

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 19 日現在

機関番号：34406

研究種目：若手研究(B)

研究期間：2014～2016

課題番号：26730105

研究課題名(和文) 正確性の担保を手動で行うリアルタイム自動翻訳技術の研究

研究課題名(英文) Study on Real-time Automatic Translation Technology using Accurate Texts Created by Users

研究代表者

福島 拓 (Fukushima, Taku)

大阪工業大学・情報科学部・講師

研究者番号：40714829

交付決定額(研究期間全体)：(直接経費) 2,900,000円

研究成果の概要(和文)：機械翻訳技術は急速に発展しているが、完全に正確な翻訳は現時点では難しい。そこで、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。用例対訳は、あらかじめ正確に翻訳された用例群である。本研究では、用例対訳を拡張して用いた自動翻訳手法の実装と評価を行った。実験から、提案手法を用いることにより正確な多言語間対話支援が可能であることを確認した。

研究成果の概要(英文)：Machine translation technology has been developed, but complete machine translations are difficult to obtain at present. A parallel text that combines example sentences and their accurate translation is used in a multilingual communication support. In this study, we propose and develop an automatic translation method using expanded parallel corpus. In the experiments, we show that the proposed method can accurately support multilingual communication.

研究分野：ヒューマンインタフェース・インタラクション

キーワード：多言語間対話支援 用例対訳 ヒューマンコンピューテーション

1. 研究開始当初の背景

多言語間対話支援を目的として、機械翻訳機の研究開発や、機械翻訳機を用いた対話支援技術の研究開発が多く行われている。しかし、機械翻訳は不正確な結果を出力する場合がある。医療や防災などの正確性が求められる分野では、翻訳誤りによって医療過誤や伝達誤りなどが発生し、それらによって生命の危険が引き起こされる可能性がある。

研究代表者は、正確な多言語間対話支援を目的として、正確な翻訳が可能な「用例対訳」に関する研究をこれまで行ってきている。しかし、用例対訳はあらかじめ人手で翻訳された文の集合であり、用例対訳と翻訳したい文とが完全一致しないと翻訳を行うことができない。このため、用例対訳が翻訳可能な文数は機械翻訳と比較して少ない状態となっている。また、既存手法では、完全一致の用例対訳が存在しない場合は、対話支援システムの利用者が類似用例対訳の検索を行うことが一般的であった。しかし、検索および選択に時間がかかるため、対話のリアルタイム性が低下するという課題が存在していた。

2. 研究の目的

本研究では、正確な多言語間対話支援が求められている分野を対象とした、自動翻訳技術の構築を目的とする。本研究では研究成果の適用先を医療分野とするが、構築手法を医療分野へ特化しないことで、他の正確性が求められる分野への波及を可能とする手法の確立を目指す。

前述の目的を達成するために、本研究では、正確な翻訳が可能であるが、翻訳可能な文数が少ない用例対訳を用いて対話支援を行う。その際、類似文や文脈情報、音声を用いて用例対訳データベースの拡張を行う。このことで、用例対訳の翻訳可能な文数を増加させる。また、データベースの拡張を行うことで、リアルタイム性の確保も同時に行う。

3. 研究の方法

本研究成果は、人を介したシステムへ適用を行う。このため、提案手法の評価のために、提案手法を適用したシステムを人に利用してもらった上で評価を行い、提案手法の改善を行うという手順の研究手法をとった。

本研究では、まず、提案するリアルタイム自動翻訳技術の利用可能性の調査を行った。その上で、各要素技術の研究開発および評価を行うことで研究活動を進めた。

4. 研究成果

研究成果として、(1)～(5)の各調査を行い、以下の知見を得た。

(1)質問回答対および類似文を用いた、リアルタイム自動翻訳技術の基礎調査

文脈情報の一つである質問回答対や類似文と、用例対訳とを用いた、応答用例対構築

モデルの提案を行った。本モデルが本研究の基礎技術にあたる。また、モデルを適用したシステムを用いて実験を行い、モデルの有用性検証を行った。

実験結果より、次の3点を明らかにした。提案モデルは、文脈に適した用例対訳の提示が可能であることを明らかにした。また、完全に一致していない用例対訳であった場合も、入力文と意味が同じであった場合は、用例対訳を利用可能にできることを明らかにした。これらは、従来の対話型用例対訳利用システムにおける課題を解決している。応答用例対の自動収集により、適切な回答選択の支援の可能性を示した。質問回答対への類似文の付与により、用例対訳を用いた自由な対話の可能性を示した。

(2)音声入力を用いた用例対訳検索手法

音声を用いた用例対訳検索では迅速な対話支援が可能である。しかし、従来手法では不適切な用例を抽出する機会が多く存在していた。そこで、用例対訳コーパスに含まれる文を否定文や疑問文に変換して追加した、拡張コーパスを用いた用例対訳検索手法を提案し、手法の効果検証を行った。

実験結果より、次の2点を明らかにした。拡張コーパスを用いることで、従来手法よりも高い適合率となる検索結果を利用者に提示可能であることを示した。正解が存在しない文を検索した場合、利用者に提示する検索結果数を減少可能であることを示した。このことで、利用者の用例対訳選択時の負担軽減が可能となった。

(3)クラウドソーシングを用いた用例対訳および質問回答対作成手法

研究成果(1)では、提案する自動翻訳技術を用いることで正確な対話支援が可能であることが示された。しかし、既存の用例対訳や質問回答対の数が少ないため、大幅に増加が必要であることも示された。そこで、人々への作業や業務の委託を行う、クラウドソーシングを用いた用例対訳や質問回答対の作成を行った。本研究では特に、多言語話者よりも人数が多い単言語話者を用いた、多言語データベースの作成を行った。

クラウドソーシングを用いた各実験およびその結果より、次の2点を明らかにした。

複数の機械翻訳機とクラウドソーシング上の単言語話者とを用いた、多言語用例対訳作成手法を提案した。実験結果より、単一の機械翻訳機を用いるよりも、複数の機械翻訳機を用いた方が、より正確な用例対訳の作成が可能であることが示された。また、最も多くの作業者が作成した対訳候補を選択する手法よりも、対訳候補のコサイン類似度を用いた抽出手法の方が、正確な対訳文を多く抽出できることが示された。機械翻訳を対象言語に翻訳の上、再度、もとの言語に翻訳を行う手法を折り返し翻訳と呼ぶ。この折り返し

し翻訳の結果をクラウドソーシング上の単言語話者に提示することで、質問回答を作成する手法を提案した。実験結果より、折り返し翻訳結果と、もとの文との一致度を作業者に提示することで、適切な質問回答が作成される可能性が示された。

(4)穴あき用例を用いた用例対訳作成手法

研究成果(3)と同様に、用例対訳の作成を目的として、穴あき用例を用いた用例対訳作成手法を構築した。穴あき用例とは、用例の一部の単語を穴あきにして入れ替え可能とした用例である。この穴あき用例の特性を活用した用例対訳作成システムを構築した。本手法では、穴あき部分に別の単語を入れた文を新たに作成し、それらの正確性判定を手で行うことで新たな用例対訳の作成を行っている。

また、システムを用いて行った用例対訳作成実験より、次の2点を明らかにした。穴あき用例と穴埋め単語の組み合わせを変えることで、多くの適切な用例作成が可能であることを示した。穴埋め単語と多言語単語辞書、用例対訳内のデータである言語間リンクを活用することで、対訳文の適切性判定が可能であることを示した。

(5)多言語対話シート作成システム

前述の要素技術を活用した、多言語対話シート作成システムの構築を行った。本システムは、医療従事者が外国人患者との対話で使用するテキスト群をまとめたものである、「対話シート」を作成・共有を行う。対話シートは、医療従事者があらかじめ作成することで、患者との対話場面における対話のリアルタイム性を確保している。また、医療従事者の対話シート作成から実際に利用するまでの時間差を用いることで、使用する用例対訳の正確性向上が可能な仕組みとしている。画面例を図1に示す。医療従事者は伝達したい言葉を図1から選択する。システムは選択されたテキストを、用例対訳を用いて翻訳して患者に提示するため、正確かつ迅速な多言語間対話支援が可能となる。

また、本システムを用いた実験結果より、次の2点を明らかにした。医療従事者が医

中国語

対話ボタンか医療従事者入力欄の翻訳をクリックすると患者言語の翻訳文が表示されます。左の言語切替ボタンで言語の変更ができます。

この病院は初めてですか？ お薬が必要ですか？ 質問はありますか？

どこが痛いですか？ 今までに大きな病気をしたことはありますか？ 保険証を見せて下さい。

これで診察は終わります。お疲れ様でした。 病名を教えてください 熱を測って下さい。

医療従事者入力欄: _____

翻訳

図1：多言語対話シート作成システム画面例

療現場で利用するテキストをあらかじめ登録して対話シートを作成するシステムにより、医療従事者と患者の間の円滑な対話支援の可能性を示した。伝達文の翻訳精度改善を行う機能による、新しい用例対訳作成の効率化の可能性を示した。

また、医療従事者による試用から、医療現場への適用可能性が示された。今後はシステムの改良の上、医療現場への適用を目指す。

5. 主な発表論文等

〔雑誌論文〕(計5件)

- 尾崎文香, 福島拓, 重野亜久里: LuPaCa: 医療従事者による用例登録を可能とした多言語対話シート作成システム, 情報処理学会論文誌, Vol.58, No.1, pp.134-142 (2017). [査読有]
<http://ci.nii.ac.jp/naid/170000131162>
- 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). [査読有]
<http://ci.nii.ac.jp/naid/110009884103>
- 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). [査読有]
<http://ci.nii.ac.jp/naid/110009867104>
- 吉野孝, 西田祥子, 福島拓: 携帯端末の通知機能を活用した用例対訳登録における利用者への貢献情報の提示による動機付けの検証, 情報処理学会論文誌, Vol.56, No.1, pp.118-125 (2015). [査読有]
<http://ci.nii.ac.jp/naid/110009867095>
- Taku Fukushima and Takashi Yoshino: Translation Repair Method for Improving Accuracy of Translated Sentences, IEICE Transactions on Information and Systems, Vol.E97-D, No.6, pp.1528-1534 (2014). [査読有]
DOI: 10.1587/transinf.E97.D.1528

〔学会発表〕(計20件)

- 福島拓, 吉野孝: 穴あき用例を活用した対訳作成手法の比較, 電子情報通信学会, 人工知能と知識処理, 2017年2月27日, 京都大学(京都府京都市)。
- 山本里美, 福島拓, 吉野孝: 医療現場を想定した会話文におけるクラウドソーシングを用いた用例対訳作成手法の効果, 電子情報通信学会, 人工知能と知識処理, 2017年2月27日, 京都大学(京都府京都市)。

3. 山本里美, 福島拓, 吉野孝: クラウドソーシングを用いた機械翻訳適応文作成における複数翻訳機の効果の検証, 情報処理学会, グループウェアとネットワークサービス研究会, 2017年1月21日, 土庄町総合会館(香川県土庄町).
4. 福島拓, 吉野孝: 穴あき用例を用いた多言語用例対訳作成手法の効果, 情報処理学会, グループウェアとネットワークサービス研究会, 2017年1月21日, 土庄町総合会館(香川県土庄町).
5. 山本里美, 福島拓, 吉野孝: 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, 2016年度情報処理学会関西支部支部大会, 2016年9月26日, 大阪大学中之島センター(大阪府大阪市).
6. 福島拓, 吉野孝: 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, 情報処理学会, マルチメディア, 分散, 協調とモバイル(DICOMO2016)シンポジウム, 2016年7月6日, 鳥羽シーサイドホテル(三重県鳥羽市).
7. 山本里美, 福島拓, 吉野孝: クラウドソーシングを用いた機械翻訳に適した文の作成に影響を及ぼす原文の特徴, 電子情報通信学会, 人工知能と知識処理, 2016年2月29日, 京都大学(京都府京都市).
8. 福島拓, 甲斐充彦: 音声入力を用いた用例対訳検索における拡張コーパスの利用, 電子情報通信学会, 人工知能と知識処理, 2016年2月29日, 京都大学(京都府京都市).
9. 尾崎文香, 福島拓, 重野亜久里: 医療分野を対象とした多言語対話シート作成システムの開発と評価, 情報処理学会, グループウェアとネットワークサービス研究会, 2016年1月21日, 天草市民センター(熊本県天草市).
10. 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, グループウェアとネットワークサービスワークショップ2015, 2015年11月28日, ホテルニュー塩原(栃木県那須塩原市).
11. 山本里美, 福島拓, 吉野孝: 後編集文と機械翻訳文との意味の近さを用いた単言語話者による用例対訳作成手法の提案, 2015年度情報処理学会関西支部支部大会, 2015年9月28日, 大阪大学中之島センター(大阪府大阪市).
12. 山本里美, 福島拓, 吉野孝, 曾根原登: 外国人観光客向けマイナーな観光地情報収集のためのクラウドソーシング利用手法の提案, ヒューマンインタフェースシンポジウム2015, 2015年9月2日, はこだて未来大学(北海道函館市).
13. 福島拓, 前島康宏, 甲斐充彦: 類似度を用いた可視化情報の提示による問診票回答収集効果の検証, 情報処理学会, マルチメディア, 分散, 協調とモバイル(DICOMO2015)シンポジウム, 2015年7月9日, ホテル安比グランド(岩手県八幡平市).
14. 福島拓, 吉野孝: クラウドソーシング上の単言語話者を対象とした翻訳パズルの提案, 人工知能学会全国大会, 2015年6月1日, はこだて未来大学(北海道函館市).
15. 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, 人工知能と知識処理, 2015年2月23日, 京都大学(京都府京都市).
16. 山本里美, 福島拓, 吉野孝: 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の多言語への適用, 電子情報通信学会, 人工知能と知識処理, 2015年2月23日, 京都大学(京都府京都市).
17. 山本里美, 福島拓, 吉野孝: クラウドソーシングを用いた応答用例対作成手法の提案, 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, 情報処理学会, グループウェアとネットワークサービス研究会, 2015年1月27日, 壱岐文化ホール(長崎県壱岐市).
18. 福島拓, 吉野孝: 翻訳パズル: クラウドソーシング上における不完全な翻訳を用いた対訳作成手法, 情報処理学会, グループウェアとネットワークサービス研究会, 2015年1月27日, 壱岐文化ホール(長崎県壱岐市).
19. 山本里美, 福島拓, 吉野孝: クラウドソーシングを用いた応答用例対作成手法の提案, 情報処理学会, マルチメディア, 分散, 協調とモバイル(DICOMO2014)シンポジウム, 2014年7月10日, ホテル泉慶(新潟県新発田市).
20. 福島拓, 吉野孝: クラウドソーシングを用いた高精度対訳作成のための低品質翻訳の活用, 情報処理学会, マルチメディア, 分散, 協調とモバイル(DICOMO2014)シンポジウム, 2014年7月10日, ホテル泉慶(新潟県新発田市).

〔その他〕
 ホームページ等
<http://taku.fukushima.jp/>

6. 研究組織
 (1) 研究代表者
 福島 拓 (FUKUSHIMA, Taku)
 大阪工業大学・情報科学部・講師
 研究者番号: 40714829