

科学研究費助成事業 研究成果報告書

令和 2 年 6 月 15 日現在

機関番号：32686

研究種目：基盤研究(B) (一般)

研究期間：2016～2019

課題番号：16H02915

研究課題名(和文) 記者会見通訳の二言語並行コーパスの構築と応用研究

研究課題名(英文) Construction of a Japanese-English parallel corpora of interpreter-mediated press conferences and applied studies

研究代表者

松下 佳世 (MATSUSHITA, Kayo)

立教大学・異文化コミュニケーション学部・准教授

研究者番号：90746679

交付決定額(研究期間全体)：(直接経費) 13,700,000円

研究成果の概要(和文)：本研究は、我が国の通訳翻訳研究の活性化を目指して、研究者が広く利用可能な日英の通訳の対訳コーパスを構築することを目的に実施された。日本には翻訳された単一言語コーパスや外国語学習者向けの英語コーパスは多数存在するが、原文と訳文が対応した「通訳」の対訳コーパスに関しては、ほぼ皆無であった。中でも、自然発話、すなわちプロ通訳者による実現場でのパフォーマンスデータを元にしたコーパスは存在しなかった。このため、本研究では、日本記者クラブ等による実際の記者会見における原発話と通訳者の訳出を、音声とその波形、文字情報を組み合わせる形でデータベース化した、動的で大規模な対訳コーパスを構築し、公開した。

研究成果の学術的意義や社会的意義

これまで、欧米に比べて日本の通訳翻訳研究、特に科学的・実証的研究が十分に発展してこなかった一因には、日本語と外国語の対訳を備えた検証可能なデータの不足があった。中でも通訳研究には、自然発話を元にした言語資源が不可欠であるが、現存する対訳コーパスは実験的環境下の小規模なものにとどまることが、研究上の大きな障害となってきた。このたび、大規模な日英通訳コーパスである「通訳データベース(JNPCコーパス)」を構築したことで、上記の障害を克服し、通訳翻訳研究の発展に寄与することができた。すでに同コーパスの多言語化など複数の後継プロジェクトも始動しており、応用研究を含めた今後の発展が期待される。

研究成果の概要(英文)：The purpose of this study was to construct a parallel corpus between Japanese and English that can be widely used by researchers of Translation and Interpreting Studies (TIS) in order to promote TIS in Japan. Although large-scale monolingual corpora of translated outputs as well as English corpora for foreign language learners already exist in Japan, interpreting corpora based on authentic data have been almost non-existent. In particular, there was no corpus utilizing natural speech based on performances by professional interpreters in real-life settings. Therefore, we constructed and publicly released a large-scale, dynamic corpus by combining video, audio, and textual information obtained from video recordings of interpreter-mediated press conferences held at the Japan National Press Club (JNPC).

研究分野：図書館情報学・人文社会情報学、通訳翻訳研究

キーワード：通訳 翻訳 コーパス 英語 記者会見

1. 研究開始当初の背景

音声言語のコミュニケーションである通訳の訳出表現を対象とする研究は、書記言語を扱う翻訳と比べて十分に発展してきたとは言い難い。これまで行われてきた通訳研究の視点は、通訳の歴史、役割、規範、産業形態などが中心で、通訳の訳出を実証的手法で論じる研究はほとんどなかった。実証的な通訳研究がなかなか進まなかった主たる原因は、分析可能なデータの入手困難にある。それでも、海外においては、実務環境下の大規模な通訳コーパスとして欧州議会の通訳コーパス (EPIC: European Parliament Interpreting Corpus) が構築され、広く研究に用いられてきたが、日本では、実験環境下で作成した名古屋大学統合音響情報研究拠点 (CIAIR) の「同時通訳データベース」が存在するにとどまっていた。

通訳記録のデータ入手が困難とされてきた理由は、主に以下の3点に集約できる。

(1) **研究分野の状況** そもそも世界で通訳が実証的な研究対象と見なされるようになったのは比較的最近のことである。欧米では、1940年代から大学における通訳教育も始まっていたが、当初は実務者の養成が中心で、実証的な通訳研究の必要性が強く意識されるようになったのは1980年代末のことであった。日本では、90年代から通訳研究が始まっていたものの、通訳研究を扱う学会が設立されたのは2000年と比較的最近のことである。

(2) **技術的制約** 通訳の訳出は、録音しない限り残らない。仮に録音が存在する場合でも、文字として書き起こさない限り、詳細な分析はできない。通訳の録音は難しいことではないが、人手による音声言語の書き起こし作業には相当の時間と労力を要する。加えて、通訳記録の書き起こしには、二言語の高い語学力に加え、研究を目的とする書き起こしに関する知識と技能が必要であるため、作業を担当できる人材も限られる。

(3) **素材の入手** 使用可能な通訳記録の入手自体にも課題がある。実務で通訳が使用される状況は、機密性の高い場面が多いうえ、権利上の問題から、研究に使用できる素材は限られている。また、権利上、使用可能な素材が得られたとしても、訳出品質、記録品質に問題がなく、研究目的に合致したデータとなると数が限られる。

本研究が実施されるに至った背景には、こうした状況に変化が生まれたことが挙げられる。具体的には、以下の通りである。

(1) 国内外で通訳翻訳研究に関するコミュニティが形成され、大学院における通訳翻訳研究の体制が整い、実証研究の必要性が高まった。

(2) 言語処理技術が飛躍的に発展したため、音声認識技術の活用により通訳記録の書き起こし作業の工数を大幅に削減できるようになった。加えて、音声波形と音声認識の結果を対応させ、同時通訳の原発話と訳出の時間的対応づけを実現することも技術的に可能となった。

(3) 本研究で利用した公益社団法人・日本記者クラブ (JNPC) の通訳付きの記者会見映像を始め、公開されている通訳付きの素材が増えた。事前に映像等の使用許可を得ることで、品質的に問題のない素材を大量に入手できる環境が整った。

2. 研究の目的

本研究の目的は、実務環境下での通訳 (同時・逐次) の大規模データベースを構築し、訳出データを必要とする通訳翻訳分野の幅広い研究に共通の研究資源を提供することにある。そのため、JNPC の通訳付き記者会見を使用し、この動画から音声波形データを作成し、これにテキストデータを対応させてデータベース化した。これを特定非営利活動法人・言語資源協会 (GSK) を通じて、研究者・教育者を対象に一般公開した (「通訳データベース (通称: JNPC コーパス)」 <https://www.gsk.or.jp/catalog/gsk2020-a/>)。今後、本研究のプロジェクトメンバーだけでなく、多くの研究者がこのコーパスを利用した研究を進めることで、通訳翻訳研究の学際的な広がりが期待できる。特に、異なる手法を得意とする通訳研究者、翻訳研究者が互いに知見を取り入れることで、より広範な研究に使用可能な汎用型データベースの構築も考えられる。応用研究の範囲は通訳翻訳研究および言語研究にとどまらず、教育現場での活用も可能である。

3. 研究の方法

上述の通り、本研究では、通訳音声を書き起こしてコーパスを構築した。構築方法は、JNPC が一般公開している動画から、原発話と通訳を複数の音声認識ソフト (speech to text) にかけて自動生成し、その後、機械の誤認識によるエラーを、人手によって修正した。完成したコーパスには、その文字情報 (トランスクリプト) の他、オリジナルの動画、音声が含まれている。これらは「ELAN」というフリーソフトウェアで視聴・閲覧できる。書き起こしに関しては、英語の場合は単語ごとに、日本語の場合は、わかち書きしない形態素ごとに記載し、それぞれに発話時間を付与した。例えば、“Thank you very much.”であれば、4単語それぞれについて発話時間 (開始と終了の時間) を記録した。また、発話者や通訳者の言い淀み (hesitation) についても、それを区別できるアノテーションを付して、できる限り記載した。これらのトランスクリプトデータは ELAN から任意のフォーマットで書き出し可能であり、利用者の目的に合わせて修正・調整が可能である。



図 1: ELAN 画面のコーパスを開いた様子

4. 研究成果

本研究の最大の成果は、JNPC コーパスの構築と公開である。このデータベースには、2010 年から 2017 年の間に日本記者クラブで行われた記者会見のうち、79 件分のデータが収録されている。内訳は、同時通訳付きのものが 71 件、逐次通訳付きのものが 8 件である。会見は最短 17 分 35 秒、最長 2 時間 7 分 12 秒で、平均時間は 58 分 21 秒である。会見は冒頭のスピーチ部分と、その後の質疑応答からなり、スピーチ部分は英日または日英一方の訳出のみ、質疑応答部分は概ね双方向の訳出となっている。

尚、JNPC コーパスは現状、日本語と英語の二言語並行コーパスだが、2020 年度からは後継プロジェクトとして、日本語と中国語、日本語とスペイン語のサブコーパスの構築が決まっている。こうした多言語化により、国際的にも貴重な研究資源が提供できるものと考えている。

本研究期間中には、通訳コーパスの構築と公開だけでなく、国内外のワークショップやシンポジウムを通じて、データの構築方法、研究利用や教育への応用などに関する知見を積極的に発信した。これらの活動を通じて、上述の多言語展開や学際的共同研究の機会が生まれたことも、大きな成果と言えよう。

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 1件/うち国際共著 2件/うちオープンアクセス 4件）

1. 著者名 山田優、松下佳世、石塚浩之、歳岡冴香、Michael Carl	4. 巻 -
2. 論文標題 記者会見通訳の二言語並行コーパスの構築	5. 発行年 2017年
3. 雑誌名 言語処理学会第23回年次大会発表論文集	6. 最初と最後の頁 1168-1171
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 該当する
1. 著者名 石塚浩之	4. 巻 第19号
2. 論文標題 サイトラにおける認知プロセス-分割・保持・組換え	5. 発行年 2018年
3. 雑誌名 通訳翻訳研究への招待	6. 最初と最後の頁 69-89
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 Kayo Matsushita	4. 巻 -
2. 論文標題 A new approach to self-practice and self-assessment: Testing the effectiveness of using a parallel corpus in interpreter training	5. 発行年 2018年
3. 雑誌名 Conference Proceedings: Ewha GSTI International Conference 2018	6. 最初と最後の頁 39-43
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 -
1. 著者名 山田優、松下佳世、石塚浩之、歳岡冴香、Michael Carl	4. 巻 -
2. 論文標題 記者会見通訳の二言語並行コーパスの構築 第二報	5. 発行年 2018年
3. 雑誌名 言語処理学会第24回年次大会発表論文集	6. 最初と最後の頁 734-737
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 該当する

1. 著者名 松下佳世、山田優、石塚浩之	4. 巻 第22号
2. 論文標題 英日・日英通訳データベース（JNPCコーパス）の概要	5. 発行年 2020年
3. 雑誌名 通訳翻訳研究への招待	6. 最初と最後の頁 -
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 -

〔学会発表〕 計21件（うち招待講演 2件 / うち国際学会 15件）

1. 発表者名 Kayo Matsushita
2. 発表標題 The influence of risk on interpreter performance: A corpus-based analysis
3. 学会等名 4th International Conference on Cognitive Research on Translation, Interpreting and Language Acquisition (国際学会)
4. 発表年 2017年

1. 発表者名 Michael Carl & Masaru Yamada
2. 発表標題 Multimodal integration of written and spoken translation production
3. 学会等名 4th International Conference on Cognitive Research on Translation, Interpreting and Language Acquisition (国際学会)
4. 発表年 2017年

1. 発表者名 山田優、松下佳世、石塚浩之、歳岡冴香、Michael Carl
2. 発表標題 記者会見通訳の二言語並行コーパスの構築
3. 学会等名 言語処理学会第23回年次大会（国際学会）
4. 発表年 2017年

1. 発表者名 Kayo Matsushita
2. 発表標題 Corpus-based research utilizing interpreter-mediated press conferences in Japan
3. 学会等名 8th Annual International Translation Conference Program (国際学会)
4. 発表年 2017年

1. 発表者名 石塚浩之
2. 発表標題 同時通訳に見る順送り訳の実態-JNPCギラード豪首相記者会見通訳より
3. 学会等名 日本通訳翻訳学会「順送りの訳」研究プロジェクト研究会
4. 発表年 2018年

1. 発表者名 Kayo Matsushita
2. 発表標題 Non-professional translators' dilemma in emerging media: A Japanese case study
3. 学会等名 4th International Conference on Non-Professional Interpreting and Translation (国際学会)
4. 発表年 2018年

1. 発表者名 Kayo Matsushita
2. 発表標題 A new approach to self-practice and self-assessment: Testing the effectiveness of using a parallel corpus in interpreter training
3. 学会等名 Ewha GSTI International Conference 2018 (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 松下佳世・山田優
2. 発表標題 プロの通訳者による実現場でのパフォーマンス・データを活用した通訳コーパスの構築とその応用可能性
3. 学会等名 日本通訳翻訳学会第19回年次大会（国際学会）
4. 発表年 2018年

1. 発表者名 松下佳世・山田優
2. 発表標題 英日通訳コーパスの活用法ワークショップ
3. 学会等名 日本通訳翻訳学会関西支部第49回例会
4. 発表年 2018年

1. 発表者名 山田優・長沼美香子
2. 発表標題 英日サイト・トランスレーションのプロセスに関する実証的研究と今後の展望
3. 学会等名 日本通訳翻訳学会「順送りの訳」研究プロジェクト研究会
4. 発表年 2018年

1. 発表者名 石塚浩之、稲生衣代、辰己明子、長沼美香子、畑上雅朗、山田優
2. 発表標題 サイトラ研究プロジェクト 活動報告
3. 学会等名 日本通訳翻訳学会第19回年次大会（国際学会）
4. 発表年 2018年

1. 発表者名 Michael Carl, Masaru Yamada, Silvia Hansen-Schirra, Moritz Schaeffer, Kayo Matsushita
2. 発表標題 Integrating written and spoken translation production in the CRITT TPR-DB
3. 学会等名 The International Association for Translation and Intercultural Studies (IATIS) 6th International Conference Workshop (国際学会)
4. 発表年 2018年

1. 発表者名 山田優、松下佳世、石塚浩之、歳岡冴香、Carl Michael
2. 発表標題 記者会見通訳の二言語並行コーパスの構築 第2報
3. 学会等名 言語処理学会第24回年次大会 (国際学会)
4. 発表年 2018年

1. 発表者名 Kayo Matsushita
2. 発表標題 Exploring corpus-based interpreting studies in a non-European context: A Japanese case study
3. 学会等名 ATISA IX Biennial Conference of the American Translation and Interpreting Studies Association (国際学会)
4. 発表年 2018年

1. 発表者名 Kayo Matsushita
2. 発表標題 Interpreters as newsmakers: How press conference interpreter's renderings affect news reports
3. 学会等名 The 7th Asia-Pacific Forum on Translation and Intercultural Studies (APFTIS) Conference (国際学会)
4. 発表年 2019年

1. 発表者名 Kayo Matsushita
2. 発表標題 Adjusting levels of apology to manage risk: A corpus-based analysis of the interpreters' performance from the Japan National Press Club Corpus
3. 学会等名 The 3rd East Asian Translation Studies Conference (国際学会)
4. 発表年 2019年

1. 発表者名 松下佳世・山田優
2. 発表標題 特別講演: 同時通訳に特有の難しさを乗り越えるには: 通訳コーパス構築プロジェクトの経験から
3. 学会等名 次世代音声言語研究シンポジウム2019 (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Kayo Matsushita
2. 発表標題 Moving beyond English: Challenges and opportunities in expanding a Japanese-English parallel corpus to include Chinese and Spanish
3. 学会等名 The 9th Congress of the European Society for Translation Studies (国際学会)
4. 発表年 2019年

1. 発表者名 岡村ゆうき・山田優
2. 発表標題 順送りと逆送りの測定方法の中間報告
3. 学会等名 日本通訳翻訳学会研究プロジェクト 第5回順送り研究会
4. 発表年 2019年

1. 発表者名 石塚浩之
2. 発表標題 同時通訳の訳出における指示表現の追加と順送りの訳出
3. 学会等名 日本通訳翻訳学会研究プロジェクト 第5回順送り研究会
4. 発表年 2019年

1. 発表者名 岡村ゆうき・山田優
2. 発表標題 順送りと逆送りの測定方法(2)：通訳データ編
3. 学会等名 日本通訳翻訳学会研究プロジェクト 第6回順送り研究会
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

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	山田 優 (YAMADA Masaru) (70645001)	関西大学・外国語学部・教授 (34416)	
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	船山 仲他 (FUNAYAMA Chuta) (10199416)	神戸市外国語大学・外国学研究所・名誉教授 (24501)	

6. 研究組織（つづき）

	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	染谷 泰正 (SOMEYA Yasumasa) (40348454)	関西大学・外国語学部・教授 (34416)	
研究 分担者	歳岡 冴香 (TOSHIOKA Saeka) (40708468)	近畿大学・文芸学部・講師 (34419)	
研究 分担者	水野 的 (MIZUNO Akira) (90350321)	青山学院大学・文学部・教授 (32601)	