

平成 22 年 5 月 14 日現在

研究種目：若手研究(B)  
 研究期間：2007～2009  
 課題番号：19700238  
 研究課題名(和文) 数理文献学的研究における新約聖書セマンティック Web の半自動構築  
 研究課題名(英文) The semi-auto development of New Testament semantic web based on the mathematical philology  
 研究代表者  
 三宅 真紀 (MIYAKE MAKI)  
 大阪大学・言語文化研究科・准教授  
 研究者番号：80448018

研究成果の概要(和文)：本研究は、新約聖書解釈学の問題に対して、数理文献学的手法に基づき、計量的な視点からの新規な解釈の方法を提示した。マタイ・マルコ・ルカ福音書の相互関係を扱う「共観福音書問題」に取り組み、テキストの類似概念の抽出や潜在的な意味構造の解析に努めた。ネットワーク分析の応用研究として、福音書ソーシャルネットワークを作成し、登場人物の相互関係や文書間の構造の特徴について考察した。ネットワーク構造の体系化は、インタラクティブに視覚化した web アプリケーションを構築した。

研究成果の概要(英文)：The primary objective of the conducted study was to develop a mathematical formulation of the mutual relationships between the first three Gospels in order to elucidate the nature of their interdependences and to gain insights into how the Gospels were actually created. This study reports on the application of some network analyses to the four social networks of the Gospels of the New Testament. In particular, the study discusses the interactions between story characters and features of the community structures. The results for the social networks are implemented as graph network structures through the development of a web-based application for interactive data visualization.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	600,000	0	600,000
2008年度	800,000	240,000	1,040,000
2009年度	700,000	210,000	910,000
総計	2,100,000	450,000	2,550,000

研究分野：数理文献学, コーパス言語学

科研費の分科・細目：情報学・情報図書館学・人文社会情報学

キーワード：文学情報システム, セマンティック Web, 新約聖書学

## 1. 研究開始当初の背景

情報化が推進され、今や多種多様なあらゆる情報を Web 上から閲覧することが可能である。これらの膨大な情報の中から、的確な情報を抽出するための検索技術や情報共有

のあり方が重要視され、その技術の進展も目覚ましい。中でも、次世代 Web として注目されるセマンティック Web は、コンピュータに情報の意味を理解させ、情報の自動処理を行うことにより、Web 上の情報の共有や再利

用の自動化が可能になり、適切な情報検索に重要な技術である。

聖書学の領域においても、情報化の流れを受けて、電子化テキストやシソーラスの Web 閲覧が可能となり、聖書学研究のコミュニティが形成されている。新約聖書の電子化版としては、Digital Nestle-Aland があり、単語毎に詳細な写本情報がある。

テキスト解釈において、意味論(semantic)的研究は重要な要素であり、語彙辞書の電子化やデータベース (DB) の構築が盛んである。旧約聖書の古典ヘブライ語に関しては、Leiden 大学を拠点とする大規模なプロジェクト (SAHD) があり、セマンティックス DB 化が進められている。また、Louw-Nida の語彙辞典は、BibieWorks や Logos の主要な聖書ソフトウェアにも組み込まれており、Semantic 領域による研究の重要性が分かる。さらに、セマンティック Web に関しても SemanticBible(2003)による約聖書の名詞オントロジ(NT Names)が公開されている。

グラフ理論に基づいたネットワーク分析は、膨大で複雑なネットワーク体系を直感的に把握しやすい方法論である。テキスト分析においては、頻度数が Zipf (1965)の法則に従うことが知られているが、大規模な言語知識資源における単語や単語グループ (概念) の関係性を理解するために、ネットワークを適応してデータの構造を体系的に把握することは重要である。最近では、Steyvers & Tenenbaum (2005)は、英語コーパスによる大規模な意味ネットワークが複雑ネットワークの性質と共通することを示し、グラフ理論の有用性を唱えている。

## 2. 研究の目的

本研究は、単語のグラフ情報を基にし、概念クラスタリングを新約聖書のテキストに適用して、計算に適用する計量データセット作成とクラスタリング計算の処理を区切り、半自動的に体系を形成することを目的とする。そして、データ別による体系化を比較しながら、単語と単語 (概念) の関係を示すような意味ネットワークを構築する適切な方法論を研究し、新約聖書の体系化をネットワーク図として描画し、Web アプリケーションを構築する。

グラフクラスタリング手法に関しては、Van Dongen (2000) が提唱した MCL (Markov Clustering) や Jung ら(2006)による、MCL の言語データへ適用で起こる問題を解消するために提案した RMCL(Recurrent Markov Clustering)を使用する。そして、結果の表示方法に関しては、聖書学研究者が扱いやすいように、単語や章のレベルに還元したネットワーク図で表示する。さらに、グラフ描画に関しても、現時

点で使用している、GraphViz による静的な図示や Mathematica のグラフ表示機能から発展して、インタラクティブなネットワーク表示ができるグラフ描画ツールを新たに導入し、大規模なネットワーク図を表示可能にする。また解析結果に対しては、聖書学者の意見を適時取り入れながら、研究の改善に生かす。

## 3. 研究の方法

分析データの作成は、新約聖書学に倣い、古典ギリシャ語の Nestle-Aland のテキストを用いる。同時に、古典ギリシャ語を解さない専門家以外によるシステム評価を考慮して、BibleWorks が提供する辞書から対応表を作成し、英対訳表示も可能にする。また、分析データは、単語の出現頻度数に加えて、共起語数やグラフ指標 (次数、クラスタリング係数) から選定したデータなど、様々なデータセットを用意して、より文献学的作業に近く、文書の内容に踏み込んだ概念グループ結果を採択する。

新約聖書のテキストの XML 化、ならびにグラフクラスタリングの結果によるネットワーク描画により、新約聖書の体系化を実現する。そして、単語や概念レベルでの無向グラフ図として、福音書意味ネットワークを構築し、Web アプリケーションとして公開する。さらに応用分析として、福音書の特徴付けるような単語や概念を抽出して、各文書の特徴を考察する。

また、本研究の成果を、文理融合研究の学会 (Digital Humanities, Psychometric Society) で発表を行い、本研究のグラフ理論の応用研究について議論を交わすことによって、さらなる発展に努める。

## 4. 研究成果

本研究は、新約聖書のテキストの XML 化、ならびにグラフクラスタリングの結果によるネットワーク描画により、新約聖書の体系化を試みた。グラフクラスタリング手法に関しては、Van Dongen (2000) が提唱した MCL (Markov Clustering) や Jung ら(2006)による、MCL の言語データへ適用で起こる問題を解消するために提案した RMCL(Recurrent Markov Clustering)を使用した。さらに、クラスタリングの結果表示に関しては、インタラクティブなネットワーク表示ができるグラフ描画ツールを導入し、大規模なネットワーク図を表示可能にするようなツールの検討を行った。

言語資源として、新約聖書の他にも、小説 (Le petit prince) や Joyce 氏による日本語連想語データベースを使用し、出現頻度数をデータとする多変量解析によるテキスト間の類似度計算に加えて、単語をノードで表



- Statistique des Données Textuelles*, Lyon, Presses Universitaires de Lyon (JADT2008), 査読有, Vol. 1, 2008, pp. 57-68.
- (9) Terry Joyce, Maki Miyake, Capturing the Structures in Association Knowledge: Application of Network Analyses to Large-Scale Databases of Japanese Word Associations, *In Proceedings of the Symposium on LKR2008*, 査読有, 2008, pp. 116-131.
- (10) Hiroyuki Akama, Maki Miyake, Jaeyoung Jung, How to Take Advantage of the Limitations with Markov Clustering?--The Foundations of Branching Markov Clustering (BMCL), *In Proceedings of the IJCNLP-2008*, 査読有, 2008, pp. 901-906.
- (11) 赤間 啓之, 三宅 真紀, 鄭 在玲, グラフクラスタリングを用いた、歴史事象のシミュレーションの可能性について、『人文科学とコンピュータシンポジウム論文集』、査読有、2007、pp. 277-282.
- (12) 三宅 真紀, 共起単語の選定を目的とするグラフクラスタリング評価の考察、『人文科学とコンピュータシンポジウム論文集』、査読有、2007、pp. 177-182.
- (13) Maki Miyake, Terry Joyce, Jaeyoung Jung, Hiroyuki Akama, Hierarchical Structure in Semantic Networks of Japanese Word Associations, *In Proceedings of the 21st Pacific Asia Conference on Language, Information and Computation (PACLIC21)*, 査読有, 2007, pp. 321-328.
- (14) Hiroyuki Akama, Jaeyoung Jung, Maki Miyake, Kenji Yoshihashi, Kikuko Nishina, Building a clustered semantic network for an Entire Large Dictionary of Japanese, *In Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics*, 査読有, 2007, pp. 308-316.
- (15) Maki Miyake, Terry Joyce, Mapping out a Semantic Network of Japanese Word Associations through a Combination of Recurrent Markov Clustering and Modularity, *In Proceedings of the 3rd Language & Technology Conference (L&TC'07)*, 査読有, 2007, pp. 114-118.
- (16) Hiroyuki Akama, Maki Miyake, Jaeyoung Jung, Remaking the Markov Clusters of Word Data, Using the Example of "The Little Prince" by Saint-Exupery, *SIG Technical Report*, 査読無, 2007, pp. 33-40.
- (17) 赤間啓之、鄭在玲、三宅真紀、近代ストア主義とメスマール主義の思想的類似性に関するグラフ言語学的分析、『情報処理学会研究報告』、査読無、Vol. 49. 2007、pp. 49-56.
- (18) 三宅真紀、ネットワーク分析におけるテキストマイニングークラスタリング係数-Markov Clusteringを中心として、『言語文化研究プロジェクト、電子化言語資料分析研究』、査読無、2007、pp. 3-16.
- (19) 赤間啓之、鄭在玲、三宅真紀、グラフ理論を利用した人文解釈学の可能性、『文理シナジー学会春季大会』、査読無、2007、p. 12.
- (20) Maki Miyake, Terry Joyce, Analysis of the semantic network structure of Japanese word associations, *In Proceedings of the IMPS2007*, 査読有, 2007, p. 22.
- (21) Jaeyoung Jung, Maki Miyake, Hiroyuki Akama, Developing a Web-based Free Association Retrieval System Applying Graph Theory, *In Proceedings of the World Conference on ED-MEDIA*, 査読有, 2007, pp. 3359-3364.
- (22) Maki Miyake, A Network Structure of the Synoptic Gospels Employing Clustering Coefficients, *Digital Humanities 2007*, 査読有, 2007, pp. 137-139.
- (23) 鄭在玲、三宅真紀、赤間啓之、潜在隣接を用いた意味ネットワークの適正なクラスタ化について、『言語処理学会第13回年次大会(NLP2007)』、査読有、2007、pp. 246-249.
- (24) Yumiko Shimizu, Maki Miyake, Hiroyuki Akama, Alta Mira, an Associative Image Retrieval System Based on Knowledge Store, *In Proceedings of the Symposium on LKR2007*, 査読無, 2007, pp. 71-74.
- (25) Maki Miyake, Visualizing Recurrent Markov Clustering, *In Proceedings of the Symposium on Large-Scale Knowledge Resources (LKR2007)*, 査読無, 2007, pp. 243-246.
- (26) 赤間啓之、鄭在玲、三宅真紀、グラフクラスタリングを用いたツシユールの概念ネットワーク解析、『情報処理学会研究報告』、査読無、Vol. 9、2007、pp. 33-40.

[学会発表] (計 9 件)

- (1) Maki Miyake, Investigating collocational variation in Charles Dickens through network analysis, *The*

29th International Conference of the Poetics and Linguistics Association 2009 conferences, 2009年7月30日, オランダ・Middleburg.

- (2) Maki Miyake, Capturing the Social Networks of the Gospels through a Graph Clustering, *Digital Humanities 2009*, 2009年6月25日, 合衆国・Maryland.
- (3) Maki Miyake, An Investigation of the Community Structures within Social Networks of the Gospels, *2008 Chicago Colloquium on Digital Humanities and Computer Science*, 2008年11月2日, 合衆国・シカゴ
- (4) Maki Miyake, Investigating word co-occurrence selection with extracted sub-networks of the Gospels Employing Clustering Coefficients, *Digital Humanities 2008*, 2008年6月27日, フィンランド・オウル大学
- (5) Maki Miyake, L'élaboration d'un réseau sémantique par le raffinement du Markov Clustering--A partir des données lexicales du roman de Saint-Exupéry, "Le petit prince", 9èmes JADT2008, 2008年3月13日, フランス・リヨン.
- (6) Maki Miyake, Mapping out a Semantic Network of Japanese Word Associations through a Combination of Recurrent Markov Clustering and Modularity, The 3rd L&TC'07, 2007年10月5日, ポーランド・ポズナニ.
- (7) Maki Miyake, Terry Joyce, Analysis of the semantic network structure of Japanese word associations, The 15th International Meeting and the 72nd Annual Meeting of the Psychometric Society, 2007年7月10日, 東京.
- (8) Maki Miyake, A Network Structure of the Synoptic Gospels Employing Clustering Coefficients, *Digital Humanities 2007*, 2007年6月7日, Urbana-Champaign, USA.
- (9) Maki Miyake, Visualizing Recurrent Markov Clustering, The Symposium on LKR2007, 2007年3月3日, 東京.

[その他]

ホームページ等

<http://www.lang.osaka-u.ac.jp/~mmiyake/>

三宅 真紀 (MIYAKE MAKI)

大阪大学・言語文化研究科・准教授

研究者番号：80448018

## 6. 研究組織

### (1) 研究代表者