

科学研究費助成事業 研究成果報告書

令和 5 年 9 月 18 日現在

機関番号：32686

研究種目：基盤研究(B) (一般)

研究期間：2020～2022

課題番号：20H04486

研究課題名(和文) 翻訳者の訳出プロセスの可視化と、翻訳・言語研究の共有基盤の構築

研究課題名(英文) Visualizing human translator's translation process for establishing a research based shared among language researchers

研究代表者

山田 優 (Yamada, Masaru)

立教大学・異文化コミュニケーション学部・教授

研究者番号：70645001

交付決定額(研究期間全体)：(直接経費) 13,600,000円

研究成果の概要(和文)：本研究は、「人間の翻訳者のような機械翻訳を作るためには、どうすればよいだろうか？」という問いに回答を与えるための手段と言語資源を提供してきた。人間翻訳者の訳出プロセスを可視化・データ化し、機械翻訳の品質向上に寄与することを目指し、研究用ツール「Trados-to-Translog-II」を開発した。また翻訳者の作業データ収集も行った。同ツールおよびデータは一般公開し、訳出プロセスに関する研究の基盤確立に貢献した。また同時通訳や機械翻訳+ポストエディットのプロセスデータ収集・分析や、外国語能力と翻訳の違いを考慮したトピックも扱い、国際会議でのワークショップも開催。学際的研究の発展を促した。

研究成果の学術的意義や社会的意義

本研究結果である翻訳プロセスの可視化・データ公開により、翻訳通訳学における研究コミュニティの拡大に大きく貢献した。また、AIなどによる自然言語処理の向上と測定方法の確立にも貢献した。さらに言語学・翻訳研究・認知科学等への学際的研究の加速度的促進への貢献が期待される。具体的な学術・社会的意義は以下の通り。人間翻訳プロセスの理解向上、および機械翻訳や大規模言語モデルを含む自然言語処理の品質の改善と測定方法の構築である。これらの事柄は人間(および人工知能)の言語処理プロセスの解明、という基礎研究の発展と、応用分野では翻訳者教育・トレーニング、および英語教育・外国語教育の発展と進化にも影響を与えた。

研究成果の概要(英文)：This research aimed to answer the question, "How can we create machine translation systems that translate like human translators?" It provided means and language resources for this purpose. We focused on visualizing and digitizing the translation process of human translators to contribute to the improvement of machine translation quality. To this end, we developed a research tool called "Trados-to-Translog-II" and also collected data on the work of translators. Both the tool and data have been made available to the public. This work has contributed to establishing a solid foundation for research on the translation process. Furthermore, we tackled topics considering the difference between foreign language skills and translation. We also collected and analyzed process data on simultaneous interpretation and, machine translation + post-editing. In addition, we hosted workshops at international conferences, promoting interdisciplinary research development.

研究分野：翻訳通訳学

キーワード：翻訳通訳学 翻訳プロセス 機械翻訳 言語処理プロセス 順送り 通訳プロセス 翻訳品質

1. 研究開始当初の背景

本研究は、「人間の翻訳者のような機械翻訳を作るためには、どうすればよいだろうか?」という問いに答えを出す目的で始まった。その背景には、機械翻訳の品質精度の不足があり、人間の翻訳者の能力との間にギャップがあったので、その差分の記述をするためにも、人間の翻訳プロセスの解明が必要であると考えられていた。人間が翻訳者になるためには、言語能力だけでなく専門能力が必要であり、これらは大部分が、「熟練知」や「経験知」という形で培われてきた。しかしながら、これらの能力を明示的に記述し共有することは困難であった。

本研究の特長は、人間の翻訳者たちの訳出プロセスを、テキスト情報だけでなく、身体反応も含めて記録し、それを可視化・データ化することにある。そして、このデータを言語資源として提供し、学際研究の共有基盤を確立することを目指した。

2. 研究の目的

本研究の目指すところは、上述の問題への具体的な解を提供し、それに必要なツールと言語リソースを開発することである。次世代の機械翻訳を現実化するには、人間の翻訳者のスキルを機械が学習可能な形式で視覚化・操作化するデータ化が必要となるだろう。しかしながら、人間の翻訳能力は「経験に基づく知識」として身につけられているため、これを外部のデータとして表現することは難しかった。

そこで、本研究の一つ目の目的は、プロの翻訳者が「何をやっているか」(翻訳プロセス)を記録し、視覚化することである。具体的には、プロの翻訳者が原文をどのように読み、どの辞書を参照し、何を検索し、どの候補から訳語を選んでいくというプロセスを記録し、データ化することである。これにより、「経験知」が「翻訳プロセス」として視覚化され、熟練翻訳者の「操作」を分析可能にする。

事実、翻訳プロセス研究という分野では、人間の翻訳プロセスの研究がすでに行われてきた。しかし、本研究が従来の研究と異なる点はプロの翻訳者が制約のない実際の作業環境で翻訳を行う様子をデータ収集することである。本研究では、収集したデータと記録ツール(ソフトウェア)を研究者に無料で提供する。

これらを通じて、本研究のもう一つの目指すところは、翻訳研究や自然言語処理、言語学などの学術的研究(次世代の機械翻訳開発も含む)のための基盤研究の共有プラットフォームを構築することである。これらの目標は、翻訳と言語学の研究者だけでなく、人間の翻訳プロセスを理解し、改善するための新しいツールを開発する技術者にも利益をもたらす。最終的には、翻訳プロセスを理解し、人間の経験知識をデータとして捉えることで、人間と機械の間のギャップを埋めることである。

3. 研究の方法

実験環境の信頼性：過去の翻訳プロセス研究では、Translog-II というツールを使って翻訳者が原文を読むときの視線動向(アイトラッキング)や訳文入力時のキーボードロギングを含む詳細なデータを記録していた。この研究でも同様のツールを使用した。Translog-II の限界は、翻訳者に人工的な実験環境を強制することで、翻訳者は自由に翻訳支援ツールを使ったり、インターネットで調査をしたりすることができず、分析は限られたデータから推測されることが多くなっていた。この研究では、これらの問題を解決し、プロの翻訳者の真実味のある作業環境を再現できるツールを開発した。

実際の作業環境を再現し、現役のプロ翻訳者の翻訳作業を記録し、視覚化することは、これまでになかった新たなアプローチであった。これにより、翻訳プロセス研究は、人工的な環境から現実に近い環境へと進化し、この分野における独自性が大きく向上する。

学際性：自然言語処理技術の進歩に伴い、人間が機械翻訳を使用して後編集を行うといった、人間と AI や機械翻訳との相互作用の調査が重要となっている。インターネット検索は翻訳作業に不可欠であり、機械翻訳や翻訳支援ツールの利用も産業翻訳で急速に増えている。その結果、翻訳作業は人間とコンピュータの相互作用(HCI)とも見なすことができる。つまり、機械翻訳がまだ不可能な部分を、人間の翻訳者が補完しているということだ。そのため、この「訳出プロセスデータ」を収集すれば、プロ翻訳者の経験知が集積された言語リソースとなり、また、人間翻訳者のような機械翻訳を開発するための手がかりも得られる。このようにして、本研究は翻訳研究と自然言語処理などの間で共通の関心事を持つことが可能となり、これは密度の高い学際的な共同作業を実現するための創造性を持つことを意味する。

4. 研究成果

3年間の研究期間を通じて、2021年度にデータ収集用のツール「Trados-to-Translog-II」を完成させ、公開した。このツールは既存の Translog-II の提供元である CRITT TPR-DB のサイトで公開した(下記参照)。

TRADOS-TO-TRANSLOG-II: ADDING GAZE AND QUALITIVITY DATA TO THE CRITT TPR-DB

Updated: May 23, 2022.

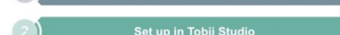


This describes a new tool in the TPR-DB that converts **Trados Studio** keylogging data (**Quality**) into **Translog-II** format and adds the converted data to the CRITT TPR-DB. The tool is also able to synchronize with the output of various eye-trackers. The Trados-TPR-DB interface provides the possibility to record translation behavior in an ecologically realistic translation environment. We are now able to investigate patterns of reading and typing activities in a widely and professionally used CAT tool, and thus to achieve a better understanding of factors that impact professional translation activity.

For using the Trados to Translog conversion tool, you need to have the platform and software listed below. The procedures for adding gaze and Quality data to the TPR-DB are the following six steps.

Required platform and software

-  TPR-DB account & management tool-Check CRITT website.
-  Trados Studio 2019 & Qualityity 2019- Check Trados Appstore.
-  Tobii Eye tracker TX 300 & Tobii Studio 3.3.2- Check instructions from Tobii.

Procedures

-  1 Download and install Qualityity
-  2 Set up in Tobii Studio
-  3 Conduct Post-editing experiment in Trados
-  4 Export Qualityity output
-  5 Export eye tracker output
-  6 Upload eye tracking and keylogging data to TPR-DB

データ収集については、このツールを使用して、英日のプロ翻訳者 8 名と英中の学生翻訳者が参加した実験を行った。これは実務翻訳環境を模倣したもので、作業指示書やポストエディットの手法を指定するなどの条件を制御した環境下で、機械翻訳とポストエディットの実験を行った。この実験は米国ケント州立大学との共同研究として実施し、結果は国際会議で発表した。

さらに、翻訳者だけでなく、訳文を読む（読者の読み）プロセスのデータも収集した。これにより、翻訳者中心のプロセスから、翻訳の受け手の視点からのデータ収集を実施し、その結果を複数の国際会議で発表した。また、同時通訳のプロセスデータの収集と分析も行った。

また、訳文の読みのプロセスにおける「困難さ (difficulties)」の詳細な分析を進展させ、外国語教育とプロ翻訳の違いを考慮したトピックも考察した。

新型コロナウイルスの影響で研究初期にはツール開発やデータ収集が難しくなるなどの問題があったが、中盤以降はこれを挽回し、研究成果を発表することができた。その成果の一つ、岡村・山田 (2020) の『順送り訳』の規範と模範：同時通訳を模範とした教育論の試論』は日本通訳翻訳学会新崎賞（奨励賞）を受賞した。2022 年度（最終年度）には、国際学会での発表や翻訳プロセス研究に特化したワークショップを開催した。

このようにして、本研究の目的は、特定の問題の解決よりも、データ収集のためのツール開発とそのツールを使っての研究、そしてそれを通じて翻訳研究や自然言語処理、言語学などの学術的研究（次世代の機械翻訳開発も含む）のための基礎研究の共有基盤を構築した。この目的は、3 年間の研究期間を通じて達成できたと言える。

今後の展望としては、開発したツールと収集したデータを更に広範囲での研究に活用し、外国語教育や翻訳教育への貢献を通じて、翻訳研究全体の発展を目指す。さらに、開発したツールを翻訳研究者コミュニティと共有し、多様な視点とアプローチからの新たな研究の展開を促すことも重要な目標とする。

5. 主な発表論文等

〔雑誌論文〕 計7件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 6件）

1. 著者名 中澤敏明	4. 巻 74
2. 論文標題 書評・解説「自動翻訳大全：終わらない英語の仕事が5分で片づく超英語術」	5. 発行年 2021年
3. 雑誌名 AAMT Journal	6. 最初と最後の頁 23-26
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 山田優	4. 巻 74
2. 論文標題 書評・解説「機械翻訳はどこへ行くのか？『機械翻訳：歴史・技術・産業』」	5. 発行年 2021年
3. 雑誌名 AAMT Journal	6. 最初と最後の頁 27-32
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 山田優	4. 巻 24
2. 論文標題 ポストエディットと持続可能な翻訳の未来	5. 発行年 2021年
3. 雑誌名 関西大学 外国語学部紀要	6. 最初と最後の頁 83-105
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 岡村ゆうき, 山田優	4. 巻 2
2. 論文標題 「順送り訳」の規範と模範：同時通訳を模範とした教育論の試論	5. 発行年 2020年
3. 雑誌名 MITIS Journal	6. 最初と最後の頁 25-48
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 渡邊里菜, 山田優	4. 巻 2
2. 論文標題 英日字幕翻訳のコーパスベース研究 プロ字幕とファンサプの比較分析	5. 発行年 2020年
3. 雑誌名 MITIS Journal	6. 最初と最後の頁 1-23
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 松下佳世	4. 巻 20
2. 論文標題 コロナ禍における遠隔通訳の実施状況調査	5. 発行年 2020年
3. 雑誌名 通訳翻訳研究	6. 最初と最後の頁 125-146
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 石塚浩之	4. 巻 2月
2. 論文標題 通訳翻訳研究の世界 第13回 通訳に関わる二つのレベル?! : 相関モデル	5. 発行年 2021年
3. 雑誌名 通訳翻訳ジャーナル	6. 最初と最後の頁 127-128
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計8件 (うち招待講演 3件 / うち国際学会 5件)

1. 発表者名 山田優
2. 発表標題 機械翻訳の進歩に接して外国語の教師が心得ておくべきこと
3. 学会等名 青山学院大学附置外国語ラボラトリー主催 公開セミナー (招待講演)
4. 発表年 2022年

1. 発表者名 山田優
2. 発表標題 機械翻訳をめぐる諸問題
3. 学会等名 日本英語表現学会 第50回全国大会（招待講演）
4. 発表年 2021年

1. 発表者名 Kayo Matsushita, Masaru Yamada
2. 発表標題 Revisiting Daniel Gile's "Effort Models": How complex multi-tasking affects the interpreters' cognitive effort in remote simultaneous interpreting (RSI)
3. 学会等名 IATIS 7th International Conference (国際学会)
4. 発表年 2021年

1. 発表者名 Masaru Yamada, Akiko Sakamoto, Toshiaki Nakazawa, Yusuke Hiraoka, Nanami Onishi
2. 発表標題 Process data supports a sustainable future for translators integrating machine translation
3. 学会等名 IATIS 7th International Conference (国際学会)
4. 発表年 2021年

1. 発表者名 Takanori Mizowaki, Masaru Yamada, Haruka Ogawa, Yuki Okamura, Yusuke Hiraoka
2. 発表標題 Syntactic cross and reading effort
3. 学会等名 ALAPP 2021 Conference (国際学会)
4. 発表年 2021年

1. 発表者名 Masaru Yamada
2. 発表標題 Post-editing and a sustainable future for translators
3. 学会等名 Translation in Transition (TT5) (招待講演) (国際学会)
4. 発表年 2020年

1. 発表者名 Kayo Matsushita
2. 発表標題 Can remote simultaneous interpreting be a game changer in the Japanese interpreting industry?
3. 学会等名 Tenth IATIS Regional Conference 2020 (国際学会)
4. 発表年 2020年

1. 発表者名 長沼美香子
2. 発表標題 情報型字幕における結束性と順送り訳
3. 学会等名 JAITS研究プロジェクト 第11回「順送りの訳」研究会
4. 発表年 2021年

〔図書〕 計3件

1. 著者名 Michael Carl	4. 発行年 2021年
2. 出版社 Springer	5. 総ページ数 726
3. 書名 Explorations in Empirical Translation Process Research	

1. 著者名 山田優、坂西優	4. 発行年 2020年
2. 出版社 三オブックス	5. 総ページ数 288
3. 書名 自動翻訳大全	

1. 著者名 ティエリー・ポイボー、高橋 聡、中澤 敏明	4. 発行年 2020年
2. 出版社 森北出版	5. 総ページ数 192
3. 書名 機械翻訳	

〔産業財産権〕

〔その他〕

CRITT TPR-DB https://sites.google.com/site/centrere translationinnovation/tpr-db

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	長沼 美香子 (Naganuma Mikako) (80460012)	神戸市外国語大学・外国語学部・教授 (24501)	

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	松下 佳世 (Matsushita Kayo) (90746679)	立教大学・異文化コミュニケーション学部・教授 (32686)	
研究分担者	石塚 浩之 (Ishizuka Hiroyuki) (40737003)	広島修道大学・人文学部・教授 (35404)	
研究分担者	中山 英樹 (Nakayama Hideki) (00643305)	東京大学・大学院情報理工学系研究科・准教授 (12601)	
研究分担者	中澤 敏明 (Nakazawa Toshiaki) (10828552)	東京大学・大学院情報理工学系研究科・客員研究員 (12601)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計1件

国際研究集会 Tenth IATIS Regional Conference	開催年 2020年～2020年
---	--------------------

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関