

# 国立国語研究所学術情報リポジトリ

## 近世・近代の日本語及び沖縄語訳聖書の平行・コーパスの構築

メタデータ	言語: Japanese 出版者: 公開日: 2023-11-24 キーワード (Ja): キーワード (En): 作成者: 宮川, 創 メールアドレス: 所属:
URL	<a href="https://doi.org/10.15084/0002000131">https://doi.org/10.15084/0002000131</a>

## 近世・近代の日本語及び沖縄語訳聖書の平行・コーパスの構築

宮川 創 (国立国語研究所研究系) †

### Construction of a Parallel Corpus of Early Modern and Modern Japanese and Okinawan Translations of the Bible

So Miyagawa (National Institute for Japanese Language and Linguistics)

#### 要旨

現存の聖書の日本語訳のなかで最も早いものはキリシタン資料における聖書の抄訳である。しかし、まとまった形での現存する最古の聖書和訳は、19世紀前半にギュツラフがシンガポールにて刊行した『約翰福音之伝』(「ヨハネによる福音書」)と『約翰上中下書』(ヨハネ書簡3通)である。この後、ベッテルハイムの沖縄語訳と漢和对訳を経て、江戸末期・明治初期にかけて、ヘボンやブラウンなどのプロテスタント訳、明治元訳、ニコライの正教会訳、ラゲのカトリック訳などが出版された。発表者はこれらの翻訳に含まれている「ヨハネによる福音書」の平行・コーパス(日本語7翻訳、沖縄語1翻訳、アイヌ語1翻訳、翻訳元の可能性が高いギリシア語批判校訂版、英訳、ラテン語訳、ドイツ語訳、中国語訳など)を作成し、JSON-LDとしてAPIを通じて外部にデータを提供できるシステムを構築した。最後に、本平行・コーパスを用いて翻訳間の影響関係を分析・視覚化するため、Rのstyloパッケージを用いて諸翻訳間のクラスター分析を行なった。

#### 1. はじめに

本稿は、江戸期と明治期を中心とした日本語および琉球語訳の聖書翻訳の平行・コーパス(並列テキスト集)を構築し、その翻訳間の影響関係を分析するものである。これまでも日本の聖書翻訳は多く研究されているが、それらを一元的にまとめたデジタル・平行・コーパスは存在していない。この研究はその初の試みとなる。このコーパスは、キリシタン資料における聖書抜粋、「ギュツラフ訳」、「ベッテルハイム沖縄語訳(琉球語訳/琉訳)」、「ヘボン/ブラウン訳」、「明治元訳」、「ニコライ/中井訳」、「ラゲ訳」、「大正改訳」など、近世・近代の多くの既存翻訳を元としている。また新たなデータとしてアイヌ語の「バチェラー訳」や「ニコライ/中井訳」の底本の一つとなった教会スラヴ語訳、「ラゲ訳」の底本となったヴルガタ訳ラテン語聖書、そのほか、プロテスタントの聖書翻訳に大きな影響を与えた、「モリソン訳」や「ブリッジマン/カルバートソン訳」の古典中国語訳などを追加している。本稿では、まず、構築中のこのコーパスの内容を述べ、次に設計、そして、最後にパイロット・スタディとして、「ヨハネによる福音書」の諸翻訳のスタイロメトリー(stylometry; 計量文体学/文体統計学)によるクラスター分析をした結果を示す。

---

† so-miyagawa@ninjal.ac.jp

1 本稿は、「言語資源ワークショップ2023」(LRW2023)で発表した内容の縮小版の予稿であり、この予稿に大幅な追記・修正を加えた論文は、『計量国語学』34巻4号(2024年3月20日発行予定)に掲載される予定である。

## 2. コーパスの内容

日本語における聖書翻訳の歴史は、近世のキリシタンの活動に端を発する。ジョン・セーリス (John Saris) による、キリシタン版日本語訳新約聖書の報告<sup>2</sup>はあるものの、その現存は確認されていない。しかし、様々な現存するキリシタン版の日本語資料には、聖書からの抜粋が収録されている。キリスト教修養書である『コンテムツス・ムンヂ』 (*Contemptus Mundi*) のキリシタン版日本語訳<sup>3</sup>は、聖書からの抜粋の日本語訳が豊富に含まれている。このテキストは複数の版が現存しているが、それらの内の一つでローマ字版 (1596 年刊) のものは、ドイツのヴォルフエンビュッテルにあるヘルツォーク・アウグスト図書館に所蔵されており<sup>4</sup>、ラテン語、ポルトガル語、そして日本語の要素が混在している。具体的には、ラテン語のウルガタ (*Vulgata*) 訳聖書からの引用があり、その後日本語訳が示されている。筆者は、ボーフム大学の研究チームと協力して、手書きテキスト自動翻刻ソフトウェア *Transkribus*<sup>5</sup>を用いて、この古典テキストのデジタル翻刻と TEI (*Text Encoding Initiative*) *Guidelines* に準拠した XML (*Extensive Markup Language*) 化を行った<sup>6</sup>。これは「日本語史研究用テキストデータ集」サイト上で公開され、CC BY 4.0 International ライセンスに基づいている<sup>7</sup>。これらのキリシタン資料における聖書翻訳の抜粋は、コーパスに追加する計画である。

一方、日本語への本格的な聖書翻訳は、19 世紀のプロテスタント宣教師カール・フリードリヒ・アウグスト・ギュツラフ (*Karl Friedrich August Gützlaff*) によって行われた。ギュツラフは特に「ヨハネによる福音書」を翻訳し、これが後の翻訳に影響を与えた。『コンテムツス・ムンヂ』や他のキリシタン版資料から抜粋された「ヨハネによる福音書」の部分は、パラレル・コーパスに追加する予定である。この「ギュツラフ訳」によって、日本語聖書翻訳の歴史と言語学的特性をさらに詳細に解析することが可能になる。本研究では、Gallica で公開されている「ギュツラフ訳」のフランス国立図書館本の「ヨハネによる福音書」を、伊波 (1997, 1998a, 1998b, 1999a, 1999b) の翻刻や岩崎 (1984) の翻刻、日本聖書教会 (2006) の翻刻と影印、ギュツラフ (2001) の影印を参考にしながら、デジタル翻刻したものである。

ジャン・バーナード・ベッテルハイム (*Jean Bernard Bettelheim*) はハンガリー出身のユダヤ教徒で、後に英国国教会に改宗し、日本で宣教活動を目指した。しかし、彼は日本本土ではなく、琉球王国 (現在の沖縄県) に到着し、1846 年から 8 年間、ほぼ軟禁状態で那覇の護国寺で過ごした。この期間中に彼は琉球語派・北琉球語群に属する沖縄語を学び、新約聖書のいくつかの部分を沖縄語に翻訳した。この聖書翻訳は、通常「琉球語訳聖書」または「琉訳聖書」と呼ばれ、国文学研究資料館の「国書データベース」で公開されている<sup>8</sup>。本稿では、「ベッテルハイム沖縄語訳」と呼称する。本パラレル・コーパスにおける「ベッテルハイム沖縄語訳」のテキストは、この「国書データベース」にて公開されている画像を、伊波 (1997, 1998a, 1998b, 1999a, 1999b) の翻刻を参考

<sup>2</sup> Saris (1900: 141) 参照。

<sup>3</sup> 現在では、『キリストに倣いて』 (*Imitatio Christi*) というタイトルで知られる。この書は、14-15 世紀の、アウグスチノ会士であったトマス・ア・ケンピス (*Thomas à Kempis*) の著作である。

<sup>4</sup> 岸本・白井 (2019) 参照。

<sup>5</sup> 宮川 (2022) 参照。

<sup>6</sup> ノイツラ・宮川 (2022) 参照。

<sup>7</sup> 「コンテムツス・ムンヂ (HAB 所蔵本)」日本語史研究用テキストデータ集 ([https://www2.ninjal.ac.jp/textdb\\_dataset/cmhab/](https://www2.ninjal.ac.jp/textdb_dataset/cmhab/), 2023 年 8 月 17 日確認)。

<sup>8</sup> <https://kokusho.nijl.ac.jp/biblio/100244876/1?ln=ja> (2023 年 8 月 17 日確認)。

にしながら、デジタル翻刻したものである。

沖縄語での翻訳は、沖縄語の動詞語尾や格助詞など、その特徴を反映しており、琉球歴史的仮名遣いが用いられている。ベッテルハイムは欧州に帰国後、沖縄語と日本語が異なることに気付き、四福音書を漢和对訳で翻訳し、ウィーンで出版した。ギュツラフとベッテルハイムという二人の宣教師による聖書翻訳は、日本及び琉球におけるキリスト教宣教と言語研究に貴重な資料となっている。特にベッテルハイムの沖縄語訳は、沖縄語や琉球文化、そして19世紀の宣教活動に対する深い理解を可能にする。

19世紀後半には、別の宣教師、ジョナサン・ゴープル (Jonathan Goble) が日本に来て、「マタイによる福音書」の日本語訳を横浜で出版した。「ゴープル訳」はデジタル化されていないが、「明治学院大学図書館デジタルアーカイブス」でその画像が公開されている<sup>9</sup>。

ジェームズ・カーティス・ヘボン (James Curtis Hepburn) は19世紀後半に活動した米国長老派教会の医療伝道宣教師で、和英語辞典『和英語林集成』の著者であり、ヘボン式ローマ字の開発者でもある。彼はシンガポールでギュツラフ訳の『約翰福音之傳』を手に入れ、その後日本へ向かい、サミュエル・ロビンス・ブラウン (Samuel Robbins Brown) と共に四福音書の和訳を行った。これらの訳は「ヘボン/ブラウン訳」と呼ばれ、当時の日本での印刷が明治政府によって禁止されたため、上海で印刷された後、日本に輸入された。この「ヘボン/ブラウン訳」は Wikisource から取得され<sup>10</sup>、ルビの形式が正規表現を用いて HTML 形式に変換されている。このテキストは概して漢字表記が少ない。

1875年には、長老派の宣教師クリストファー・カロザース (Christopher Carrothers) 名義で、日本人信徒の加藤九郎が翻訳し、編集した『畧解新約聖書』も出版されたが、この中には「ルカによる福音書」および「ヨハネによる福音書」は含まれていない<sup>11</sup>。

ヘボンやブラウンが属していたのは、カルヴァン派の系譜を継ぐ教派で、横浜を中心に活躍していた。一方、バプテスト派のネイサン・ブラウンは、洗礼 (彼の用語では「浸礼 (しづめ)」) に対する神学的違いから、ヘボンやサミュエル・R・ブラウン (Samuel R Brown) と対立した関係にあった。

このように、19世紀後半の日本では、プロテスタントの宣教師たちが聖書の翻訳において多大な貢献を果たしていた。ヘボンを中心としたプロテスタント宣教師グループは、1872年に標準的な日本語訳を作成するための翻訳委員会を設立した。ヘボン、サミュエル・R・ブラウン、アメリカン・ボードのダニエル・クロスビー・グリーン (Daniel Crosby Greene) や英国聖公会のジョン・パイパー (John Piper) やウィリアム・ボール・ライト (William Ball Wright)、バプテスト派のネイサン・ブラウンなども参加した。しかし、翻訳の際に使用する用語について意見が分かれ、特に「洗礼」か「浸礼 (しづめ)」かでバプテスト派のネイサン・ブラウンが脱会した。この委員会による新約聖書の日本語訳は、1875年から順次出版され、これは「明治元訳」と呼ばれている<sup>12</sup>。

<sup>9</sup> [https://mgda.meijigakuin.ac.jp/bible/book\\_image/1871matgbl2](https://mgda.meijigakuin.ac.jp/bible/book_image/1871matgbl2) (2023年8月17日確認)。

<sup>10</sup> <https://ja.wikisource.org/wiki/%E6%96%B0%E7%B4%84%E8%81%96%E6%9B%B8%EF%BC%88%E3%83%98%E3%83%9C%E3%83%B3%E8%A8%B3%EF%BC%89> (2023年9月30日閲覧)。

<sup>11</sup> 『畧解新約聖書 第一冊馬太』: [https://mgda.meijigakuin.ac.jp/bible/book\\_image/1875matcar](https://mgda.meijigakuin.ac.jp/bible/book_image/1875matcar) (2023年8月17日確認)。『畧解新約聖書 第二冊馬可』: [https://mgda.meijigakuin.ac.jp/bible/book\\_image/1876marcar](https://mgda.meijigakuin.ac.jp/bible/book_image/1876marcar) (2023年8月17日確認)。

<sup>12</sup> <https://ja.wikisource.org/wiki/%E6%98%8E%E6%B2%BB%E5%85%83%E8%A8%B3%E6%>

正教会では、ロシア正教会の宣教師ニコライ（Николай；出生名 Касаткин カサートキン）が 1861 年に日本での宣教を開始し、漢学者で正教徒のパウエル中井木菟麻呂と共に新約聖書を翻訳した。この訳は 1901 年に公刊され、「ニコライ／中井訳」と呼ばれている。このニコライ／中井訳のデータは Wikisource から取得され<sup>13</sup>、2023 年 8 月 17 日時点で一部テキストが不足しているため、それを補完する作業を進めている。

ローマ・カトリック教会からは、1897 年には高橋五郎が司祭のミシェル・スタイシェン（Michael Steichen）の口述から和訳した「スタイシェン／高橋五郎訳」が出ており、これは「ヨハネによる福音書」を含む四福音書の日本語文語訳である。また、1910 年にエミール・ラゲ（Émile Ragué）による翻訳、通称「ラゲ訳」が出版されており、この翻訳のデジタル・テキストも Wikisource で公開されている<sup>14</sup>。

このように、明治時代の日本においては多くの宣教師や学者が関わり、多角的な視点から聖書の日本語訳が行われた。それぞれの教派が独自の方法で翻訳を行い、日本にキリスト教を広めるとともに、日本語としても重要な文献をした。

大正時代以降で重要な文語訳聖書として、1910 年に設立された改訳委員会による 1917 年の『改訳 新約聖書』（「大正改訳」）がある。この「大正改訳」は、翻訳の誤りなど様々な問題があった明治元訳の新約聖書の改訳である。大正改訳では、外国人宣教師だけでなく、多数の日本人キリスト者も参加した。大正改訳は Wikisource でルビなし版として公開されているが、ルビ付き版は「日本語の聖書」というホームページで公開されていることが明らかにされている<sup>15</sup>。

次に、1928 年に永井直治によって発行された『新契約聖書』がある。永井版は、他の翻訳を参照せずに、Stephanus 版の Textus Receptus を底本とした日本語文語訳であり、内村鑑三の序言がついており、国立国会図書館デジタルコレクションからデジタル翻刻されている<sup>16</sup>。全体として、明治以降の聖書の文語体日本語翻訳は、多様な試みと改訂が行われ、デジタル化されたテキストも存在するが<sup>17</sup>、ルビや漢字の読みに一貫性や明確性がない場合もあるという点が強調される。

以下は、今回取り上げなかった聖書翻訳を含めた、1945 年までの言語別の聖書翻訳のリストである。聖書が書別で出版されたり、分冊で出版されたりしている場合は、「ヨ

---

96%B0%E7%B4%84%E8%81%96%E6%9B%B8\_(%E6%98%8E%E6%B2%BB37%E5%B9%B4)  
(2023 年 8 月 17 日確認)。

<sup>13</sup> <https://ja.wikisource.org/wiki/%E6%88%91%E4%B8%BB%E3%82%A4%E3%82%A4%E3%82%B9%E3%82%B9%E3%83%8F%E3%83%AA%E3%82%B9%E3%83%88%E3%82%B9%E3%81%AE%E6%96%B0%E7%B4%84> (2023 年 8 月 17 日確認)。

<sup>14</sup> <https://ja.wikisource.org/wiki/%E6%88%91%E4%B8%BB%E3%82%A4%E3%82%A8%E3%82%BA%E3%82%B9%E3%82%AD%E3%83%AA%E3%82%B9%E3%83%88%E3%81%AE%E6%96%B0%E7%B4%84%E8%81%96%E6%9B%B8> (2023 年 9 月 30 日確認)。

<sup>15</sup> ルビなし版は、[https://ja.wikisource.org/wiki/%E5%A4%A7%E6%AD%A3%E6%94%B9%E8%A8%B3%E6%96%B0%E7%B4%84%E8%81%96%E6%9B%B8\\_\(%E3%83%AB%E3%83%93%E4%BB%98\)](https://ja.wikisource.org/wiki/%E5%A4%A7%E6%AD%A3%E6%94%B9%E8%A8%B3%E6%96%B0%E7%B4%84%E8%81%96%E6%9B%B8_(%E3%83%AB%E3%83%93%E4%BB%98)) (2023 年 8 月 17 日閲覧)。ルビあり版は、<http://bible.salterrae.net/taisho/xml/john.xml> (2023 年 8 月 17 日閲覧)。

<sup>16</sup> <https://dl.ndl.go.jp/pid/1109881> (2023 年 8 月 17 日閲覧)。

<sup>17</sup> <https://ja.wikisource.org/wiki/%E6%96%B0%E5%A5%91%E7%B4%84%E8%81%96%E6%9B%B8> (2023 年 9 月 30 日閲覧)。

ハネによる福音書」の出版年で代表させてある。「ヨハネによる福音書」が含まれない場合は、福音書を中心とした他の書の出版年で代表させてある。

#### 日本語

- 16/7 世紀：キリシタン訳（未発見，ジョン・セーリスによる報告<sup>18</sup>）
- 16/7 世紀：キリシタン資料中の聖書引用
- 1837 年：ギュツラフ訳（「ヨハネによる福音書」及び「ヨハネの手紙一・二・三」）
- 1850 年：ウィリアムズ訳（「マタイによる福音書」など）
- 1871 年：ゴープル訳（「マタイによる福音書」）
- 1872 年：ヘボン／ブラウン訳（「マタイによる福音書」，「マルコによる福音書」，「ヨハネによる福音書」）
- 1873 年：ベッテルハイム漢和对訳（「マタイによる福音書」，「マルコによる福音書」，「ルカによる福音書」，「ヨハネによる福音書」）
- 1875 年：加藤九郎／カロザース訳（「マタイによる福音書」，「マルコによる福音書」）
- 1879 年：N・ブラウン訳（新約聖書）
- 1880 年：明治元訳（新約聖書；後に旧約聖書）
- 1881 年：井深梶之助訳（「マルコによる福音書」）
- 1897 年：スタイシェン／高橋五郎訳（「マタイによる福音書」，「マルコによる福音書」，「ルカによる福音書」，「ヨハネによる福音書」）
- 1899 年：中川藤四郎等訳（詩篇）
- 1901 年：ニコライ／中井訳（新約聖書，「詩篇」など）
- 1910 年：ラゲ訳（新約聖書）
- 1905/11/14 年：左近義弼訳（「マタイによる福音書」など）
- 1917 年：大正改訳（新約聖書）
- 1928 年：永井直治訳（新約聖書）
- 1944 年：塚本虎二訳（新約聖書）

#### 沖縄語

- 1855 年：ベッテルハイム沖縄語訳（四福音書など）

---

<sup>18</sup> Saris (1900: 141) 参照.

## アイヌ語

- 1897年：バチェラー訳（新約聖書、「詩篇」など）

これらの聖書のうち、「ウィリアムズ訳」（1850年）は主に「マタイによる福音書」と「創世記」からなり、「ヨハネによる福音書」は5章9節まで試訳として現存している。「ゴープル訳」（1871年）は「マタイによる福音書」のみである。「ベッテルハイム漢和対訳」（1873年）の「ヨハネによる福音書」は見つかっていない。他の三福音書および使徒行伝は既に発見されている。「カロザース／加藤訳」（1875年）は、「マタイによる福音書」と「マルコによる福音書」のみである。井深梶之助訳（1881年）は、「マルコ福音書」のみである。

### 3. コーパスの設計

初期段階にある本コーパス・プロジェクトは、新約聖書のうち、まず『ヨハネによる福音書』の書翻訳のデジタル翻刻のコーパス化に取り組んだ。

このコーパスの設計には Omeka S というデジタル・アーカイブのコンテンツ・マネジメント・システムを使用した（図1）。Omeka S は文化遺産のデジタル・アーカイブに特化しており、リンクト・オープン・データ・モジュールを実装することで、データを、API (application programming interface) を通して JSON-LD 形式で出力できるように設計されている。

#### 新日本語・琉球語ヘクサブラ ヨハネによる福音書1章1節

タイトル	新日本語・琉球語ヘクサブラ ヨハネによる福音書1章1節
口語訳聖書	初めに言があった。言は神と共にあった。言は神であった。
ギュツラフ訳	ハジマリニ カシコイモノ ゴザル。コノ カシコイモノ ゴクラクトモノ ゴザル。コノ カシコイモノ フゴクラク。
ベッテルハイム訳	ハジマリニ カシコイモノ ヲテ、コノカシコイモノヤ シヤウテイト トモニヲタン カノ カシコイモノヤ シヤウテイド。
ヘボン・ブラウン訳	元始(はじめ) 言靈(ことだま) あり 言靈(ことだま) は 神(かみ) とともにあり 言靈(ことだま) は 神(かみ) なり
明治元訳	太初(はじめ) に 道(ことば) あり 道(ことば) は 神(かみ) と 偕(とも) にあり 道(ことば) は 即(すなはち) 神(かみ) なり
ニコライ・中井訳	太初(はじめ) に 言(ことば) 有(あ)り、言(ことば) は 神(かみ) と 共(とも) に 在(あ)り、言(ことば) は 即(すなはち) 神(かみ) なり。
ラゲ訳	元始(はじめ) に 御言(みことば) あり、御言(みことば) 神(かみ) の 御許(おんもと) に 在(あ)り、御言(みことば) は 神(かみ) にて ありたり。
永井直治訳	初に言ありき、また言は神と偕にありき、また言は神なりき。
大正改訳聖書	太初に言あり、言は神と偕にあり、言は神なりき。
Textus Receptus	εν αρχη ην ο λογος και ο λογος ην προς τον θεον και θεος ην ο λογος
King James Version	In the beginning was the Word, and the Word was with God, and the Word was God.
書名	ヨハネによる福音書
章番号	1
節番号、節の番号	1



図1 Omeka S による本パラレル・コーパスのオンラインでの表示

聖書翻訳の多様性を考慮する上で、パラレルコーパスの作成は言語学、神学、文化研

究など多くの学術分野で重要な貢献を果たす可能性がある。まず、メタデータの構造化は研究の基盤を形成する。この目的のために、JSON-LD や RDF といった形式が使用される。これらはメタデータを機械可読かつ相互運用可能な形で整理し、各翻訳の文脈や特性、関連するリソースに対するリンクをする。続いて、Linked Open Data (LOD) の導入により、聖書の各翻訳が存在する広がりの中で他の文献や歴史的事件、地理的場所といった外部リソースとどのように関連しているかを明示する。

次に、デジタルアーカイブ専用に作られたコンテンツ管理システム (Content Management System: CMS) の Omeka S が一元管理と公開のプラットフォームとして機能する<sup>19</sup>。これは大規模なデータベースを整理し、多様な視点からのアクセスを可能とするものである。その上で、API を通じてこのパラレル・コーパスは外部の研究者やデベロッパーにも開かれ、新たな分析手法やアプリケーションの開発が促進される。

さらに、研究成果の信頼性と拡散も重要な側面である。この点で、Omeka S プラットフォーム内で各リソースに明示される引用情報と、SNS シェア機能によって、研究成果はより広範な公衆に届けられる。最後に、柔軟な検索機能が大量の翻訳データを効率的に探索するための手段をする。

以上の手法は聖書翻訳のパラレル・コーパス作成において多層的な貢献を果たす。これにより、言語学、神学、文化研究など多様な研究分野で、聖書翻訳に対するより高度で包括的な理解が可能となる。

#### 4. スタイロメトリーによるクラスター分析

以上のように作成したパラレル・コーパスは、言語学的・文学的・宗教学的・歴史学的など様々な分野で使用できる。この技術によって、コーパスのデータは広く共有されやすくなる。本節では、パイロット・スタディとして簡易な計量文体学 (スタイロメトリー) の分析方法の 1 つであるクラスター分析を行う。日本語と琉球語で翻訳された『ヨハネによる福音書』の計量的類似度分析について報告する。すなわち、スタイロメトリーの手法を用いて、聖書翻訳の成り立ちが文体統計のクラスター分析にどのように反映されるかを検討する。この研究で使用した主要なツールは、R 言語用のパッケージである *stylo* (Eder et al. 2016) である。スタイロメトリーは、テキスト間の文体や類似度を統計解析する分野であり、著者推定や年代推定などに用いられる。翻訳間の影響関係の分析では、スタイロメトリー (文体学の統計解析) を用いている。この手法によって、各聖書翻訳の文体の類似度が試験的に計測され、それを基に翻訳の影響関係を計量的に解析する。このような分析は、翻訳文献の影響力や変遷を理解する上で非常に有用なものとなる。全体として、この研究はデジタル・ヒューマニティーズの手法を活用して聖書翻訳の歴史と影響関係を新たな角度から研究する試みであり、今後の研究に対する重要な基礎となると言える。

*stylo* パッケージは主成分分析、クラスター分析、ネットワーク分析など多くの統計手法をしている。本研究では、クラスター分析を主要な統計手法として用いた。R 言語の実行環境としては *Rstudio* を使用した。入力データはプレーン・テキストであり、言語設定は CJK (中国語・日本語・韓国語) に設定した。分析単位は文字で、*n-gram* は 2 に設定した。

初回の分析では、テキストは主に文字表記 (カタカナか漢字仮名交じり表記か) によってクラスター化されてしまった。この結果は文体の類似度ではなく、文字表記に依存していることを示している。この問題を解消するために、すべてのテキストをカタカナに統一し、再分析を行った。

以下の図 2 は、このカタカナ化を行った後の、*stylo* によるクラスター分析の結果で

<sup>19</sup> 宮川 (2023) 参照。

ある。

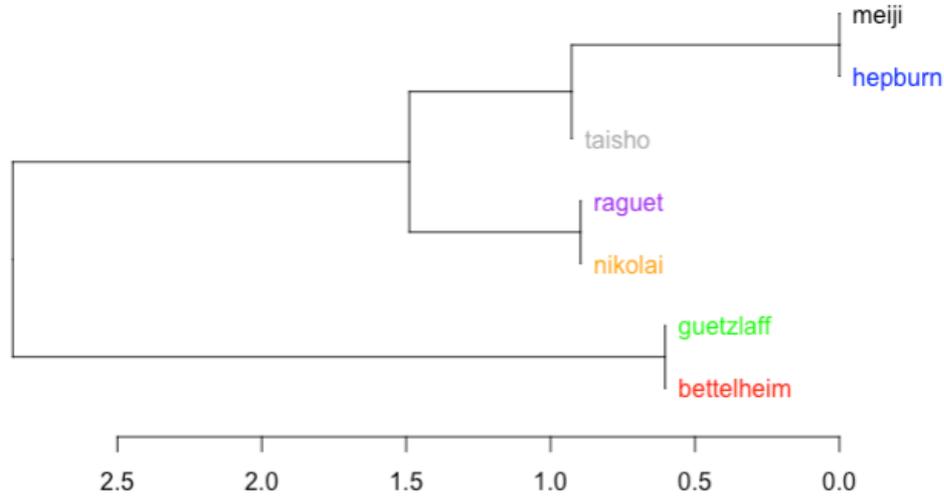


図2 「ヨハネによる福音書」の諸翻訳のクラスター分析

「明治元訳」(meiji)と「ヘボン／ブラウン訳」(hepburn)が最も近いことが示されているが、これは、翻訳者が同じであるため、説明がつく。次に「ギュツラフ訳」(guetzlaff)と「ベッテルハイム沖縄誤訳」(bettelheim)が近いことが示されている。「ギュツラフ訳」は知多半島の方言の影響を受けた言語変種で書かれている一方、「ベッテルハイム沖縄語訳」は、当時の琉球王国の沖縄語で書かれている。言語の違いによる助詞や活用語尾の違いは大きいものの、「ベッテルハイム沖縄語訳」自身が「ギュツラフ訳」を参考にして書かれたため<sup>20</sup>、その類似度を説明できる。次に「ラゲ訳」(raguet)と「ニコライ／中井訳」(nikolai)の近さが示されている。これらは、異なる教派による翻訳であり（「ラゲ訳」はローマ・カトリック教会、「ニコライ／中井訳」はロシア正教会）、かつ、異なる底本の聖書（「ラゲ訳」はラテン語訳聖書、「ニコライ／中井訳」はギリシア語聖書および教会スラブ語訳聖書）から訳されている。しかし、「ニコライ／中井訳」は1901年、「ラゲ訳」は1910年と、比較的近い時期の、漢文訓読文の影響を受けた日本語文語で書かれているため、それらの類似度はある程度説明がつく。次に、類似度が高いのは、「明治元訳」(meiji)と「ヘボン／ブラウン訳」(hepburn)のクラスターと「大正改訳」(taisho)である。「大正改訳」は、「明治元訳」の改訳であるため、容易にその類似度は説明がつく。最後に、「明治元訳」・「ヘボン／ブラウン訳」・「大正改訳」・「ラゲ訳」・「ニコライ／中井訳」がクラスターをなしているが、これらは、日本語文語で書かれているという点で、残りの、標準的な日本語文語が用いられていない「ギュツラフ訳」・「ベッテルハイム沖縄語訳」のクラスターと大きく異なっていることの説明がつく。

このように「ヨハネによる福音書」の諸翻訳において、文体統計のクラスター分析は、翻訳の成り立ちによるグループ分けを明確に示すことができた。すなわち、スタイロメトリーのクラスター分析は日本語と琉球語の「ヨハネによる福音書」の計量的な類似度を分析する有用なツールであることが確認された。ただし、初回の分析では文字表記が大きな影響を与えたため、このような外部要因を制御することが重要であると示された。今後は、この手法をさらに発展させ、多言語・多文化のテキスト分析に応用する可能性が考えられる。

<sup>20</sup> 伊波 (1997) 参照。

以上のように、*stylo* パッケージは多様な分析手法をしており、特に文字表記などの外部要因を制御した上で、高度なテキスト分析が可能であることが示された。この研究は、日本語と琉球語、さらには他の CJK 言語に対するスタイロメトリーの応用可能性を拓く一例である。

## 5. 結論と展望

このパイロット・スタディは、スタイロメトリーを用いて聖書翻訳の文体と成り立ちを解析するための基礎を築いた。今後は、現在構築中の聖書翻訳の追加を完了させるとともに、ラテン語訳聖書や教会スラブ語訳や、日本語聖書翻訳に影響を与えたと先行研究が考えているブリッジマン／カルバートソン訳やモリソン訳などの漢訳聖書を追加する。また、ある程度構築できている「ヨハネによる福音書」のみならず、他の新約聖書の諸書、さらには旧約聖書の諸書にも広げていく。そして、また、スタイロメトリー分析も、クラスター分析だけでなく、主成分分析や多変量解析など *stylo* に装備されている他の分析方法の活用法を模索したり、変数を変更しながら結果の比較を考察したりする予定である。

## 謝 辞

本研究は、人間文化研究機構共同研究プロジェクト（2022 年度～）共創先導プロジェクト共創促進研究「学術知デジタルライブラリの構築」国語研拠点（代表：高田智和）、国立国語研究所共同研究プロジェクト（2022 年度～）広領域連携型基幹研究「異分野融合による総合書物学の拡張的研究」国語研ユニット「古辞書類に基づく語彙資源の拡張と語彙・表記の史的変遷」（代表：高田智和）、国立国語研究所共同研究プロジェクト（2022 年度～）機関拠点型基幹研究「開かれた言語資源による日本語の実証的・応用的研究」基幹型「多様な語彙資源を統合した研究活用基盤の共創」統括班（代表：小木曾智信）、日本学術振興会科学研究費助成事業基盤研究（A）23H00007「日本語諸方言の形態素解析用辞書の構築と活用」（代表：小木曾智信）、日本学術振興会科学研究費助成事業基盤研究（B）19H01265「多言語による日本語学用語辞典および日琉諸語の用例に対するグロス規範の作成」（代表：マシュー・ジスク）の支援を受けている。

## 文 献

- Maciej Eder, Jan Rybicki, and Mike Kestemont (2016). “Stylometry with R: a package for computational text analysis”, *R Journal*, 8:1. pp. 107–21. Available at <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html> (accessed August 17, 2023).
- John Saris (1900). *The voyage of Captain John Saris to Japan, 1613*, Hakluyt Society. Available at <https://archive.org/details/captainjvoyageof00saririch> (accessed August 17, 2023).
- 伊波和正 (1997). 「ベッテルハイム著『琉訳聖書』の分析:『ヨハネ福音書』(『欽定訳聖書』・ギュツラフ日本語訳・ベッテルハイム琉球語訳)比較対照: 資料編 I(1-5 章)」『沖縄国際大学外国語研究』 2 :1, pp. 155–206.
- (1998a). 「ベッテルハイム著『琉訳聖書』の分析:『ヨハネ福音書』(『欽定訳聖書』・ギュツラフ日本語訳・ベッテルハイム琉球語訳)比較対照: 資料編 II(6-8 章)」『沖縄国際大学外国語研究』 2:2, pp. 235–278.
- (1998b). 「ベッテルハイム著『英琉聖書』の分析:『ヨハネ福音書』(『欽定訳聖書』・ギュツラフ日本語訳・ベッテルハイム琉球語訳)比較対象: 資料篇 iii(9-12 章)」『沖縄国際大学外国語研究』 3:1, pp. 255–305.
- (1999a). 「ベッテルハイム著『琉訳聖書』の分析:『ヨハネ副音書』(『欽定訳聖書』・ギュツラフ日本語訳・ベッテルハイム琉球語訳)比較対象: 資料編 IV(13-17 章)」

- 『沖縄国際大学外国語研究』 3:2, pp. 179–190.
- (1999b). 「ベッテルハイム著『琉訳聖書』の分析: 『ヨハネ福音書』 (『欽定訳聖書』・ギュツラフ日本語訳・ベッテルハイム琉球語訳)比較対照 : 資料編 V(18-21 章)」 『沖縄国際大学外国語研究』 4:1, pp. 129–168.
- 岩崎摂子 (1984). 『善徳纂「約翰福音之伝」: 本文ならびに総索引』 桜楓社.
- 岸本恵実・白井純 (2019). 「新出本・ヘルツォーク・アウグスト図書館蔵ローマ字本『コンテムツスムンヂ』 (1596 年天草刊) について」 『大阪大学大学院文学研究科紀要』 59, pp. 37–53.
- カール・フリードリッヒ・アウグスト・ギュツラフ (2001). 『約翰福音之伝・約翰上中下書』 覆刻版, 新教出版社.
- 日本聖書協会 (2006). 『ギュツラフ訳 ヨハネによる福音書 - 現代版, 語句の解説つき 抜粋朗読 CD つき』 日本聖書協会.
- ノイツラ・ゾフィー・宮川創 (2022). 「HTR プログラム Transkribus による日本語キリタン版『コンテムツス・ムンヂ』のデジタルアーカイブ化」 『デジタルアーカイブ学会誌』 6:s3, pp. s123–s126.
- 宮川創 (2022). 「Transkribus による手書きテキスト資料の自動翻刻」 一般財団法人人文情報学研究所 (監修)・石田友梨・大向一輝・小風綾乃・永崎研宣・宮川創・渡邊要一郎 (編) 『人文学のためのテキストデータ構築入門 TEI ガイドラインに準拠した取り組みにむけて』 文学通信, pp. 56-109.
- 宮川創 (2023). 「Omeka S を用いた言語資源デジタルアーカイブの構築」 宮川創 『日琉諸語の記述・保存研究』 1, pp. 43-59.