

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 14 日現在

機関番号：14401

研究種目：基盤研究(C) (一般)

研究期間：2013～2016

課題番号：25330383

研究課題名(和文)近代ギリシャ語新約聖書校訂本の異同識別に関する数理的比較研究

研究課題名(英文)A comparative computational study on variant readings of modern Greek critical editions in the New Testament

研究代表者

三宅 真紀 (MIYAKE, Maki)

大阪大学・言語文化研究科(言語文化専攻)・准教授

研究者番号：80448018

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：本研究は、新約聖書正文批判研究におけるギリシャ語写本、および校訂本の類似性に関する研究を行った。テキストの読みの違い(異読)を数量化して表された文書間の距離計算手法について注目し、それぞれの異読距離が持つ特徴、および、文書間の類似・相違性について分析を行った。同時に、読みの違いをインタラクティブに比較するためのWebアプリケーションの開発を行い、異読の特徴について視覚的な把握が可能になった。

研究成果の概要(英文)：The study aims to investigate the similarities of Greek manuscripts and the modern critical editions of the Greek New Testament. Focusing on the measurement of distances that quantify differences between variant readings in given texts, we analyzed the characteristics of each variant distance between manuscripts of texts and then the similarities and differences between the texts. Simultaneously, we developed an interactive web application for comparing differences between the texts, which visually provides insight into the distinctive features of variant readings.

研究分野：計算言語学

キーワード：テキスト計量分析 新約聖書

1. 研究開始当初の背景

情報科学技術の革新的な発展およびデジタル・メディアの普及は、従来の人文学研究のあり方にも多大な影響を与えてきた。欧米の研究機関では、学際的な複合領域としてデジタル・ヒューマニティーズが開拓され、人文学と情報学、工学等の関連諸科学との連携体制を取りながら、学際的研究の組織的な取り組みへと進展している。

聖書学においても、特に写本研究において、情報科学技術を駆使した大規模プロジェクトが推進されてきた。近年では、図書館・博物館に所蔵されている歴史的写本を対象に、最新スキャン技術によるデジタル化が世界的な規模で進められている。また、様々な言語の聖書電子化テキストやシソーラスのオンライン閲覧・検索システムの構築に関するプロジェクトも数多く存在する。なかでも、ドイツ・ミュンスター大学の新約聖書本文研究の Digital Nestle-Aland プロジェクトは、Nestle-Aland 第 27 版のギリシャ語新約聖書のデジタル校訂本の開発を行っている。

2. 研究の目的

本研究は、近代ギリシャ語新約聖書の主要な校訂本を分析の対象として、校訂本間の類似・相違性について解析し、諸本の比較分析を行う。テキストの大部分が一致している諸校訂本の差異を明確に表すために、文書間の微小な差異を計る適切な統計指標について調査を行う。そして、校訂本間の相違点を考察し、統計指標に基づいた比較分析結果を提示する。さらに分析の視野を広げ、諸本間の影響関係を基準にして、参照した写本の傾向についても考察を行う。

また、諸本比較データによる有益な本文批評研究を実現するために、異本比較研究を目的とした扱える分析ツールを開発する。諸校訂本の特徴や異同箇所を体系的に把握しやすいように、インタラクティブな操作が容易に行え、視覚的効果を考慮したインターフェースを設計し、諸本間の類似度指標値とデジタル校訂本が連動した Web アプリケーションを作成する。

3. 研究の方法

まず、分析に使用する校訂本を選定は、本文批評ならびに翻訳聖書への影響力の 2 つの観点から行う。プロトタイプ研究には、現代の聖書学で最高水準とされる Nestle-Aland 27 版(1993)を使用し、その NA27 版と比較するために、共通する箇所が多くみられる Westcott-Hort(1881)校訂本を扱う。一方で、多くの異本箇所が存在すると知られている von Soden(1913)を対象とする。また、本文批評研究ではあまり取り上げられることは少ないが、翻訳聖書に多大な影響を及ぼした

校訂本である Scrivener(1894)版を使用する。分析の進度に伴い、分析対象の校訂本数も段階的に増やしていく。

諸校訂本の比較分析に適用する方法論は、情報検索におけるベクトル空間モデルに基づく類似度計算法や文字列の類似度尺度として使用されている編集距離を応用しながら、諸校訂本の差異を明らかに示す。

分析結果に基づいた校訂本の違いを視覚化するために、Web アプリケーションの作成する。R 言語の Shiny パッケージを使用して、諸本間の類似度指標値の遷移グラフをインタラクティブに描画する機能を搭載する。

さらに、本研究で提案した異読距離法の妥当性を示すために、写本の系統研究において、ベンチマークデータとして広く浸透している「人工写本」テキストを対象として、異読から作成された異読距離行列による写本の系統樹推定を行った。

4. 研究成果

(1) 類似したテキストの差異を計る手法に重点をおいたプロトタイプ研究の成果として、「3. 研究の方法」で示した 4 つの校訂本を分析対象とし、共通性の多い文章に埋もれている差異を効率よく抽出するための手法について調べた。

テキストの異同箇所を特定する計量分析として、諸校訂本の特徴語彙に基づくミクロ的な視点、および、テキスト全体の使用単語の類似度計算によるマクロ的な観点から諸本間の相違点について観察をした。

また、ミクロな分析には、対数尤度比検定などの手法を適用して、校訂本の特徴語彙の抽出を試みた。個々の特徴語彙は、出現する本文を逐次参照しながら語彙の使用傾向の違いを観察した。他方、節や n-gram 単位などで小単位に区分されたテキスト情報から、単語の類似度計算を行った。

校訂本の類似度計算をもとに、校訂本間の差異について比較を行った。さらに、欽定本 KJV などの英語版についても類似度計算を適用した。その結果、ギリシャ語の校訂本においては、大半の部分が共通している文章から、わずかな差異の部分を抽出する手法が求められたのに対して、英語版においては、語彙のバリエーションや解釈の違い等から生じられる、圧倒的な差異の多さが確認された。

(2) 校訂本のテキスト間の差異を視覚化する Web アプリケーションの開発を行った。テキスト間の類似度を分析する手法には、編集距離から測る方法と、単語のベクトル空間モデルから計算する方法を実装した。この類似度計算は R で行っているため、R の Web アプリケーション用のパッケージである Shiny を利用して開発を行った。テキストの類似度比較にあたっては、文(節)単位の比較を行い、図 1 で示すようなテキスト全体の類似度を俯

瞰的にみることのできるマクロ的な機能と、一文内を ngram で区切って小さな範囲においてテキストの違いを動的に確認する機能の 2 種類を用意し、マクロ的・ミクロ的な分析を可能にしている。さらに、4 つの校訂本の文書全体を通して類似・異同箇所を特定するための描画機能もつけている。(図 2)

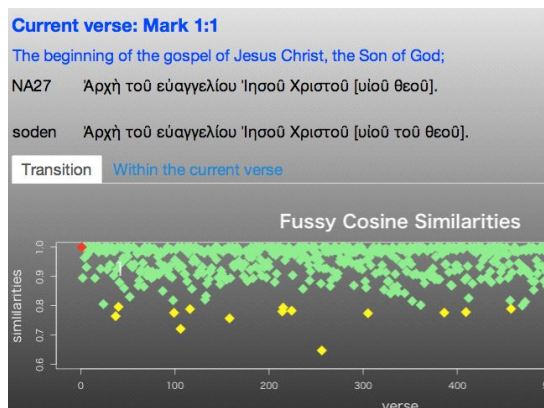


図 1：校訂本の類似度

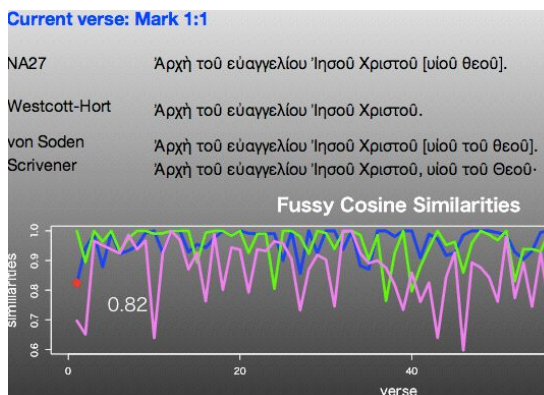


図 2：4 校訂本の類似度比較

(3)ギリシャ語新約聖書の校訂本間における差異の研究の対象を写本に拡張した研究を進めた。分析対象には校訂本が参照したギリシャ語写本の異読情報を利用し、諸写本・校訂本の差異を調査を行った。近代聖書学の正文批判研究で重要視されている大文字写本・小文字写本群の 23 写本の他に、異読情報を参照した新約聖書文献研究センター (CNNTS) の新約聖書正文批評アパルトゥスで扱われた 4 種類の校訂本を分析に追加した。使用した校訂本は、正文批評版とテキストゥス・レセプトゥスに種別することができ、正文批評版には、世界聖書協会連合による UBS 第 3 版、および、米国聖書文学学会発行の SBL 版を使用した。テキストゥス・レセプトゥス版には、オックスフォード 1873 版、および、Robinson & Pierpont (2005) が編纂したテキストをもとにしたマジョリティ・テキスト版 (Robinson, 2005) を使用した。異読箇所の文字情報を直接反映させる方法として、基準本の SBL 版と各文書との読みの異なり程度を測るために正規化レーベンシュタイン距離を計算し、多次元尺度構成法を

適用して、図 3 に表すような 2 次元平面上に可視化された写本の位置情報を観察しながら、諸写本・校訂本の違いを考察した。異読データには、正文批判研究で作成された異読コード、ならびに、文字列の異なり程度を表す指標である編集距離の 2 種類の尺度を利用し、両者における多次元尺度図の結果について比較を行った。

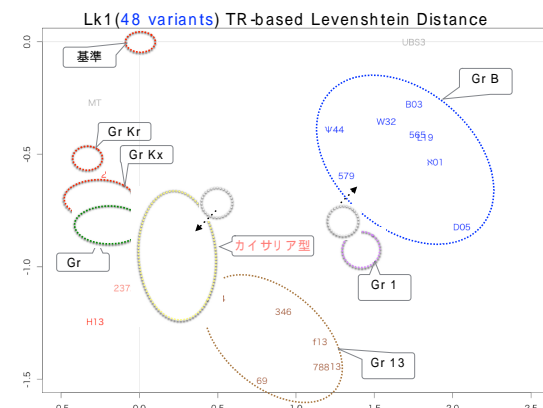


図 3：多次元尺度図による写本の違い

(4)ギリシャ語新約聖書の写本・校訂本間の写本の読みの違いをもとに数量化した写本間写本間の類似・相違性に関する調査結果を踏まえて、異読距離計算の手法の妥当性について分析を行った。分析対象のテキストには、本研究の写本・校訂本では、適応する異読距離法の精度を計算することが不可能であるため、写本の系統研究において、ベンチマークデータとして広く浸透している「人工写本」テキストを対象として、異読から作成された異読距離行列による写本の系統樹推定を行った。人工写本データには、系統解析法の適用を目的として人為的に作成された、Spencer et al. (2004) の Parzival を使用した。異読箇所の比較として、校合様式の異なる 2 種類のデータセットを使用した。そして、読みのアラインメント (並行箇所) の違いによる系統樹の推定結果の相違点を確認した。同時に、異読距離については、異読の並行箇所の文字の不一致度から測る手法、異読の文字の違いに基づいた計算法の 2 つの手法を適用した。これらの校合様式と異読距離計算法を組み合わせた距離行列データセットを作成し、Roos & Heikkila (2009) のベンチマークテストを参考にして系統樹推定を行った。そして、本当の系統関係と推定系統樹とを比較しながら、系統樹の精度を調べ、それぞれの異読距離が持つ特徴について考察した。系統樹推定には、階層的クラスター分析と近接結合による系統樹推定を行い、正解系統樹と比較しながら精度を求めた。精度の結果から、異読並行・距離計算法に関わらず、ベンチマークテストの報告と同様に、近接接合法の適用が高い精度であること確認した。編集距離による異読距離行列に関しては、異読並行様式により精度が大きく変化することを確認した。

5. 主な発表論文等

〔雑誌論文〕(計6件)

- (1) 三宅 真紀, 異読距離から測る人工写本の系統樹推定 校合様式の違いによる推定結果の比較, 『統計数理研究所共同リポート』, 386, 31-44, 2017.
- (2) 三宅 真紀, 新約聖書写本の類似性に関する予備研究 基準本の違いによる異読の計量分析, 『言語文化研究プロジェクト: 電子化言語資料分析研究 2015-2016』, 55-70, 2016.
- (3) 三宅 真紀, ネットワークによるディケンズ・レキシコンの視覚化 辺媒介性に基づくグラフクラスラリング手法の適用, 『言語文化研究プロジェクト: 電子化言語資料分析研究 2014-2015』, 29-40, 2015.
- (4) Hiroyuki Akama, Maki Miyake, Jaeyoung Jung, Brian Murphy, Using Graph Components Derived from an Associative Concept Dictionary to Predict fMRI Neural Activation Patterns that Represent the Meaning of Nouns, PLOS ONE, 10(4): e0125725, 2015.
- (5) 永崎 研宣, 三宅 真紀, 苜米地 等流, A.Charles Muller, 下田 正弘, 人文学資料としてのテキスト構造化の意義を再考する: 大正新脩大藏經における脚注の解析とLinked Data化をめぐる, 『じんもんこん 2013 論文集』, 239-246, 2013.
- (6) 三宅 真紀, Dickens の言語・文体研究を目的とするオンライン KWIC コンコーダンスの開発, 『言語文化研究プロジェクト: 電子化言語資料分析研究 2012-2013』, 25-32, 2013.

〔学会発表〕(計9件)

- (1) 三宅 真紀, 人工写本 Parzival を利用した校合データ様式と異読距離の考察: 新約聖書写本比較に向けて, 言語研究と統計 2017, 2017年03月28日, 統計数理研究所(東京都立川市)
- (2) 三宅 真紀, 編集距離による新約聖書正文批判研究の試み Profile Method との比較, 言語研究と統計 2016, 2016年03月15日, 統計数理研究所(東京都立川市)
- (3) Maki Miyake, Developing an interactive web application for comparing minor differences between modern critical editions of the New Testament, Digital Humanities at Oxford Summer School, University of Oxford, 2015年07月20日, University of Oxford (Oxford, UK).
- (4) 三宅 真紀, ネットワーク図からみるディケンズ・レキシコン "word(s)" を含む見出し語を中心にして, 言語研究と統計 2015, 2015年03月23日, 統計数理

研究所

- (5) 三宅 真紀, これからの DLD の機能について カードの活用と整理, The Dickens Lexicon Digital 公開シンポジウム, 2014年11月08日, 広島文化学園大学
- (6) 三宅 真紀, テキストマイニングの基礎と実践 新約聖書校訂本の比較研究を事例に, 文書研究とデジタル技術(文書ネットワーク共同研究会)招待講演, 2014年09月23日, 国立民族学博物館
- (7) Maki Miyake, Developing a heuristics tool for comparing differences between critical texts of the New Testament, Poetics and Linguistics Association (PALA2014), 2014年07月16日, Slovenia (The University of Maribor)
- (8) Maki Miyake, Investigating Minor Textual Differences Using Approximate String Matching Algorithms, Digital Humanities Australasia 2014, 2014年03月19日, オーストラリア(西オーストラリア大学).
- (9) Maki Miyake, Different Characteristics of Variant Readings Based on Comparison of Major Textual Similarity Measures, Japanese Association for Digital Humanities 2013, 2013年09月17日, 立命館大学.

〔図書〕(計1件)

- (1) 異読距離からみる新約聖書ギリシャ語写本間の相違 マルコ福音書第1章を事例として, 『言語研究と量的アプローチ』石川有香・石川慎一郎・前田忠彦他全7名(編), 金星堂, 74-86, 2016.

6. 研究組織

(1) 研究代表者

三宅 真紀 (Miyake Maki)

大阪大学・言語文化研究科言語文化専攻・准教授

研究者番号: 80448018