>	<b>0.62</b>	1.00

- ・表中の A, B, C はそれぞれ評価手法 A(5 段階型), 評価手法 B(チェックボックス型), 評価手法 C(複合型) を指す。
- ・B1 は「わかりやすい」, B2 は「対訳がほしい」, B3 は「表現がごちない」が評価軸である。  
「どれにも当てはまらない」は省略している。
- ・C1 は「書き言葉-話し言葉」, C2 は「あまり使わない-よく使う」, C3 は「流暢でない-流暢」が評価軸である。
- ・相関係数が 0.5 以上のものを太字に, 相関係数が-0.5 以下のものを斜体にしてある。

に, 2 値の評価ではなく評価の度合いが選択可能な評価手法を用意する必要があると考えられる。

また, 「もう少し選択肢を増やしたほうがいい」と言う意見が多く見られた。しかし, 評価項目を増やすと評価するために必要な時間が増え, 評価者の負荷も増えるため, 評価項目を増やすことは難しいと考えられる。

## 5.2 評価手法の相関

本節では, 各評価手法によって評価された用例の評価結果について考察する。

同一の評価手法と用例群で評価した人数を表 4 に示す。評価手法と用例群の組み合わせは 9 つあるが, 表 4 からすべての組み合わせで 10 人以上が評価していることが分かる。なお, 表 4 の人数が異なっているのは一部の被験者の評価ログ取得に失敗したためである。

各評価手法間の評価の相関を表 5 に示す。表 5 は, 用例ごとに各評価手法による評価結果の平均を出し, その平均を評価手法間で相関関係があるかどうか確認した結果である。な

表 6 評価手法 A と評価手法 C2 の評価が大きく異なった用例とその評価値

Table 6 Parallel texts and their evaluation values with greatly different evaluation in evaluation method A and evaluation method C2.

用例	A	C2	A-C2
1 治療費はどれくらいかかりますか?	2.82	4.67	-1.85
2 保険証をわすれました。	3.45	4.73	-1.27
3 寒くなるかと膝がしくしく痛みます	4.00	1.58	2.42
4 障害年金はうけられますか?	4.55	2.00	2.55

- ・評価手法 C2 の評価軸は「あまり使わない-よく使う」である。
- ・“A” “C2” の列は, それぞれ評価手法 A と評価手法 C2 の平均である。
- ・“A-C2” の列は, 評価手法 A の平均から評価手法 C2 の平均を引いたものである。

お, 評価手法 A は 5 段階評価 (1~5) の平均, 評価手法 B は (チェックされた数/評価者数), 評価手法 C は左端の値を 1, 右端の値を 5 に変換した 1~5 の値の平均を用いている。

表 5 より, 評価手法 A は他の評価手法と相関関係がないことが分かる。また, 評価手法 C の各評価軸間と, 評価手法 B1(わかりやすい) と評価手法 C の各評価軸との間には正の相関関係があることが分かる。

評価手法 A は評価者それぞれが自分で評価軸を決めることになるため, 評価者の違いによるばらつきが大きくなっている可能性が考えられる。また, 評価手法 B1 と評価手法 C2(あまり使わない-よく使う), 評価手法 C3(流暢でない-流暢) のように同傾向の評価軸は同様の評価が行われている。さらに, 評価手法 C1(書き言葉-話し言葉) と他の評価軸との相関関係から, 口語的な用例はわかりやすく, よく使う用例であると評価される傾向が見られた。

評価手法 A の平均と評価手法 C2 の平均が大きく異なった用例を表 6 に示す。表 6 の 1 と 2 は, 評価手法 C で「よく使う」と評価されたが評価手法 A の平均が低かったもの, 表 6 の 3 と 4 は, 評価手法 C で「あまり使わない」と評価されたが評価手法 A の平均が高かったものの例である。これは, 評価者によって評価軸が異なることが原因であると考えられる。用例の評価では前述の通り「実際に使用するか」「表現や言葉遣いは適切か」の判断を行う必要があるが, これらの判断が可能な評価手法 C と相関関係がない評価手法 A は, 用例の評価手法として適切でないと考えられる。このため, 評価結果を用例の正確性の確保に使用するためには, 評価手法 A のように曖昧になる可能性を含む評価手法ではなく, 評価手法 C のように評価軸を適切に用意した評価手法を用いる必要があると考えられる。

なお, 今回の実験では評価者が 10 名程度と比較的少なかったため, 評価手法間の差が大きく出た可能性がある。しかし, 評価手法の適用先である用例の評価は多数の用例を評価する必要があり, 多くの評価が一つの用例に集まる可能性は低い。このため, 少数の評価で用

例の正確性の確保を行うためには評価手法Cのように評価軸を複数用意する必要があると  
考えられる。

## 6. おわりに

本稿は、用例の正確性の確保を行う機能として、3つの評価手法をシステムに導入し、比  
較実験により用例の正確性の確保が可能な評価手法の確認を行った。

本実験の結果として、以下の知見を得た。

- (1) 評価者の意図を汲み取るために、評価の度合いが選択可能な評価手法を用意する必  
要がある
- (2) 少数の評価で正確性の確保をするために、評価軸を複数もつ評価手法を用意する必  
要がある

今後、医療従事者による評価においても同様の結果が得られるか確認し、用例の評価手法  
の確立を目指す。また、用例間の評価手法についても検討を行う。

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：平成20年末現在における外国人登録者統計について、  
<http://www.moj.go.jp/PRESS/090710-1/090710-1.html>
- 2) 法務省：平成20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 (確定版)、  
<http://www.moj.go.jp/PRESS/090326-1/030326-1.html>
- 3) 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)。
- 4) Takano, Y. and Noda, A.: A temporary decline of thinking ability during foreign  
language processing, *Journal of Cross-Cultural Psychology*, 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,  
pp.1-13(1994).
- 6) Kyong-Jee Kim and Curtis J. Bonk: Cross-Cultural Comparisons of Online Col-  
laboration, *Journal of Computer Mediated Communication*, Vol.8, No.1(2002).
- 7) Toru Ishida: Language Grid: An Infrastructure for Intercultural Collaboration,  
*IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp.96-  
100(2006).
- 8) Satoshi Sakai, Masaki Gotou, Masahiro Tanaka, Rieko Inaba, Yohei Murakami,

- Takashi Yoshino, Yoshihiko Hayashi, Yasuhiko Kitamura, Yumiko Mori, Toshiyuki  
Takasaki, Yoshie Naya, Aguri Shigeno, Shigeo Matsubara and Toru Ishida: Lan-  
guage Grid Association: Action Research on Supporting the Multicultural Society,  
*International Conference on Informatics Education and Research for Knowledge-  
Circulating Society (ICKS-08)*(2008).
- 9) 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)。
  - 10) 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)*.
  - 11) 福島拓, 宮部真衣, 吉野孝, 重野亜久里:医療分野を対象とした多言語用例対訳収集 Web  
システム TackPad の開発, *マルチメディア, 分散, 協調とモバイル (DICOMO2008)*  
シンポジウム, pp.1030-1036 (2008)。
  - 12) 福島拓, 吉野孝, 重野亜久里：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,  
2009-GN-70(21), pp.121-126(2009)*.
  - 13) Takashi Yoshino, Kunikazu Fujii and Tomohiro Shigenobu: Availability of Web In-  
formation for Intercultural Communication, *Pacific Rim International Conference  
on Artificial Intelligence (PRICAI-08)*, pp.923-932(2008).
  - 14) 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)*.
  - 15) 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, *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)*.
  - 16) 渡辺弘美：ウェブを変える10の破壊的トレンド, *ソフトバンククリエイティブ (2007)*.
  - 17) 山澤美由起, 吉村宏樹, 増市博：Amazon レビュー文の有用性判別実験, *情報処理学  
会研究報告, 自然言語処理研究会, 2006-NL-173-(3), pp.15-20(2006)*.