Vol.2015-GN-93 No.38
Vol.2015-CDS-12 No.38
Vol.2015-DCC-9 No.38
2015/1/27

クラウドソーシングにおける機械翻訳文の評価結果を活用した用例対訳作成手法の提案

山本 里美[1]　福島 拓[2]　吉野 孝[1]

概要：現在, グローバル化によって多言語間コミュニケーションの機会が増加している．多言語間コミュニケーションの支援として，機械翻訳や用例対訳などが用いられている．正確な情報の共有が重要となる医療分野などでは，十分に正確性の確保された用例対訳が使用されている．しかし，正確な情報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の収集は困難である．我々は，現在，機械翻訳文をクラウドソーシングにおいて評価・訂正を依頼することで，単言語話者であっても用例対訳作成を行うことができる手法の研究を行っている．これまで，クラウドソーシング作業者に会話文の形式で機械翻訳文を提示することで，応答用例対を作成できることが分かった．しかし，会話文を用いた応答用例対作成手法では，用例対訳にも使用可能な機械翻訳文や訂正文が作成されていても，用例対訳としての利用可能性の評価を専門家に依頼して行う必要があるため，応答用例対作成手法のみでは用例対訳の作成は行えなかった．本稿では，評価後の機械翻訳文を評価値によって分類し，評価値ごとに処理をかえることで，用例対訳と応答用例対を同時に作成する手法を提案する．

キーワード：クラウドソーシング，多言語間コミュニケーション，用例対訳，機械翻訳

1. はじめに

現在，世界的なグローバル化によって，多言語間コミュニケーションの機会が増加している．日本でも，外国人流学生数の増加 [1] や，訪日外国人数の増加 [2]，日本政府が 2020 年に訪日外国人数 2500 万人という目標を掲げている [3] ことから，今後も訪日外国人数は増加すると考えられる．また，それに伴い多言語間コミュニケーションの機会も増加すると考えられる．しかし，一般的に多言語を十分に習得することは難しく，日本語を理解できない外国人と日本人との間で，正確な情報の共有を行うことは非常に困難である，

その影響が顕著に現れる分野の 1 つに医療分野があり，医療通訳者による支援以外に，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として用例対訳や機械翻訳などの言語資源を組み合わせて利用することができる言語グリッド [*1] の活動などが行われている．現在，我々は，多言語用例対訳共有システム TackPad（タックパッド）[*2] による用例対訳の収集を行っている．TackPad には多くの用例や用例対訳 [*3] が収集されているが，医療現場で用いるために必要とされる数 [*4] には足りておらず，また，正確性評価が十分にされていないものも多い．専門家に依頼して用例対訳や応答用例対の作成を行う場合，十分な数の用例対訳を多言語において収集するためには，専門家の数と費用が多く必要であるため，必要な人的リソースと費用の確保が困難であると考えられる．

そこで我々は，多くの作業者を低コストで確保できるクラウドソーシングを用いた，単言語話者による用例対訳作成手法に関する研究を行っている．本稿では，応答用例対作成手法として提案された既存手法について，評価結果による分類を行うことで用例対訳と応答用例対を作成する手法を提案し，評価実験を行った．評価実験の結果より，本手法において，専門家に依頼することなく用例対訳や応答用例対として使用可能な機械翻訳文や訂正文が取得できる可能性があることがわかった．なお，用例対訳は，用例を多言語に翻訳した多言語コーパスのことを指し，応答用例対は，用例対訳の中でも質問とそ

[1] 和歌山大学システム工学部
Faculty of Systems Engineering, Wakayama University
Sakaedani 930, Wakayama, 640–8510 Japan
[2] 静岡大学大学院工学研究科
Graduate School of Engineering, Shizuoka University
Johoku 3–5–1, Naka-ku, Hamamatsu, 432–8561 Japan

[*1] http://langrid.org/jp/

[*2] http://med.tackpad.net/

[*3] 本稿では，正確性の確保が行われた多言語の用例の対を「用例対訳」，正確性の確保が行われていない多言語の対を「多言語テキストペア」とする

[*4] 1 言語あたり 3 万〜5 万文

© 2015 Information Processing Society of Japan

Vol.2015-GN-93 No.38
Vol.2015-CDS-12 No.38
Vol.2015-DCC-9 No.38
2015/1/27

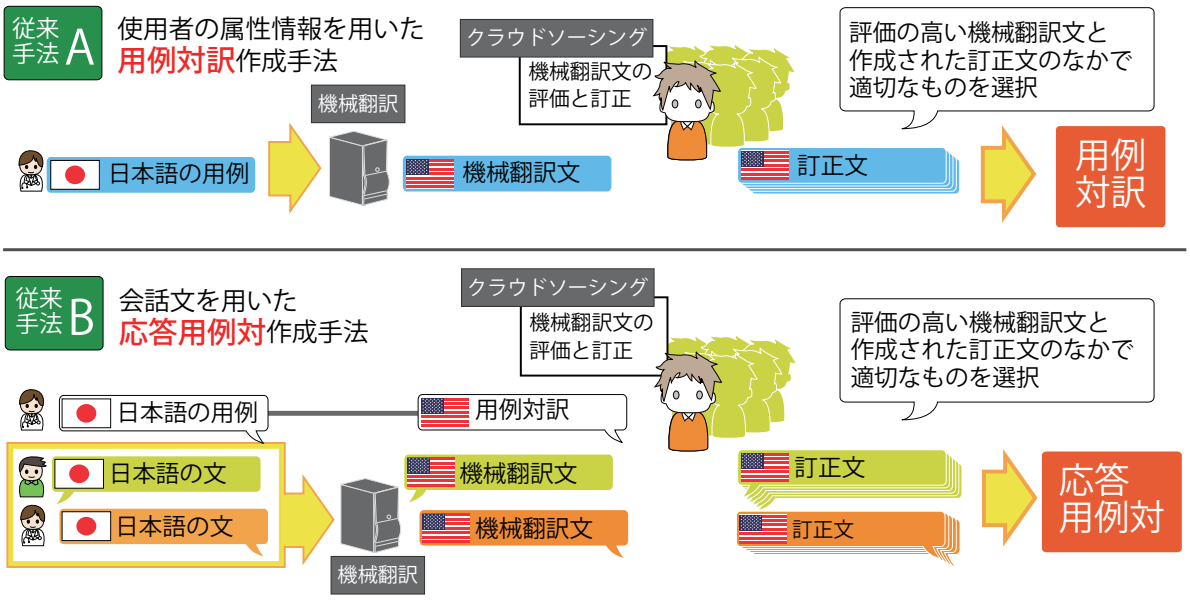


図 1 従来手法の概要

の回答の対のことを指す．

2. 関連研究

現在，クラウドソーシングを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．多言語テキストの正確性評価 [4] や多言語話者による用例対訳作成 [5] を行う研究や翻訳対象の文やクラウドソーシング作業者の特徴によって翻訳文を操作して品質の良い翻訳文を得る研究 [6]，クラウドソーシングと機械翻訳を併用してより品質の高い対訳コーパスや翻訳結果を取得する研究 [7], [8] などがある．多くの研究において，クラウドソーシングの作業者は非専門家を対象としているが，多言語を理解することができることが前提である．そのため，英語や中国語など，話者の多い言語についての対訳コーパス作成においては，多くのデータを収集可能だが，話者の少ない言語では，その言語を理解できる作業者が少ないため，十分な数の対訳コーパスの作成は困難であると考えられる．

我々は，正確性の低い機械翻訳文であっても，翻訳後の言語を母語とする人であれば，翻訳前の用例の意図を推測し，元々の意図にあった訂正を行うことが可能なのではないかと考えた．そこで，機械翻訳の翻訳後の言語を母語とするクラウドソーシング上の作業者に評価と訂正の依頼を行うことで，翻訳前の文と取得した訂正文から用例対訳を作成する手法を提案した [9]．単言語話者によって作成された訂正文と，機械翻訳前の用例を組み合わせることによって，用例対訳を作成することができれば，多くの用例が必要となる用例対訳の課題の 1 つである人的リソースの不足に対する解決方法の 1 つとなる可能性がある．しかし，全ての正確性の低い機械翻訳文を訂正することは困難である．

そのため，評価と訂正を行うクラウドソーシングの作業者に提示する情報によって，訂正精度の向上をはかる研究を行っている [10]．本稿では，取得した評価結果を評価値によって分類することで，専門家に評価を依頼することなく用例対訳や応答用例対として使用可能となる文の取得を目指す．

3. 提案手法と従来手法との比較

本章では，提案手法と従来手法の比較について述べる．3.1 節で従来手法の概要について説明し，3.2 節で従来手法における評価実験の結果について述べる．また，3.3 節で後述の従来手法 A，従来手法 B の比較を行い，3.4 節で従来手法と提案手法の比較を行う．

3.1 従来手法の概要

図 1 に従来手法の概要を示す．クラウドソーシング作業者に 1 文のみを提示して用例対訳を作成する手法を従来手法 A，会話文を提示して応答用例対を作成する手法を従来手法 B とする．

図 1 の従来手法 A では，1 つの用例に対して用例対訳を作成することを目的とし，従来手法 B では，会話文を提示することで，疑問文から始まる会話文に対する応答用例対の作成を目的としている．

従来手法 A では，用例対訳作成タスクに，用例の使用者の属性情報を使用する．属性情報とは，医療従事者や病院利用者などのことである．用例の使用者の属性情報は，従来手法 B における応答用例対作成にも使用している．

機械翻訳文の評価として流暢性の評価を 5 段階で行っている．流暢性評価の基準は，文献 [11] の評価基準 [*5]

[*5] 評価段階は，1：Incomprehensible(理解できない)，2：Dis-

© 2015 Information Processing Society of Japan

表 1 従来手法において作成された訂正文の例

	評価値	翻訳前の文	機械翻訳文	作業者によって作成された訂正文
(1)	5	頭が痛いです	It is a headache	I habe a headache.
(2)	1	ズキンズキンとした痛みがします。	Pain was Zukinzukin will.	Pain was severe.
(3)	1	ズキンズキンとした痛みがします。	Pain was Zukinzukin will.	The pain was in the heel.
(4)	2	体が今どういう状態かわかります。	Body will know what state now.	We will know your physical state now.
(5)	4	あまりありません。	I do not have much.	I do not have much of an appetite.
(6)	3	余りありません。	It is not too much.	I do not have any appetite.

表 2 **表 1** の (4)〜(6) をそれぞれに含む会話文

		1 文目（用例対訳）	2 文目（機械翻訳）	3 文目（機械翻訳）
(4)	日本語	検査の結果から何が分かりますか。	体が今どういう状態かわかります。	それに合わせて薬が出るんですね。
	英語	What can you tell from the test results?	Body will know what state now.	Medicine I'm out to match it.
(5)	日本語	食欲はありますか？	あまりありません。	朝ごはんを食べてきましたか？
	英語	How's your appetite?	I do not have much.	Did you have to eat breakfast?
(6)	日本語	食欲はありますか？	余りありません。	症状はずっと続いていますか。
	英語	How's your appetite?	It is not too much.	Symptoms Do you have followed all the way.

・1 文目は TackPad に用例対訳が登録されている用例を用いた．
・2 文目以降の英語が機械翻訳を用いた文であり，評価と訂正の対象とした．

を参考にした．従来手法 B では，文法などの他に，会話の流れとして適切かどうかということも加味して評価を行った．

3.2 従来手法における評価実験の結果

本節では，従来手法に A と従来手法 B における評価実験の結果について述べる．従来手法 A，B のどちらも，日本語の文に対して機械翻訳を用いて英語に翻訳し，英語を母語とする作業者に機械翻訳文の評価と訂正を依頼している．表 1 に，従来手法 A，B において作成された機械翻訳と訂正文の一部を示す．また，**表 2** に，従来手法 B で用いた，表 1(4)〜(6) をそれぞれに含む会話文を示す．なお，表中の評価値は，機械翻訳文に対して行われた 5 段階評価における中央値である．

従来手法 A を用いた用例対訳作成では，表 1(1) のような文の使用者に合わせて主語を変更するよう訂正や表 1(2) のように，精度の低い機械翻訳文を，意味を推測して別の表現を用いる訂正が行われた．このことから，精度の低い機械翻訳文であっても，作業者は文の意図を推測して機械翻訳文の訂正を行い，機械翻訳前の文の意図に近い対訳の作成が行えると考えられる．しかし，用例の使用者の属性情報を提示するだけでは，状況の限定が不足しているため，表 1(3) のように，作業者が状況を限定しすぎるような例があった．

そのため，機械翻訳文の訂正を行う作業者に提示する情報量を増やし，作業者がより正確に機械翻訳前の文の意図に沿った訂正が行えるような手法として，会話文を用いた従来手法 B を考えた．

しかし，従来手法 B では，会話文を提示して修正を依頼するため，表 1(4) のように過不足なく訂正が行われる場合と，表 1(5) のように会話の中での文としては適切だが，1 文の用例の対訳としては不適切な訂正文が作成される場合があった．

用例対訳としてではなく，応答用例対として使用可能な訂正文の取得が可能であったため，会話文を用いる従来手法 B では，用例対訳作成手法ではなく，会話文の対訳として使用される応答用例対の作成手法として提案し，評価実験を実験を行った．

従来手法 B を用いて作成した機械翻訳文や訂正文に対して，専門家に評価を依頼したところ，用例対訳として使用可能な文が存在することが分かった．専門家による評価は，機械翻訳前の文と機械翻訳文や訂正文が同じ意味であるかどうかの判断を 5 段階 [*6] で行う．3 名の専門家に評価を依頼したところ，評価値 5 となった機械翻訳文や訂正文の数は，3 名の専門家でそれぞれ異なり，平均は 623 文中 156 文，3 名全員が評価値を 5 とした機械翻訳文や訂正文は 21 文あった．また，表 1(6) のように，会話文の中での文としては不自然だが，1 つ文としてみると適切な翻訳が行われている機械翻訳文もある．しかし，従来手法 B は応答用例対の作成を目的として設計しているため，用例対訳として使用可能な訂正文と，応答用例対としてのみ使用可能な訂正文を区別する処理は

fluent 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)

[*6] 1 : まったく違う意味，2 : 雰囲気は残っているが，元の意味は分からない，3 : 意味はだいたいつかめる，4 : 文法などに多少問題があるが，だいたい同じ意味，5 : 同じ意味

© 2015 Information Processing Society of Japan

表 3　表 1(4)〜(6) をそれぞれに含む会話文

文 ID	機械翻訳文	評価値	
		従来手法 A	従来手法 B
13	It is not too much.	5	3
173	Now my husband because towards us, you get paid when you get.	2	1
238	What did you eat yesterday?	5	5

・表中の評価値とは，各文におけるすべての評価値の中央値である．

行っていない．そのため，従来手法 B では用例対訳として使用可能な機械翻訳文や訂正文は取得できるが，用例対訳として使用可能であるという判断ができないため，用例対訳として使用するには，応答用例対として使用可能であるかどうかの評価とは別に専門家に評価を依頼する必要がある．

3.3　従来手法 A と従来手法 B の比較

本項では，従来手法 A と従来手法 B に対して同じ用例のデータを用いて行った実験結果について述べる．使用したデータは，従来手法 B の評価実験において作成した会話文 38 組である．会話文の作成には Yahoo!クラウドソーシング [*7] を用い，1 文目は TackPad にすでに登録されている疑問文の用例であり，Yahoo!クラウドソーシングの作業者には，1 文目に続くように 2 文目と 3 文目の作成を依頼した．

従来手法 A では会話文として作成されたデータを 1 文ずつのデータとし，76 文の日本語の文から用例対訳作成を行った．従来手法 B では，会話文として作成されたデータを用い，1 文目は TackPad に登録されている正確性の確保された用例対訳，2,3 文目は機械翻訳文とし，38 組の応答用例対の作成を行った．

従来手法 B の評価実験で用いた機械翻訳を従来手法 A に使用して用例対訳を作成したところ，同じ機械翻訳文の評価であっても評価値や訂正文が異なることが分かった．各手法における評価値の例を**表 3** に示す．なお，表 3 の文 ID は，取得したデータを整理するために割り振ったものである．表 3 より，同じ機械翻訳文であっても，文 ID173 や 238 のように評価値が近い場合と，文 ID13 のように評価値が異なる場合があることが分かった．また，文 ID13 は表 3 より，1 つの文としての評価は高いため機械翻訳そのものは正しく行われていると考えられる．よって，用例対訳として使用可能である可能性がある．

3.4　従来手法と提案手法の比較

3.3 節より，従来手法 B を用いて取得した機械翻訳文に対する評価値について，後述の 4 章で述べる分類によって，機械翻訳文と訂正文の分類を行なうことで，用例対訳に使用可能な機械翻訳文や訂正文と，応答用例対に使用可能な機械翻訳文や訂正文を機械的に判定することが可能なのではないかと考えた．

従来手法 A は用例対訳作成を目的とし，従来手法 B は応答用例対作成を目的としていたため，従来手法 B において取得した応答用例対データの中に，用例対訳として使用可能なデータがあったとしても，それを判別する方法がなかったため，従来手法 B における用例対訳作成は行えなかった．提案手法では，従来手法 B において取得した評価データをもとに機械翻訳文とそれに対する訂正文を 1 つのグループとして分類し，それぞれの分類にあわせた処理を行うことで，用例対訳作成に使用可能なデータと応答用例対作成に使用可能なデータを機械的に判定することができる可能性がある．これにより，従来手法ではできなかった，1 つの評価データから，用例対訳と応答用例対を同時に作成していくことが可能となると考えられる．また，同じデータを用いて用例対訳と応答用例対が作成できれば，不正なデータを重複して使用することが減るため，より低いコストで用例対訳や応答用例対を作成できる可能性がある．

4.　評価値による分類における機械翻訳文の特徴の仮説

本章では，従来手法 B において使用した機械翻訳文に対する評価値ごとの分類と，それぞれの分類における機械翻訳文の特徴についての仮説を述べる．なお，分類における評価値は，各機械翻訳文に対するすべての評価値の中央値とする．評価値による分類と，分類ごとに考えられる機械翻訳文の特徴は以下である．

分類 1　評価値 5 の機械翻訳
分類 1 は，精度の高い機械翻訳が行われた機械翻訳文である．この文は，用例対訳としても，応答用例対として使用可能である．

分類 2　評価値 4 または 4.5 の機械翻訳
分類 2 は，ある程度正確な機械翻訳が行われた機械翻訳文である．
用例対訳としても，応答用例対としても使用可能である．しかし，会話文中で使用する場合には一部流暢でないところがある．

分類 3　評価値 3 または 3.5 の機械翻訳
分類 3 は，文法などに誤りはないが，翻訳語の言語として不自然な表現が用いられている，もしくは会話の中での文としては不自然な機械翻訳文である．

分類 4　評価値 2 または 2.5 の機械翻訳
分類 4 は，精度の低い機械翻訳が行われた機械翻訳文である．多くの作業者に文の意図が伝わっていない．

分類 5　評価値 1 または 1.5 の機械翻訳

[*7] http://crowdsourcing.yahoo.co.jp/

© 2015 Information Processing Society of Japan

Vol.2015-GN-93 No.38
Vol.2015-CDS-12 No.38
Vol.2015-DCC-9 No.38
2015/1/27

分類 5 は，精度が非常に低い機械翻訳が行われた．

分類 3 の機械翻訳文は，文法には問題がないが，翻訳後の言語の話者にとって不自然な表現が使われている場合と，会話の中での文として不自然な場合の 2 通りが考えられる．会話の中での文として不自然な場合には従来手法 A を用いると高い評価となる可能性がある．

5. 提案手法

本章では，従来手法 B より取得したデータを 4 章で述べた分類によって，その後の処理を変えることで，用例対訳と応答用例対を作成する手法について述べる．

5.1 分類ごとの処理

本節では，4 章で述べた分類ごとの機械翻訳文と訂正文の処理について述べる．タスク画面の例を**図 2** に示す．

4 章の分類 1〜3 における処理について，以下に示す．

分類 1 分類 1 は，機械翻訳の精度が高い可能性があるため，機械翻訳文はそのまま用例対訳としても応答用例対としても使用可能であると考えられる．そのため，分類 1 の機械翻訳文や訂正文は，特に処理を行わず，後述の**図 3**(5) において，用例対訳や応答用例対として使用可能かどうかの評価を行う．

分類 2 分類 2 は，分類 1 の機械翻訳文程ではないが，ある程度精度の高い機械翻訳が行われたと考えられるため，機械翻訳文や訂正文が，会話文中において適切な文であるかどうかの判定を行う．適切な文であるかどうかの判定はクラウドソーシングを用いて行う．タスクの画面例を図 2 に示す．作業者には，会話を行う人物の属性情報と，会話文を提示し，機械翻訳文や訂正文から，会話文中において適切な文を選択する．

分類 3 分類 3 は，作業者にある程度文の意図が伝わっている可能性が高いと考えられるため，訂正文は会話文中の文として使用可能であると考えられる．よって，訂正文は分類 2 と同様，会話文中において適切な文であるかどうかの判定を行う．また，機械翻訳文は会話文中の文として不適切であっても，正しい機械翻訳が行われている可能性があるため，機械翻訳文に対しては従来手法 A を用いて評価を行い，従来手法 B と同様に分類ごとに処理を行う．なお，分類 3 の機械翻訳文については処理を行わない．

5.2 提案手法の概要

本節では，従来手法 B において取得した評価結果を評価値によって分類してそれぞれに異なる処理を行うことによって用例対訳と応答用例対を作成する手法について説明する．図 3 に提案手法の流れを示す．本手法は以下の 5 ステップで構成されている．

Step 1 従来手法 B を用いた応答用例対作成

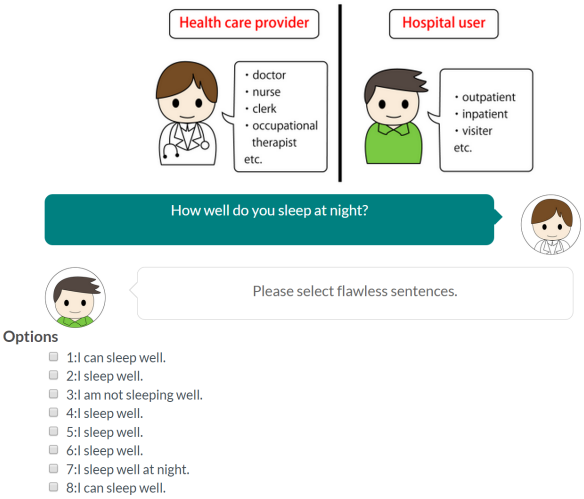


図 2 適切な文を選択するタスク画面

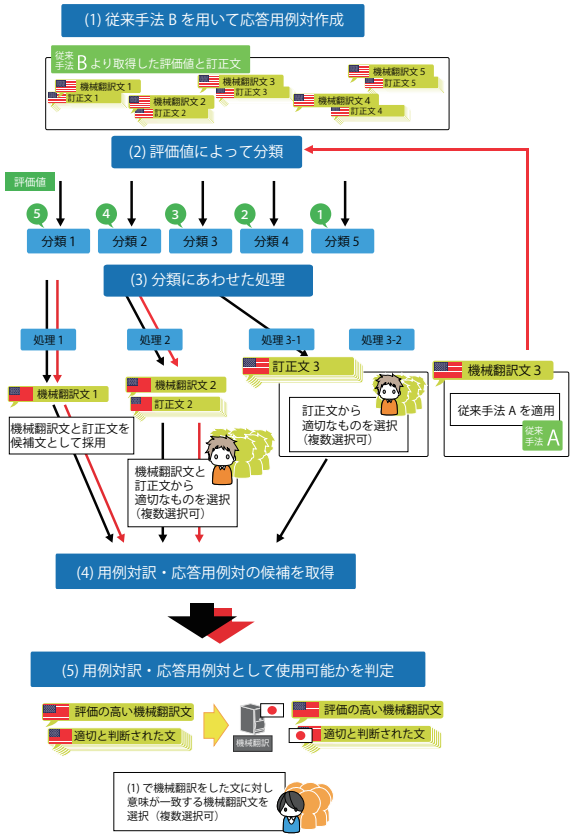


図 3 評価結果による分類ごとの処理の流れ

図 3(1) で，従来手法 B による応答用例対作成を行う．

Step 2 評価値による分類

図 3(2) で，Step 1 で取得した評価データより，機械翻訳文や訂正文を評価値によって分類する．分類は 4 章で述べた分類を用いる．

Step 3 分類ごとの処理

図 3(3) で，分類 1〜3 に対してそれぞれに 5.1 節で述べた処理を適用する．機械翻訳文と訂正文を 1 つのグループとして扱う．分類 2，3 に対する処理で

© 2015 Information Processing Society of Japan



図 4　日本人による機械翻訳文や訂正文の評価タスク画面

は，Step 1 での機械翻訳後の言語を母語とするクラウドソーシングの作業者に適切な機械翻訳文や訂正文の選択を依頼する．なお，選択は複数選択を可能とする．

Step 4　用例対訳・応答用例対の候補を取得
図 3(4) で，Step 3 の結果より，用例対訳や，応答用例対として使用可能であると考えられる機械翻訳文や用例対訳を取得する．分類 2，3 における候補文は，選択タスクを行った作業者の過半数が適切と判断した機械翻訳文や訂正文とする．

Step 5　用例対訳・応用用例対として使用可能かを判定
図 3(5) で，Step 4 で取得した候補に対して，クラウドソーシングを用いて用例対訳や応答用例対として使用可能かどうかの判定を行う．判定の方法はタスク画面の例を**図 4** に示す．

Step 5 のクラウドソーシングを用いた，用例対訳や応答用例対として使用可能かの判定方法として，本手法では，作成された用例対訳や応答用例対の候補文に対して，会話文の作成を行ったときの対象の言語に機械翻訳を行い，本手法を適用する前の文と，候補文を機械翻訳した文の比較を依頼する．本手法を適用する前の文と，候補文の機械翻訳文で意味が異ならなければ，用例対訳や応答用例対の作成が正しく行われていると判断できる可能性がある．

6.　評価実験

本章では，5 章で述べた手法の評価実験について述べる．

6.1　実験用データセット

本節では，実験に使用するデータについて述べる．本実験では，従来手法 B の評価実験において取得した会話文 38 組とその評価結果を使用した．使用した会話文は Yahoo!クラウドソーシング [*8] において作成を依頼したものである．なお，機械翻訳は Google 翻訳 [*9] を用い，英語に翻訳して評価と訂正を行った．機械翻訳文の評価と訂正は CrowdFlower[*10] において国を United States として登録している作業者に依頼した．また，1 文につき 10 人の作業者に評価と訂正を依頼したため，取得した評価データは 760 件である [*11]．取得した訂正文数は 1 文につき最大 10 文であり，“none”や “base sentence”，“NA”などの不適切な文は使用しなかった．また，重複している訂正文は 1 つの訂正文とし，区別をしなかった．なお，本実験では作成した会話文のうち 2 文目についてのみ本手法を適用した．そのため，実際に分類を行った機械翻訳文の数は 38 文であり，取得した評価データの総数は 380 件である．

6.2　実験手順

本節では，評価実験の実験手順について述べる．

本手法における分類の結果，38 文の機械翻訳文中，分類 1 は 5 文，分類 2 は 14 文，分類 3 は 10 文となった．分類 1 の 5 文は機械翻訳文，訂正文ともに後述の日本人作業者による評価を行い，分類 2，分類 3 は 5.2 で述べた処理後に取得した用例対訳や応答用例対の候補文に対して，日本人作業者による評価を行う．

用例対訳や応答用例対の候補文の取得には CrowdFlower を用い，1 つの機械翻訳文と訂正文のグループあたり 10 件の評価を行った．不正な訂正文を除いた結果，有用な訂正文が存在しない機械翻訳文もあった．そのため，分類した機械翻訳文の数と，候補文を作成した機械翻訳文の数は一致しない．

本実験では，作成された機械翻訳文や訂正文に対する評価方法として Yahoo!クラウドソーシングを用いて日本人の作業者に評価タスクを依頼した．Yahoo!クラウドソーシングの作業者には，図 3(4) で取得した用例対訳や応答用例対の候補となる機械翻訳文や訂正文に対して，機械翻訳を用いて日本語に翻訳し，図 3(1) で用いる，クラウドソーシングによって作成された日本語の文と比較し，置き換えても意味が変化しないものを選択するタスクを依頼した．1 つの設問につき 10 名の作業者に評価を依頼し，1 タスクあたりの設問は 5 問とした．

7.　実験結果と考察

本章では，6 章における実験の結果について考察を行う．本実験では，作成した 38 組の会話文のうち，2 文目のみを使用したため，実際に用例対訳や応答用例対の作成を行った文は 38 文である．

[*8] http://crowdsourcing.yahoo.co.jp/
[*9] https://translate.google.com/
[*10] http://crowdflower.com/
[*11] 会話文の 2 文目と 3 文目それぞれに評価と訂正文の作成を行った．

© 2015 Information Processing Society of Japan

本手法を用いて取得した，用例対訳として使用可能だと考えられる機械翻訳文や訂正文は 14 文，応答用例対として使用可能だと考えられる機械翻訳文や訂正文は 26 文であった．本実験において使用したデータや取得したデータの種類とその数を**表 4** に示す．表 4(12)，(13) が，本実験において最終的に取得できた応答用例対や用例対訳として使用可能性がある機械翻訳文や訂正文の数である．

大学生 3 名に従来手法 B で作成された機械翻訳文と訂正文について，応答用例対として使用可能かどうかの判定を依頼した．大学生 3 名には，本実験で使用した 38 組の日本語の会話文と，作成された機械翻訳文の訂正文のデータを提示し，会話文中の訳として正しいものを選択してもらった．本実験において取得した，表 4(12) の 26 文は，その全てが，大学生 3 名が選択した機械翻訳文や訂正文と一致した．これは，本手法において正確性の高い応答用例対を取得できた可能性を示す．

また，専門家 3 名に従来手法 B で作成された機械翻訳文と訂正文について，用例対訳として使用可能かどうかの評価を依頼した．用例対訳として使用可能かどうかの評価では，大学生に依頼した応答用例対としての評価とは異なり，機械翻訳文や訂正文が，機械翻訳前の文に対して意味が同じかどうかに関して 5 段階 [*12] の評価を依頼した．評価が 5 となった機械翻訳文や訂正文は用例対訳として使用可能であると考えられる．本実験において取得した，表 4(13) の 14 文は，その全てが，専門家による評価によって用例対訳として使用可能であると判断された文と一致した．これは，本手法において正確性の高い用例対訳を取得できた可能性を示す．

しかし，日本語への機械翻訳の精度が悪く，応答用例対や用例対訳として使用可能であると判断されなかった機械翻訳文や訂正文も存在した．その例を**表 5** に示す．

文 ID167 は，“diet”の “d”が大文字になっているため，“diet”の翻訳が「食事」や「食事療法」ではなく「国会」と翻訳されてしまっている．しかし，weblio 英語翻訳 [*13] を用いた場合の翻訳結果は「ダイエットほど重要であるものはない。」となり，SDL の無料翻訳サービス [*14] では「食事療法は最も重要な事である。」という結果になるため，これらの機械翻訳サービスを用いて評価を行っていた場合，作成された訂正文が使用可能であると判断される可能性がある．

これらのことから，本手法を用いて用例対訳や応答用例対を作成する場合，作成された機械翻訳文や訂正文から，用例対訳や応答用例対として使用可能な全ての文を取得することは難しいが，用例対訳として使用可能な文と応答用例対として使用可能な文の両方を作成できる点において，本手法は有用であると考えられる．

表 4 本実験で使用または取得したデータの種類とその数

	データの種類	データの数
(1)	英語に機械翻訳を行った文	76 文
(2)	1 文あたりの評価回数	10 回
(3)	取得した評価データ	760 件
(4)	図 3(2) 以降の処理を行った機械翻訳文	8 文
(5)	(4) について取得した評価データ	380 件
(6)	(5) から不正な訂正文を除いた訂正文	276 件
(7)	分類 1 に分類された機械翻訳文	5 文
(8)	分類 2 に分類された機械翻訳文	14 文
(9)	分類 3 に分類された機械翻訳文	10 文
(10)	分類 3 へ分類後，従来手法 A によって分類 1 に分類された機械翻訳文	5 文
(11)	分類 3 へ分類後，従来手法 A によって分類 2 に分類された機械翻訳文	0 文
(12)	応答用例対として使用可能と考えられる機械翻訳文や訂正文	26 文
(13)	用例対訳として使用可能と考えられる機械翻訳文や訂正文	14 文

・(3) 不正な訂正文とは，“none”や “base sentence”，“NA”などのことであり，機械的に除外した．

8. おわりに

応答用例対作成手法として提案された手法について，評価結果による分類を行うことで用例対訳と応答用例対を作成する手法を提案し，評価実験を行った．評価実験の結果より，本手法において，専門家に依頼することなく用例対訳や応答用例対として使用可能な機械翻訳文や訂正文が取得できる可能性があることがわかった．

しかし，本実験では用例対訳や応答用例対として使用可能な全ての文を取得することは出来なかった．今後は，本手法では取得できなかったが用例対訳や応答用例対として使用可能であった訂正文を取得するための処理やタスクの設計についての検討を行う．

謝辞 本研究の一部は，JST A-STEP「多段クラウドソーシングを活用した多言語用例対訳プラットフォームの構築」，JSPS 科研費 24220002 および 26730105 の助成を受けた．

参考文献

[1] 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 2014-12-23).
[2] 日本政府観光局 (JNTO)：訪日外客数の動向 (オンライン)，入手先 ⟨http://www.jnto.go.jp/jpn/reference/tourism_data/visitor_trends/⟩ (参照 2014-12-23).
[3] 法務省：訪日外国人 2500 万人時代の出入国管理行政検討会議 (オンライン)，入手先

[*12] 評価段階は，1:全く違う意味，2:雰囲気は残っているが，元の意味は分からない，3:意味はだいたいつかめる，4:文法などに多少問題があるが，だいたい同じ意味，5:同じ意味
[*13] http://translate.weblio.jp/
[*14] http://www.freetranslation.com/ja/

© 2015 Information Processing Society of Japan

表 5　日本語への翻訳精度が低い文の例

文 ID	従来手法 B に用いた文	翻訳前の文	翻訳後の文
55	仰向けです。	On my back.	私の背中に。
167	まずは食生活です。	Diet is the most important thing.	国会は最も重要なものです。

・従来手法 B に用いた文とは，6.1 節で述べた，クラウドソーシング作業者によって作成された文である．

⟨http://www.moj.go.jp/nyuukokukanri/kouhou/nyuukokukanri01_00103.html⟩ (参照 2014-12-23).

[4] Callison-Burch, C.: Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk, *Proceedings of EMNLP 2009*, pp. 286–295 (2009).

[5] Negri, M. and Mehdad, Y.: Creating a Bi-lingual Entailment Corpus through Translations with Mechanical Turk: $100 for a 10-day Rush, *Proceedings of the NAACL HLT 2010*, pp. 212–216 (2010).

[6] F.Zaidan, O. and Callison-Burch, C.: Crowdsourcing Translation:Professional Quality from Non-Professionals, ProceedingHLT ‘11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Vol.1, pp. 1220–1229 (2011).

[7] Vamshi Ambati，Stephan Vogel : Can Crowds Build Parallel Corpora for Machine Translation Systems?，CSLDAMT ’10 Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk，pp. 62–65 (2010).

[8] Philip Resnik，Olivia Buzec，Yakov Kronrod，Chang Hu，Alexander J. Quinn，Benjamin B. Bederson: Using Targeted Paraphrasing and Monolingual Crowdsourcing to Improve Translation, Transactions on Intelligent Systems and Technology (TIST), Vol.4, No.3, Article No.38 (2013).

[9] 福島拓, 吉野孝: クラウドソーシング上の単言語話者に依頼可能な多言語用例対訳作成手法の提案と評価, 言語処理学会第 19 回年次大会 (NLP2013), pp. 302–305 (2013).

[10] 山本 里美，福島 拓，吉野 孝 : クラウドソーシング上における使用者の属性情報を用いた用例対訳生成手法の提案，電子情報通信学会技術報告，AI2013-37，pp. 7–12 (2014).

[11] 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).

© 2015 Information Processing Society of Japan