G-22

機械翻訳適応文を用いた単言語話者による機械翻訳文の評価と訂正

Evaluation and Correction of Machine Translation by Monolingual Speaker using Machine Translation Adaptation Sentences

山本　里美 †　　福島　拓 ‡　　吉野　孝 †
Satomi Yamamoto　Taku Fukushima　Takashi Yoshino

1. はじめに

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は増加している．また，インターネットを用いた多言語間コミュニケーション支援の方法として，機械翻訳や多言語用例対訳などがある．多言語用例対訳とは，正確性の確保された多言語の対のことで，正確な情報の共有が重要となる医療分野などで用いられる．しかし，正確な情報の共有のために必要とされる用例対訳の数は多く，十分な数の用例対訳を多言語において収集することは困難である [1].

我々は，現在，クラウドソーシングと機械翻訳を用い，単言語話者であっても用例対訳作成が行える手法に関する研究を行っている．これまでの実験により，クラウドソーシング上の単言語話者に，会話文形式で機械翻訳文を提示し，その機械翻訳文の評価と訂正を依頼することで，一部の用例について用例対訳や応答用例対[*1]が作成できることがわかった [2]．しかし，文献 [2] の実験において作成された訂正文の約 7 割は，専門家によって不正確な対訳だと判断された．この原因の 1 つとして，不適切な機械翻訳が行われた場合に，クラウドソーシングの作業者が機械翻訳文の意味を推測できず，適切な訂正が行えないことがあると考えられる．

そこで，我々は，従来の用例対訳作成手法に，単言語話者が機械翻訳文の意味を推測しやすくなるような機械翻訳文を作成するための「機械翻訳適応文」を用いることを提案し，適応文作成実験を行った [3]．機械翻訳適応文は，折り返し翻訳文との一致度が高い文のことであり，実験の結果，すべての用例に対して，元の用例を使用する場合以上の折り返し翻訳文との一致度を示す機械翻訳適応文を作成できることがわかった．

本稿では，従来の用例対訳作成手法に機械翻訳適応文を用いて対訳作成実験を行い，機械翻訳適応文の効果の検証を行う．

2. 関連研究

現在，クラウドソーシング用いて多言語データを収集する研究が多く行われている．Omar らによる，翻訳対象文やクラウドソーシング作業者の特徴をもとに翻訳文を分析し，品質の良い翻訳文を作成する研究 [4] や，複数の属性の作業者を組み合わせによって精度の高い翻訳文を作成する研究 [5] などがある．これらの研究により，クラウドソーシングを用いることで，専門家による翻訳の質に近い翻訳文を作成できることが分かったが，これらの研究では，クラウドソーシングの作業者が 2 言語以上を理解できることを前提としている．しかし，クラウドソーシング上の多言語話者の数は少なく，特に，話者の少ない言語を理解できる多言語話者をクラウドソーシング上で確保することは困難である．

クラウドソーシング上の単言語話者を対象にした多言語データ収集の研究に，児童向け図書を対象にした，Chan らによる “Crowdsourced Monolingual Translation”がある [6]．これは，翻訳対象文の言語を母語とする作業者と，翻訳対象言語を母語とする作業者が，文への注釈付けや，機械翻訳文の修正を行うことで翻訳を行う手法である．複数の単言語話者が，繰り返し注釈付けと機械翻訳文の修正を行うことで，1 人の多言語話者による修正と同程度の質の翻訳文を取得することができる．しかし，この手法では，翻訳対象文と同じ意味を翻訳文が示すかどうかの評価をしておらず，正確性の確保が必要とされる用例対訳作成に用いることは困難である．

そこで我々は，クラウドソーシング上の単言語話者と機械翻訳を用いることで，用例対訳候補文の作成，正確性評価を行う手法について研究を行っている．

3. クラウドソーシングを用いた用例対訳作成手法

本章では，現在我々が提案している，クラウドソーシング上の単言語話者を用いた用例対訳作成手法について述べる．図 1 に，用例対訳作成の流れを示す．

本手法は，図 1 のように，3 つの Step で構成されている．各 Step では，それぞれ異なるクラウドソーシングのタスクを用いることで，翻訳や正確性評価の処理を行う．

図 1 の Step 1 の機械翻訳適応文作成では，翻訳対象文を機械翻訳に適した文に書き直すことで，図 1 の Step 2 の機械翻訳文の評価と訂正において，作業者が原文の意味を推測しやすい機械翻訳文を作成する．機械翻訳適応文の作成には，折り返し翻訳を用い，折り返し翻訳文との一致度を作業者に提示することで作業の支援を行っている．なお，折り返し翻訳文との一致度の評価には，機械翻訳の評価指標として用いられる RIBES（ライビーズ）[7] を用いている．

図 1 の Step 2 の機械翻訳文の評価と訂正では，翻訳対象言語のを母語とする作業者に，その言語に翻訳した機械翻訳文の評価と訂正を依頼することで，用例対訳候補文となる文を取得する．なお，本手法では，疑問文とその回答，それに続く 1 文の，3 文からなる会話文を使用する．これは，会話文を提示することで，図 1 の Step 2 の機械翻訳文の訂正を行う作業者は，文脈から翻訳前の用例の意図を推測しやすくなり，機械翻訳文の訂正精度が向上すると考えたためである．また，機械翻訳文の評価は，文献 [8] の評価基

† 和歌山大学, Wakayama University
‡ 大阪工業大学, Osaka Institute of Technology

[*1] 質問とその回答の対，またその類似文からなる用例対訳のことであり，会話の支援に用いられる．

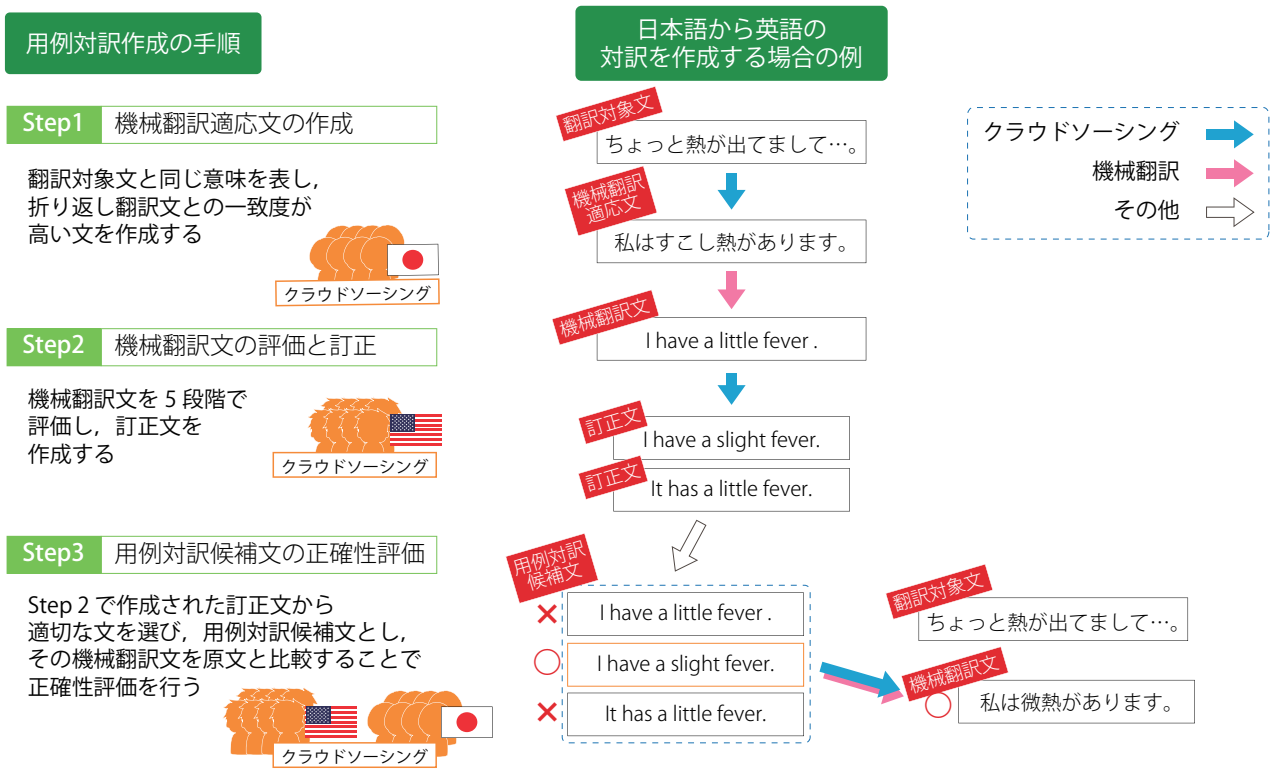

図 1: クラウドソーシング上の単言語話者を用いた用例対訳作成手法の流れ

表 1: 実験用タスクの設定

項目	設定
評価する機械翻訳文数	76 文
1 タスクあたりの設問数	1 問
1 文あたりの評価回数	10 回
総設問数（総タスク数）	760 問
1 タスクあたりの報酬	$0.02
作業者の所在地	United States

準[*1] を参考にしている．

また，図 1 の Step 3 の用例対訳候補文の評価では，翻訳後の言語を母語とする作業者と原文の言語を母語とする作業者にそれぞれタスクを依頼し，評価を行う．翻訳後の言語を母語とする作業者によって，文としての正しさを評価し，原文の言語を母語とする作業者によって翻訳後の文の意味が原文と同じかどうかの評価を行う．

4. 機械翻訳適応文を用いた対訳作成実験

本章では，用例対訳作成手法における機械翻訳適応文の効果を検証するための実験方法について述べる．本実験では，機械翻訳文の評価と訂正において，原文をそのまま用いた場合と，機械翻訳適応文を用いた場合の結果を比較する．これは，図 1 の Step 1 を行わなかった場合と，行った場合での比較である．

4.1 使用したデータセット

本実験では，文献 [3] の実験で用いた会話文と，その会話文を用いて作成した機械翻訳適応文を使用した．文献 [3] では，38 組の会話文の 2 文目を翻訳対象文として，機械翻訳適応文作成を行い，各文に対して 10 文の機械翻訳適応文を取得した．本実験では，文献 [3] で取得した各 10 文の中から，最も機械翻訳に適した 1 文を用いて用例対訳作成を行う．そのため，実験で用例対訳作成を行うのは，原文の 38 文と，その原文に対する機械翻訳適応文 38 文の，計 76 文である・

なお，文献 [3] で作成された機械翻訳適応文が機械翻訳に適しているかどうかの評価には，クラウドソーシングによる評価と，折り返し翻訳文との RIBES の値（折り返し翻訳文との一致度）を用いた．文献 [3] で作成された機械翻訳適応文の中には，折り返し翻訳文との RIBES の値は高いが，原文と大きく意味が異なるものも含まれていた．そのため，単純に折り返し翻訳文との RIBES の値のみで評価することが困難だと考え，クラウドソーシングを用いて，原文との意味の近さの評価を行った．

クラウドソーシングでは，原文と比較して意味が異なる文を選択するように指示した．これにより，10 文の中から，原文と大きく意味が異なる文を除外できる．さらに，原文と意味が異なる文が除外された機械翻訳適応文の中から，最も折り返し翻訳文との RIBES の値が高い文を取得する．これにより，原文に対して，最も適切な機械翻訳適応文の抽出を行う．

[*1] 評価段階は，1 : Incomprehensible(理解できない)，2 : Disfluent English(流暢でない)，3 : Non-native English(非母語言語)，4 : Good English(良い英語)，5 : Flawless English(完璧な英語)

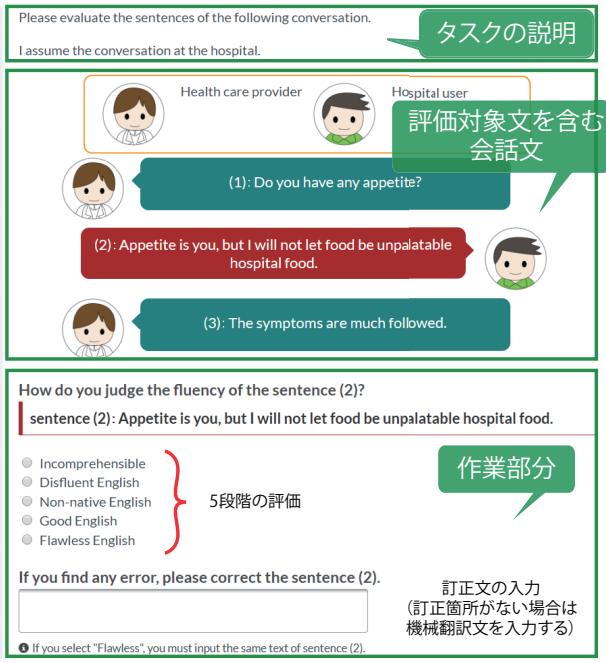

図 2: 機械翻訳文の評価と修正タスクの画面例

4.2 タスクの設計

本実験では，クラウドソーシングサービスとして CrowdFlower を用い，所在地を United States とする作業者に対して，機械翻訳文の評価と訂正タスクを依頼する．表 1 に，タスクの設定を示す．4.1 節で述べたデータを用いて実験を行うため，評価する機械翻訳文の数は，原文を機械翻訳した 38 文と，原文から作成した機械翻訳適応文 38 文を機械翻訳したものの，計 76 文となる．

また，作業者に提示するタスク画面を図 2 に示す．作業者には機械翻訳文を含む会話文や，その文の利用者の属性情報（医療従事者または病院利用者）を提示する．作業者は，提示された会話文の 2 文目について，3. 章で述べた基準によって 5 段階評価を行い，評価対象文に訂正が必要な場合には訂正文の作成を行う．なお，本稿では，今後，原文を用いた用例対訳作成を Task 1 （原），図 1 の Step 1 を行い，機械翻訳適応文を用いた用例対訳作成を Task 2 （機）とする．

5. 実験結果と考察

Task 1 （原）と Task 2 （機）について，取得した機械翻訳文の評価値や訂正文を比較する．5.1 節で取得した訂正文の評価結果について述べる．また，5.2 節で，機械翻訳適応文の効果について述べ，5.3 節と 5.4 節で，今後の課題について述べる．

5.1 訂正文の評価

取得した訂正文について，著者の 1 人が評価を行った．評価は，以下の軸を設定し，作業者によって作成された訂正文が適切であるかを判断した．

条件 1 機械翻訳文の訂正文として適切である．

条件 2 原文の意味を大きく変更していない．

表 2: 各タスクにおける評価結果

項目	Task 1 （原）	Task 2 （機）
取得した訂正文の数	380 文	380 文
適切な訂正文の数	216 文	254 文
機械翻訳文の評価値（平均）	3.57	3.67

表 3: Task 2 （機）の結果を Task 1 （原）と比較

Task 1 （原）と比較する項目	文の数
評価値が向上した	14 文
評価値が低下した	12 文
適切な訂正文数が増えた	23 文
適切な訂正文数が減った	7 文

・各文の数は実験で使用した 38 文中で該当する文の数である．

評価結果を，表 2 に示し，Task 1 （原）と Task 2 （機）で比較した結果を表 3 に示す．評価の結果，Task 1 （原）で取得できた適切な文は 216 文，Task 2 （機）で取得できた適切な文は 254 文で，原文を用いる場合よりも 18%増加している．なお，文ごとに訂正文数の変化をみると，取得した適切な訂正文数が原文を用いる場合よりも増加したのは 38 文中 23 文，変化しなかったのが 38 文中 8 文だった．このことから，原文の 8 割は，機械翻訳適応文を用いた方が適切な訂正文が多く作成された．

また，Task 2 （機）では，5.3 節や，5.4 節で述べる 3 文以外の 35 文すべてにおいて，適切な訂正文を 1 文以上取得することができた．Task 1 （原）も同様に，適切な訂正文を 1 文以上取得することができたのは 35 文だった．

5.2 機械翻訳適応文の効果

機械翻訳適応文を用いることで，機械翻訳が適切に行われる可能性が高くなる．そのため，Task 1 （原）よりも Task 2 （機）の方が機械翻訳文の評価値が高くなる傾向にあると考えた．しかし，Task 1 （原）の方が機械翻訳文の評価値（中央値）が高かった文は 38 文中 14 文，低かった文は 38 文中 12 文だった．また，Task 1 （原）における評価値の平均は 3.57，Task 2 （機）における評価値の平均は 3.67 だった．38 文の原文それぞれにおいて，機械翻訳適応文を用いた場合との機械翻訳文の評価値では違いが見られたが，全体の平均に差はなく，機械翻訳文の評価値に，機械翻訳に使用する文による差は見られなかった．このことから，機械翻訳が適切に行える可能性が高い機械翻訳適応文であっても，機械翻訳文の評価が必ずしも向上するわけではないことがわかった．

評価値が向上した文の例を表 4 に示す．なお，文を区別するため，翻訳対象の 38 文それぞれに ID を付与した．文 ID 1 に示す文は，会話文では “水虫の感染源はどこですか？” に続く文である．どちらも英語の文としては正しいが，原文の場合は，会話文中の表現として不自然なため，低い評価値になっていると考えられる．しかし，すべての作業者

表 4: Task 2 （機）で機械翻訳文の評価値が向上した文の例

文 ID	項目	Task 1 （原文を使用）	Task 2 （機械翻訳適応文を使用）
1	日本語	ペットからだと考えられます。	これは、ペットからだと考えられます。
	機械翻訳文	It believed it from the pet.	This is most likely from the pet.
	折り返し翻訳文との RIBES の値	0.74	0.97
	機械翻訳文の評価値	3	5
2	日本語	仕事中に発作が起きてしまわないかがとても心配です。	私は、仕事中に発作が起きてしまうのがとても心配です。
	機械翻訳文	It is very concerned about attacks or not Shimawa happening at work.	I am very worried that the attack would have occurred during the work.
	折り返し翻訳文との RIBES の値	0.67	0.81
	機械翻訳文の評価値	2	4

が訂正文として機械翻訳文をそのまま入力していた．そのため，5.1 節の条件 2 より，Task 1 （原）では，適切な訂正文を取得できないという結果になった．なお，Task 2 （機）においても，Task 1 （原）と同様に，ほとんどの作業者が訂正文として，機械翻訳文を使用していた．しかし，実験で使用したタスクでは，作業者に，評価値が 5(Flawless English) の場合は，訂正文として機械翻訳文を入力するように指示している．文 ID 1 における Task 2 （機）の機械翻訳文の評価値は 5 であり，適切な翻訳が行われているため，作業者は機械翻訳文を訂正文として用いたと考えられる．文 ID 1 に示す結果より，機械翻訳適応文作成において，作業者がより会話文中で自然な文になるような文の作成を行うことによって，評価値が高い機械翻訳文を取得できる可能性がある．

また，文 ID 2 の原文の機械翻訳文のように，不正確な機械翻訳が行われる場合がある．機械翻訳適応文作成の主な目的は，このような不正確な機械翻訳を防ぐことである．不正確な機械翻訳が行われるような原文と，機械翻訳適応文での機械翻訳文の評価結果の比較より，機械翻訳適応文を用いることで，機械翻訳文の質が向上し，翻訳後の言語を母語とする作業者からの評価が高い機械翻訳文を取得できる可能性があると考えられる．

なお，機械翻訳文の評価値が Task 1 （原）よりも Task 2 （機）の方が低い文は 12 文あった．しかし，その 12 文中 6 文において，Task 2 （機）の方が Task 1 （原）よりも多く，適切な訂正文を取得している．例えば，機械翻訳文が “Body state of.” や “There is a thing to wear the slippers of others to the source of infection of athlete’s foot” では，作業者による評価値は 1 や 2.5 といった結果で，原文を用いる場合より評価値が低かったが，正確な訂正文が多く作成されていた．機械翻訳の評価値は Task 1 （原）と比較して低いが，適切な訂正文は多く作成された例を表 5 に示す．表 5 に示す文は，会話文において，“検査の結果から何が分かりますか。”に続く文である．そのため，原文の “体が今どういう状態かわかります。 ”が，機械翻訳適応文では “体の状態。”という表現に変更されていても，会話での応答として適切なため，“State of your body.”といって訂正文も適切なものとして評価した．

これらのことから，機械翻訳適応文を用いた場合，機械翻訳文の評価値が低いような，不正確な機械翻訳が行われている場合でも，クラウドソーシングの作業者が意味を推測しやすい機械翻訳文を作成できる可能性があると考えられる．

5.3 両方のタスクで適切な訂正文が取得できない例

なお，Task 1 （原）と Task 2 （機）の両方で適切な訂正文が取得できなかった文は 38 文中 2 文あった．その 2 文の原文は “仰向けです。”と “革靴などは避け、できるだけ足に負担の少ない物がよいでしょう。”である．“仰向けです。”は，Task 1 （原）と Task 2 （機）で同様の機械翻訳文 (“It is his back.”) が用いられており，どちらの作業者も，ほとんどが，訂正文の動作の主語として “He（彼）”を用いていた．このため，条件 1 を満たすような適切な訂正は行われていたが，条件 2 を満たさない訂正文となっていた．このように，機械翻訳によって，文中で明示されていない動作主などが補完される場合，適切な翻訳が行われない場合がある．これは，機械翻訳適応文作成の段階で，作業者に動作主を明示するよう指示することで改善できる可能性がある．

5.4 Task 2 （機）で適切な訂正文が取得できない例

Task 1 （原）では適切な訂正文を取得できたが，Task 2 （機）では取得できなかった文が 1 文あった．この 1 文について表 6 に示す．表 6 で示した例は，原文と機械翻訳適応文のどちらも “主人”という単語を使っているが，原文での機械翻訳では “husband”，機械翻訳適応文での機械翻訳では “master”と翻訳されている．これによって，Task 2 （機）では，前述した評価基準の条件 2 を満たさない訂正文が作成された．このように，日本語では同じ表現をするが，英語では異なる表現になる単語については，折り返し翻訳を用いた機械翻訳適応文作成では対応できない限界がある．

6. おわりに

本稿では，クラウドソーシング上の単言語話者を用いた用例対訳作成において，機械翻訳が適切に行えるような，機械翻訳適応文を用いることによる効果の検証を行った．機

表 5: 評価値は低いが適切な訂正文を多く取得できた例

項目	Task 1 （原文を使用）	Task 2 （機械翻訳適応文を使用）
日本語	体が今どういう状態かわかります。	体の状態。
機械翻訳文	Body will know what state now.	Body state of.
折り返し翻訳文との RIBES の値	0.76	0.93
機械翻訳文の評価値	2.5	1
適切な訂正文の数	1	7
訂正文（一部）	It will tell you what kind of state the body is in.	Overall condition.
	The body will tell you what it's state is.	State of your body.
	Listen to your body	the state of the body.

・Task 2 （機）の方が機械翻訳文の評価値は低いが，取得できた適切な訂正文の数は多い．

表 6: Task 1 （原）でのみ適切な訂正文を取得できた例

項目	Task 1 （原文を使用）	Task 2 （機械翻訳適応文を使用）
日本語	今主人がこちらに向かっているので、着いたら払ってもらいます。	現在主人がこちらに向かっている為、彼が到着した後、彼がその支払いをします。
機械翻訳文	Now because my husband is headed here, have them pay when you get.	Since the current master is heading here, after he arrived, he will make that payment.
折り返し翻訳文との RIBES の値	0.55	0.96
機械翻訳文の評価値	3	2
訂正文（一部）	Now, my husband is headed here. He will pay when he gets here.	When the boss returns he will make the payment.
	Since my husband is headed here, please have him pay when he gets here.	the master will pay it when he gets here
	My husband will pay when he gets here.	He will make the payment after the current master arrives.

・機械翻訳適応文は，折り返し翻訳文との RIBES の値は高いが，機械翻訳文が，“主人” が “master” と翻訳されており，原文が示す “夫” という意味を変えてしまっている．そのため，作成された訂正文はすべて 5.1 節の条件 2 を満たさない．

械翻訳適応文を用いた用例対訳作成実験を行い，以下の結果を得た．

- 機械翻訳に適した機械翻訳適応文を用いて機械翻訳文の評価を行う場合でも，必ずしも評価値が原文を用いる場合よりも高くなるとは限らない．
- 機械翻訳適応文を用いて，機械翻訳文の訂正をクラウドソーシングに依頼する場合，取得できる適切な訂正文の数は，原文を用いる場合よりも増加する．
- 原文と同じ意味を示す機械翻訳適応文を作成する場合，折り返し翻訳を用いる手法では対応できない限界があるため，単語単位で原文との意味の一致を評価したり，主語や動作主を明示したりする必要がある．

今後は，機械翻訳適応文作成や，抽出におけるタスクの再検討を行い，機械翻訳文の訂正精度向上を目指す．また，本実験で作成した機械翻訳文の訂正文に対して，クラウドソーシングを用いた正確性評価を行い，クラウドソーシング上の単言語話者による評価データを取得する．

参考文献

[1] 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).

[2] 山本 里美, 福島 拓, 吉野 孝: クラウドソーシングにおける会話文を用いた応答用例対作成手法の提案, 情報処理学会論文誌, Vol. 56, No. 3, pp. 1080–1089 (2015).

[3] 山本 里美, 福島 拓, 吉野 孝: クラウドソーシング上の単言語話者による用例対訳作成手法への折り返し翻訳利用の提案, 情報処理学会, GN ワークショップ 2015, No.6, pp.1–7 (2015).

[4] Omar F. Zaidan and Chris Callison-Burch: Crowdsourcing Translation: Professional Quality from on-Professionals, '11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Vol. 1, pp. 1220–1229 (2011).

[5] Vamshi Ambati, Stephan Vogel, Jaime Carbonell: Collaborative Workflow for Crowdsourcing Translation, CSCW '12 Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, pp. 1191–1194 (2012).

[6] Chang Hu, Philip Resnik, and Benjamin B. Bederson.: Crowdsourced Monolingual Translation, ACM Trans. Comput.-Hum. Interact, vol. 21, No. 4, pp. 1–35 (2014).

[7] 平尾 努, 磯崎 秀樹, 須藤 克仁, Kevin Duh, 塚田 元, 永田 昌明, "語順の相関に基づく機械翻訳の自動評価法," 自然言語処理, Vol. 21, No. 3, pp. 421–444, 2014.

[8] Kevin Walker, Moussa Bamba, David Miller, Xisoyi Ma, Chris Cieri, and George Doddington, Multiple-Translation Arabic (MTA) Part 1, In Linguistic Data Consortium, Philadelphia (2003).