様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

# 科学研究費助成事業　　研究成果報告書

令和　５ 年　 ５ 月 ２３ 日現在

機関番号：３２６８６
研究種目：基盤研究(C)（一般）
研究期間：2020～2022
課題番号：２０Ｋ１２５６３

研究課題名（和文）記者会見通訳データを用いた日中・日西並行コーパスの構築と応用研究

研究課題名（英文）Compilation of Japanese-Chinese and Japanese-Spanish parallel corpora using press conference interpretation data and applied research

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・教授

研究者番号：９０７４６６７９

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：本研究は、令和2年度から令和4年度にかけて実施され、すでに構築済みの「通訳データベース（JNPCコーパス）」の対象言語を広げる形で新たに日本語と中国語、日本語とスペイン語の間の通訳の対訳コーパスを加えた。日中10件（同時通訳8件、逐次通訳2件）、日西11件（同時通訳6件、逐次通訳5件）が収録された。同コーパスは、「通訳データベース（JNPCコーパス）日中・日西サブコーパス」として、特定非営利活動法人 言語資源協会（GSK)を通じて令和5年4月に公開された。また、同コーパスを用いて日英、日中、日西の訳出を比較した研究成果の一部を国内外で口頭発表し、日本通訳翻訳学会の学会誌にも投稿した。

研究成果の学術的意義や社会的意義
本研究は、我が国の通訳研究の活性化ならびに、将来的なAI（人工知能）による自動通訳の精度向上を目的として、すでに構築済みの日本語と英語の間の大規模な通訳コーパス（JNPCコーパス）の対象言語を広げ、新たに日本語と中国語、日本語とスペイン語の間の通訳の対訳コーパスを構築し、サブコーパスとして加えるものである。構築したサブコーパスは、英中西の各言語を専門とする通訳研究者が応用研究に用いるほか、通訳の実技練習の教材や、自主学習用の素材として活用する。

研究成果の概要（英文）：This study was conducted from FY2020 to FY2022 to expand the target languages of the JNPC Corpus by adding new bilingual corpora for interpretation between Japanese and Chinese, and between Japanese and Spanish. Ten Japanese-Chinese (8 simultaneous interpretation and 2 consecutive interpretation) and 11 Japanese-Spanish (6 simultaneous interpretation and 5 consecutive interpretation) datasets were included. This corpus was released as the "Japanese-Chinese and Japanese-Spanish Sub-Corpora of the Interpretation Database (JNPC Corpus)" in April 2023 by Gengo-Shigen-Kyokai (GSK) (literal meaning “Language Resources Association”), a non-profit organization. Some of the results of the preliminary research comparing Japanese-English, Japanese-Chinese, and Japanese-Spanish interpretations using the corpora were presented both domestically and internationally, and a paper was published in the journal Interpreting and Translation Studies, No.22.

研究分野：Translation and Interpreting Studies

キーワード：通訳　コーパス　記者会見　中国語　スペイン語

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

## １．研究開始当初の背景

本研究は、我が国の通訳研究の活性化ならびに、将来的な AI（人工知能）による自動通訳の精度向上を目的として、日本語と中国語、日本語とスペイン語の間の通訳の対訳コーパスを構築するために、令和 2 年度から 4 年度にかけて実施された。具体的には、日本記者クラブで定期的に行われている通訳付きの記者会見における原発話と通訳者の訳出、および記者会見映像をもとに新たに録音した日中、日西通訳者の訳出を、映像、音声、音声波形、文字情報を組み合わせた形でデータベース化した。

本研究を実施した背景には、日本語と外国語の間の通訳行為を研究するための言語資源の不足があった。翻訳研究においては 1990 年代からコーパスを用いた研究が盛んに行われており、日本にも小規模なものを含めれば多数の対訳コーパスが存在する。こうした言語資源を活用する形で、近年我が国でも機械翻訳の研究が積極的に進められてきた。近年では、Google が導入して話題となった AI によるニューラル機械翻訳（NMT：Neural Machine Translation）の登場により、言語的な特徴が大きく異なる日本語と英語の間でも機械翻訳の精度が飛躍的な高まりを見せた。NMT の精度を高めるためには、学習用の対訳コーパスが大量に必要であり、これは AI 技術を用いた自動通訳にも当てはまる。さらに、同時通訳においては、話者の発言（情報の入力）と通訳者の訳出（情報の出力）をほぼ同時的に行うために、話の先を予測（anticipation）したり、重複部分を省略（omission）したりといった様々な操作が行われており、人間の通訳者が実際にどのように同時性を生み出しているのかを解明し、AI に学ばせるには、自然発話をもとにした大規模なコーパスが不可欠となる。

このような現状を踏まえ、研究代表者・松下と研究分担者 6 名からなる研究チームは、平成 28 年度から「記者会見通訳の二言語並行コーパスの構築と応用研究」（16H02915）を実施し、日本記者クラブが YouTube に通訳音声付きで公開している映像素材を用いた日英通訳コーパスを構築した。このコーパスは「通訳データベース（JNPC コーパス）」（https://www.gsk.or.jp/catalog/gsk2020-a/）として、令和 2 年 4 月から特定非営利活動法人・言語資源協会（GSK）を通じて研究・教育用に公開されている。しかし、言語資源が不足しているのは日英の言語ペアのみではない。世界で最も母語話者が多く、日本にも多数が暮らす中国語を見ても、研究者が自らの研究のために独力で集めたもの以外の一定の規模を持つ日中通訳コーパスは本研究を始めた段階では存在していなかった。世界で 2 番目に母語話者が多いスペイン語の場合も、上記の規模の日西通訳コーパスは存在せず、東京外国語大学のスペイン語学ゼミが約 20 年前に調査した日本語の擬声語・擬態語（オノマトペ）とそれに対応するスペイン語の翻訳コーパスがあるにとどまっていた。

## ２．研究の目的

上記の背景を踏まえ、本研究では、JNPC コーパスの対象を中国語とスペイン語に広げ、新たに日中・日西の 2 種類のサブコーパスを構築し、英語以外の言語による通訳研究の発展と、将来の AI 自動通訳開発にとって欠かせない言語資源を提供することを目的とした。また、JNPC コーパスに収録されている記者会見のうち、原発話が日本語であるもの 4 本を新たに中国語とスペイン語に訳してサブコーパスに加えることで、日本語から英中西 3 言語への訳出を比較した応用研究を実施することを目指した。

## ３．研究の方法

本研究は、令和 2 年度から令和 3 年度までと、最終年度である令和 4 年度の二段階に分けて実施した。第一段階では、次段落で説明する方法でサブコーパスの構築を行った。第二段階は応用研究の実施ならびに成果の発表で、この詳細は「4．研究成果」で具体的に述べる。

本コーパスの構築にあたり、研究チームはまず平成 21 年 9 月から YouTube で公開されている日本記者クラブの記者会見ビデオのうち、中国語とスペイン語の通訳が付いているものを特定し、コーパス化すべき会見の選定を行った。会見の内容や長さに加え、スペイン語の場合は複数の国、異なる通訳者のものであること、さらに中国語の場合は経済・歴史・文学・国際関係など分野の多様性や異なるスピーカーで同じ通訳者ペアのものなど、研究用コーパスとしての有用性を考慮し、総合的に判断した。その結果、中国語で行われた 6 つの記者会見（同時通訳 4 件、逐次通訳 2 件）と、スペイン語で行われた 7 つの記者会見（同時通訳 2 件、逐次通訳 5 件）を選んだ。その上で、会見ビデオの音声を新たに文字に書き起こし、動画、音声データと共にアノテーションソフト「ELAN」にエクスポートしてコーパス化した。さらに JNPC コーパスにある記者会見のうち、日本語で行われた 4 つの会見について、プロの通訳者に依頼して日中・日西の同時通訳音声を新たに録音し、コーパスに加えた。テクストデータについては、自動音声認識（ASR）を用いて生成したのち、各言語を専門とする作業者による校正を 3 回繰り返した。完成した 2 種

類のサブコーパスは、「通訳データベース（JNPC コーパス）日中・日西サブコーパス」として、令和 5 年 4 月に GSK を通じて公開された（https://www.gsk.or.jp/catalog/gsk2023-a/）。

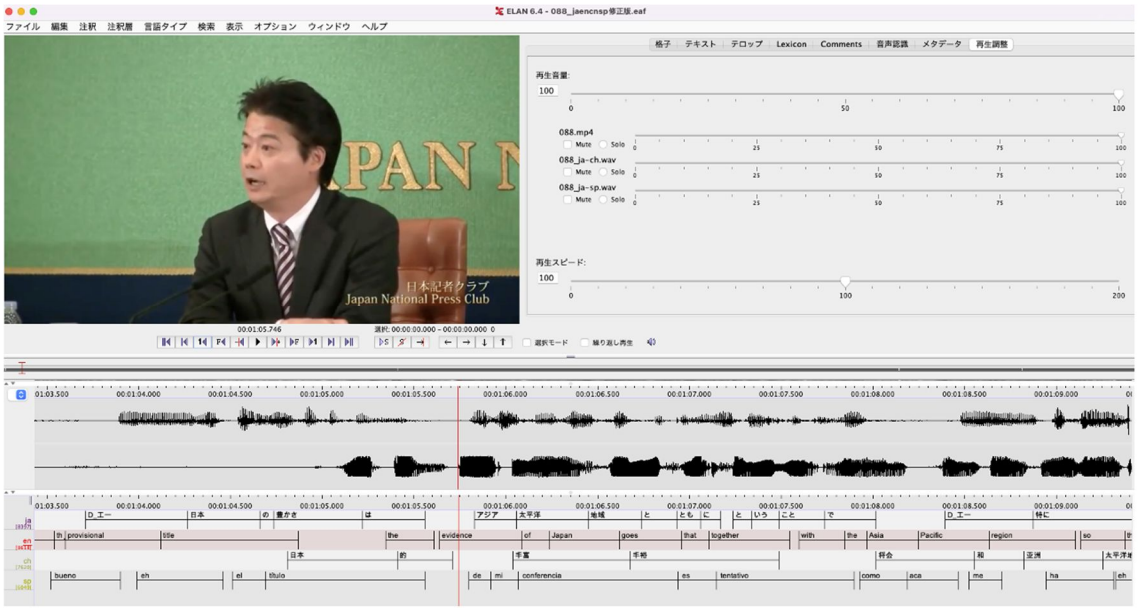


図 1．サブコーパスの収録データの一例

## 4．研究成果

上述の通り、日中・日西のサブコーパスを構築し、広く一般に公開できたことは、本研究の最大の成果である。JNPC コーパス本体と同様、GSK にサブコーパス利用の申請が出された場合は、本研究チームが審査した上で、希望者に実費相当の費用のみで配布される仕組みで、すでに複数の申請があり、研究・教育での活用が始まっている。

また研究期間の後半では、同コーパスの最終化作業と並行して、応用研究も実施した。具体的には、JNPC コーパスに収録された会見のうち、2011 年と 2014 年に行われた外務大臣の記者会見データを用いて、日英・日中・日西の訳出データを比較した。実際の会見では、2 名の日英通訳者が日英双方向に訳出を行なっていたため、同じ会見のビデオを用いて新たに収録した日中、日西通訳者による訳出を、日英のものと比較した。

最初に日英と日中の比較を行い、その成果は令和 4 年 7 月にフランス・パリで開かれた国際学会 The 4th East Asian Translation Studies Conference (EATS4)において、研究代表者・松下と研究分担者・古川が “Diverging Strategies: Key Findings from a Comparative Study of Chinese and English Interpretation Using the Japan National Press Club Interpreting Corpus” と題して共同発表した。

さらに、研究分担者・吉田を加えて、日本語から英語、中国語、スペイン語への訳出を比較し、その分析結果を令和 4 年 9 月にオンラインで実施された日本通訳翻訳学会第 23 回年次大会で「多言語通訳コーパスを活用した日英・日中・日西の訳出比較に基づく初期的考察」と題して共同発表した。この内容をまとめた論考は同年、「多言語通訳コーパスを活用した日英・日中・日西の訳出比較」のタイトルで、『通訳翻訳研究 22 巻』に掲載された（https://doi.org/10.50837/its.2205）。

研究分担者・吉田はさらに、令和 5 年 3 月にスペイン・サラマンカ大学で行った招待講演で、構築したコーパスの概要と、比較研究の内容を「多言語通訳コーパス（日・西・中・英）から見る文化仲介方略の異なりに関する考察（スペイン語の原題：Evaluación de las diferencias estratégicas de mediación cultural a través de un corpus de interpretación multilingüe japonés-español/chino/inglés）」と題して紹介した。

これらの成果に加えて、令和 3 年度には、研究代表者・松下がトルコのボアズィチ大学翻訳通訳学部との国際共同研究を通じて、日英・日中・日西の訳出データ比較に用いたのと同じ記者会見のうち一つをトルコ語に通訳したデータを録音し、日中・日西サブコーパスと同じ手法で ELAN 化した。このデータは現時点では公開されていないものの、今後 JNPC コーパスをさらに多言語化していく上で、活用する予定である。

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　0件／うちオープンアクセス　0件）

| | |
|---|---|
| １．著者名<br>松下 佳世、古川 典代、吉田 理加 | ４．巻<br>22 |
| ２．論文標題<br>多言語通訳コーパスを活用した日英 ・ 日中 ・ 日西の訳出比較 | ５．発行年<br>2022年 |
| ３．雑誌名<br>通訳翻訳研究 | ６．最初と最後の頁<br>75～89 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>10.50837/its.2205 | 査読の有無<br>有 |
| オープンアクセス<br>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<br>－ |

〔学会発表〕　計3件（うち招待講演　1件／うち国際学会　1件）

| |
|---|
| １．発表者名<br>Kayo Matsushita and Michiyo Furukawa |
| ２．発表標題<br>Diverging Strategies: Key Findings from a Comparative Study of Chinese and English Interpretation Using the Japan National Press Club Interpreting Corpus |
| ３．学会等名<br>The 4th East Asian Translation Studies Conference (EATS4)（国際学会） |
| ４．発表年<br>2022年 |

| |
|---|
| １．発表者名<br>松下佳世・古川典代・吉田理加 |
| ２．発表標題<br>多言語通訳コーパスを活用した　日英・日中・日西の訳出比較に基づく初期的考察 |
| ３．学会等名<br>日本通訳翻訳学会第23回年次大会 |
| ４．発表年<br>2022年 |

| |
|---|
| １．発表者名<br>Rika Yoshida |
| ２．発表標題<br>「多言語通訳コーパス（日・西・中・英）から見る文化仲介方略の異なりに関する考察」（原題はスペイン語） |
| ３．学会等名<br>サラマンカ大学「翻訳と異文化仲介」修士課程院生向け講座「翻訳と異文化仲介における研究の新潮流」（原題はスペイン語）（招待講演）（招待講演） |
| ４．発表年<br>2023年 |

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

| | 氏名<br>（ローマ字氏名）<br>（研究者番号） | 所属研究機関・部局・職<br>（機関番号） | 備考 |
|---|---|---|---|
| 研究分担者 | 古川　典代<br>(Furukawa Michiyo)<br>(70411270) | 神戸松蔭女子学院大学・文学部・教授<br>(34513) | |
| 研究分担者 | 吉田　理加<br>(Yoshida Rika)<br>(20761951) | 愛知県立大学・外国語学部・准教授<br>(23901) | |

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0件

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 | | | |
|---|---|---|---|---|
| トルコ | ボアズィチ大学翻訳通訳学部 | | | |