

## 科学研究費助成事業 研究成果報告書

平成 26 年 5 月 27 日現在

機関番号：12608

研究種目：若手研究(B)

研究期間：2010～2013

課題番号：22700256

研究課題名(和文)レトリカルデータベースシステムの構築による計量的修辞分析手法の確立

研究課題名(英文) Establishment of quantitative analysis method for rhetoric based on the development of rhetorical database system

研究代表者

村井 源 (Murai, Hajime)

東京工業大学・社会理工学研究科・助教

研究者番号：70452018

交付決定額(研究期間全体)：(直接経費) 2,900,000円、(間接経費) 870,000円

研究成果の概要(和文)：聖書など古典テキストに特徴的に見られる、構造的な記述(キアスムス、パラレリズムなど)を科学的に分析するための基盤を構築した。まず個々の構造を納めるためのデータ構造を設計しソフトウェアを開発した。次に聖書の構造的な記述を(旧約39巻、新約27巻、旧約続編11巻の、約1500の構造と約7700の対応テキスト箇所)を分析し、データベースに納めた。また分析結果が妥当かを統計的に判定するためのプログラムを開発した。得られたテキスト中の構造のリストはホームページ上([http://www.valdes.titech.ac.jp/~h\\_murai/bible/bible.html](http://www.valdes.titech.ac.jp/~h_murai/bible/bible.html))で公開している。

研究成果の概要(英文)：In classic text such as the Bible, there are many the structural descriptions of texts. Those are called Chiasm or Parallelism. In order to analyze scientifically those structures, the foundation was constructed. At first, design of the data structures was done for storing each structure in database. The software for that database was also developed. Then, structural descriptions in the Bible (about 1500 structures and 7700 pairs of corresponding texts in the 39 books of the Old Testament, 27 books of the New Testament, and 11 books of Apocrypha) were analyzed and those were stored in that database. Moreover, the program for evaluating statistically the validity of the result of analysis was developed. The list of the resultant structures in the text was published on the website ([http://www.valdes.titech.ac.jp/~h\\_murai/bible/bible\\_e.html](http://www.valdes.titech.ac.jp/~h_murai/bible/bible_e.html)).

研究分野：図書館情報学・人文社会情報学

科研費の分科・細目：人文社会情報学

キーワード：聖書 修辞構造 キアスムス 物語構造 計量分析

## 1. 研究開始当初の背景

情報処理技術を用いた自動的なテキストの意味分析は、近年の様々な研究によって大きく進歩してきた。テキストを形態素分割、係り受け分析、機械可読なオントロジーの利用などによって、自動要約や自動分類など様々な応用技術が開発されてきた。しかしこれらの技術が従来主に取り扱ってきたのは、科学の学術論文、新聞、特許文書などの簡潔で逐語的な表現が中心となる比較的分析の容易なテキストコーパスであった。近年 Blog などの Web テキストも対象として扱われるようになったが、その多くはテキスト自体の意味分析ではなく、特定商品の評判分析のように Web をデータソースとして用いて必要な情報をマイニング的に抽出するものである。

一方で、Digital Humanities などと呼ばれる、従来情報技術であまり扱われてこなかった人文系の情報資源を工学的に研究する分野も現れてきてはいる。しかし、抽象的表現と複雑な修辭に満ちた人文系テキストの工学的分析は困難であり、計量文体学やデジタルコンコルダンスなどの、単語レベルでの分析は盛んに行われてきたが、より上位のテキスト構造である命題や文のレベルでの自動的意味分析は、テキストに用いられた修辭法の自動的特定が困難であることなどが原因で立ち遅れている。機械的に修辭構造を記述する枠組みとしては Rhetorical Structure Analysis や TEI などいくつかのメタデータ記述法が対応しているが、これらは人手での分析を前提としている。しかしながら、人文系テキストの意味分析を行う上では、テキスト中の命題や文のレベルでの意味的構造や、様々な修辭法の定量的な分析が必要不可欠である。このため、テキスト内の修辭構造に基づく意味関係や、一つのセンテンスに含まれる複数の命題間の関係性などの修辭的な情報を科学的・計量的に分析する手法の開発が望まれる。

このような計量的・自動的な情報処理技術による妥当な意味分析を実現するためには、人間が従来行ってきたテキストの読解・解釈結果との相違の定量的な検証が必要となると考えられる。しかしながら、Digital Humanities におけるテキスト電子化プロジェクトの主眼は対象テキストの電子化と効率的利用であり、対象テキストを人間が解釈した結果である関連テキストの収集はあまり重視されていない。さらに、これら関連テキスト群とプロジェクトの対象テキストの関係性を定量的に分析した研究は少ない。

研究代表者は古典テキストの分野において、テキスト中の小部分間の関係から構築される古典的な修辭構造の分析や、異本テキスト間の相違関係からのプロット抽出などの研究を行ってきており、過去の研究で分析してきたテキストにおける修辭構造のリスト約 4000 件を電子データとして保有している。

一方で計量的なテキスト分析に関しては、Web 上で利用可能なテキスト分析用の汎用サーバクライアントモデルのソフトウェアを開発し現在公開中である。また、これらに合わせて、文学・政治・宗教・芸術分野など様々な人文系のテキストの計量的な分析を行ってきており、人文系テキストの電子化と、計量的に分析する上でのノウハウの蓄積を行ってきた。さらに特定のテキストとそれを人間が解釈した結果生成される関連テキストとの関係性を、引用解析の手法を用いて定量的に分析し、人間がテキストの解釈を行った結果として得られた概念を計量的に提示する研究を進めてきた。

研究代表者はこれらの研究を進める中で、高度な修辭法が多用される人文系テキストでは、単語レベルでの計量分析のみでは、結果の再現性を担保した科学的な分析を実現できる一方で、概して表面的なレベルにとどまりやすく、人間が行う場合のような深いテキスト解釈が行えないことを痛感した。そこで従来の単語レベルの意味分析を土台としつつも、より大きなテキスト中の単位を計量的に扱うために、修辭法・レトリックに関連する修辭的な情報の機械による定量的な分析と人手による高度な分析を、同じフレームワーク内で蓄積し、相違点を明らかにすることで、テキストの自動意味分析における新たな一歩を踏み出すことができるのではないかと思うに至った。

## 2. 研究の目的

研究期間内に行うことは大きく、[1]レトリカルデータベースシステムの開発、[2]文・命題の修辭的情報の分析システムの開発、[3]過去の人文研究によるテキスト解釈結果の電子化と分析、の三つに分けられる。以下に各々の内容を説明する。

[1] レトリカルデータベースシステムの開発  
機械処理を用いて自動的に得られたテキスト内の要素の修辭的な関係性と、人文的手法によって特定されたテキスト内の構造や解釈を並行的に蓄積可能なオープンソースの Web アプリケーションを開発する。複数の解釈が並立することが一般的な人文テキストの実情に合わせて、特定の個所の修辭的情報に対して、相違する複数のデータを入力主体の ID 付きで蓄積することにより、網羅的な解釈の提示と同時に、特定の学派や手法に基づいた場合のテキストの解釈のみを選択的に提示することも可能とする。また、近年広く成果を上げつつある集合知の手法を取り入れて、ID を取得すれば誰でも自由にデータの蓄積が可能なオープンなシステムとして公開する。

[2] 文・命題の修辭的情報の分析システムの開発

テキスト中に含まれる、文や命題などのテ

キストの小部分の、他の文・命題との関係性を自動的に推定するアルゴリズムを開発する。具体的には、係り受け分析の結果に基づき命題を抽出し、テキスト中での他の命題との関係性を、オントロジー等の機械可読辞書による表現分析や、人手による分析結果の機械的学習に基づいて推定する。また命題の結合としての各文のテキスト中での相互の関係性も同様に、多様な表現の機械可読辞書による分析と、学習データに基づいて推定を行う。機械学習用のデータとしては、[3]によって得られるテキストの人手による内容分析の結果を用いる。

### [3] 過去の人文研究によるテキスト解釈結果の電子化と分析

過去の人文研究によって得られたテキスト解釈結果を、解釈の対象となるテキストと合わせて電子化し、[2]で得られる機械による自動推定の結果と合わせて[1]で得られるレトリカルデータベースシステムへと入力する。入力された人文的成果のデータは、計量的な分析によって機械的に得られた[2]の推定結果と比較され、その相違を[2]の研究へフィードバックして機械による自動推定の精度向上を図ると同時に、人間によるテキスト解釈プロセスのアルゴリズム的な明示化をはかる。

### 3. 研究の方法

修辭的情報を扱うためのデータベースシステムを、すでに開発しているサーバクライアントモデルのテキスト分析ソフトウェアをベースにして開発する。データ形式は一般的に用いられる意味タグ付きテキスト形式ではなく、対象となる部分テキストを自由にかつ並行的に指定可能なレトリカルデータベースに基づくものとする。このようにすることで、人文系のテキストに多々見られる多層的な意味解釈、異なる視点に基づく複数の意味解釈の並立を機械的に処理可能となる。開発されたシステムに従来の人文研究で得られたデータを入力し、さらに集合知の利用により多くのデータの蓄積を目指す。また、修辭的情報の自動分析を行うために、まず文の命題への自動分割と命題間の関係性の自動推定を行うアルゴリズムを、人手による修辭的情報の分析結果の機械学習に基づいて開発する。人手と自動分析によって得られた修辭的情報の相違を分析し、人間の解釈のメカニズムを究明すると同時に自動テキスト解釈のための基礎データを得る。

### 4. 研究成果

#### [1]ソフトウェア開発

レトリカルデータベースシステムを開発するうえで、研究代表者らがすでに開発したWeb上で動作するサーバクライアントモデル

のテキスト分析ソフトウェアやクライアントモデルのテキスト分析ソフトウェア(Digital Hermeneutics System)を土台とした。これらの二つのソフトウェアは何れもリレーショナルデータベースを用いた構造であるが、新規に開発するレトリカルデータベースと、すでに構築した古典テキストのデータベースシステムが連動するシステムを構築した。また、これに合わせてレトリカルデータベースの閲覧画面を付加した。

また開発したデータベースシステムに古典的修辭構造蓄積を行った。具体的にはすでに分析していた修辭構造を中心にデータを入力し、特定のテキスト箇所ごとの修辭構造の一覧と、特定の修辭構造における関連テキスト箇所の一覧が表示できるようになった。また、データベースの操作によって得られた一覧のリスト部分をHTML化して研究代表者ホームページ上にて公開を行った。

#### [2]レトリカルデータの蓄積と分析

テキストの修辭的構造の中で、古典的な修辭構造(新約・旧約聖書)のデータ化を進め、日本語版のデータに合わせて英語版の作成を進めた。テキスト全体の大きなサイズの修辭構造に合わせて、テキスト中の小物語区分の中に存在する文や命題を単位とする修辭構造の分析とデータベース化を進めた。旧約聖書中39巻、新約聖書中27巻、旧約聖書続編11巻分のテキスト全体に亘る修辭構造の分析を終え、約1500の修辭構造と約7700の対応テキスト箇所の対を抽出した。またテキストの小物語区分(pericope)における修辭構造を約1900抽出できている。また、研究代表者ホームページ([http://www.valdes.titech.ac.jp/~h\\_murai/bible/bible.html](http://www.valdes.titech.ac.jp/~h_murai/bible/bible.html))上にテキスト全体に亘る修辭構造に合わせて、命題や文やフレーズの単位からなるそれらの小さいレベルの修辭構造も公開を行った。

また上述の蓄積された修辭構造の妥当性を数値的に検証するためのアルゴリズムの検討を行い、ランダムなテキスト間の組み合わせにランダムなテキストの分割位置に基づく修辭構造作成のプログラムを実装し、検討対象修辭構造との比較を行った。結果としてほぼすべての、テキスト全体に亘るレベルの修辭構造に関しては5%有意水準での有意差が複数種類の単語やフレーズの組み合わせに関して抽出された。また既に得られた種々の修辭構造は聖書全巻(一部除く)を通じて非常に規則的で統一的な形式となっていることが明らかとなった。形式的な体系性と数値的妥当性より検討対象修辭構造は著者あるいは最終編集者の意図した構造であると結論できる。また修辭構造全体だけでなく、それらを構成する個々の部分に関する統計学的な有意性を分析するアルゴリズムの検討を行っている。

[3]古典的修辞構造の対応関係の分析に向けての語彙分析

古典的な修辞構造の小物語区分間の対応関係进行分析するため、複数の翻訳間やテキスト内の語彙の対応関係の揺らぎを計量的に抽出することによって翻訳と原語間での語彙のネットワークを構築した。構築した語彙ネットワークは類似の語義の単語をまとめたり、単語が示す概念の関係性を計量的に図る指標として用いることが可能であり、小物語区分間の類似性を判定する一つの基盤として利用可能と考えられる。

[4]物語構造の分析とデータの蓄積

古典的修辞構造に合わせてより大きい単位である物語構造もプロットの物語理論に基づいた分析を星新一のショートショート全集に対して行いデータ化した。またデータ化された物語構造に対して計量的な分析を施して物語パターン特徴を抽出した。

また、村上春樹の長編小説をデータとして単語の出現位置の差異から物語構造を抽出した。

[5]文・命題の自動分析

映画・演劇の批評テキストを対象に、何に対する批評であるかという分類を、対象を特徴的に示す単語に基づいて、文単位で自動的に行う分析を実施した。分析の結果、映画と演劇での批評における視点の相違や批評対象間の関係性の相違が数値的に明らかになった。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 7 件)

全件 査読有

1. Hajime Murai, “Exegetical Science for the Interpretation of the Bible: Algorithms and Software for Quantitative Analysis of Christian Documents”, Springer International Publishing, Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing Studies in Computational Intelligence, Volume 492, pp. 67-86, 2013.
2. 村井源, 川島隆徳, “総合芸術の批評における批評対象の特徴—映画と演劇の批評の計量的比較—”, 情報知識学会誌, Vol. 22, No. 3, pp. 203-222, 2012.
3. 工藤 彰, 村井源, 往住 彰文, “共通語の布置と変化に基づく並行形式小説の物語構造”, 情報知識学会誌, Vol. 22, No. 3, pp. 187-202, 2012.
4. 村井源, 川島隆徳, 工藤 彰, “映画と演劇の批評文における固有名の関係性と

役割の計量分析”, 情報知識学会誌, Vol. 22, No. 1, pp.23-43, 2012.

5. 村井源, 松本 斉子, 佐藤 知恵, 往住 彰文, “物語構造の計量分析に向けて—星新一のショートショート—の物語構造の特徴—”, 情報知識学会誌, Vol. 21, No. 1, pp.6-17, 2011.
6. 工藤 彰, 村井源, 往住 彰文, “計量分析による村上春樹長篇の関係性と歴史的変遷”, 情報知識学会誌, Vol. 21, No. 1, pp.18-36, 2011.
7. 村井源, “漸近的对応語彙推定法に基づく翻訳文の解釈的特徴の抽出—日本語翻訳聖書の計量的比較—”, 情報知識学会誌, Vol. 20, No. 3, pp.293-310, 2010.

[学会発表](計 21 件)

全件 査読有

1. 村井源, “登場人物の言動に着目した物語プロットデータベースのためのデータ構造試案”, 人文科学とコンピュータシンポジウム論文集, IPSJ Symposium Series Vol. 2013, No. 4, pp. 255-260, 2013/12/14, 京都大学.
2. 高橋 椋一, 村井源, 猪原 健弘, “脚本執筆における概念構造の抽出—映画・演劇脚本術の計量的比較—”, 人文科学とコンピュータシンポジウム論文集, IPSJ Symposium Series Vol. 2013, No. 4, pp. 1-6, 2013/12/12, 京都大学.
3. Hajime Murai, “Validation of Hierarchical Rhetorical Structure by Combination of Quantitative Evaluation and Traditional Rhetorical Analysis”, Proceedings of JADH 2013 Conference, pp. 50-51, 2013/9/20, Ritsumeikan University.
4. 松原 仁, 佐藤 理史, 赤石 美奈, 角 薫, 迎山 和司, 中島 秀之, 瀬名 秀明, 村井源, 大塚 裕子, “コンピュータに星新一のようなショートショートを創作させる試み”, An attempt at automatic composition of Shin'ichi Hoshi-like short short stories, The 27th Annual Conference of the Japanese Society for Artificial Intelligence, 2D1-1, 2013/6/5, 富山国際会議場.
5. 村井源, “物語プロットデータベースのためのデータ構造の検討”, 情報知識学会第21回年次大会, Vol. 23, No. 2, pp. 308-315, 2013/5/26, お茶の水女子大学.
6. 村井源, “漸近的对応語彙推定法を用いた異言語テキスト間の概念ネットワーク”, 人文科学とコンピュータシンポジウム論文集, IPSJ Symposium Series Vol. 2012, No. 7, pp. 61-68, 2012/11/17, 北海道大学.
7. Hajime Murai, “Evaluating Validity of Classic Rhetorical Structure by

- Comparing with Random Division and Random Combination”, Proceedings of JADH 2012 Conference, p. 5, 2012/9/16, Tokyo University.
8. Hajime Murai, “Introducing Scientific Methods for the Interpretation of the Bible: Quantitative Analysis of Christian Documents”, 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, pp.391-398, 2012/8/8, Kyoto Campus Plaza.
  9. 村井源, “デジタルアーカイブを用いた古典修辞構造検証手法の検討について”, 人文科学とコンピューターシンポジウム, ISPJ Symposium Series Vol. 2011, No. 8, pp. 211-218, 2011/12/11, 龍谷大学.
  10. 工藤彰, 村井源, 往住彰文, “村上春樹『1Q84』における因子分析を用いたチャプターの特徴と共起ネットワーク”, 人文科学とコンピューターシンポジウム, ISPJ Symposium Series Vol. 2011, No. 8, pp. 133-140, 2011/12/10, 龍谷大学.
  11. Hajime Murai, “Implementing Rhetorical Structure Database System for Digital Archive”, Osaka Symposium on Digital Humanities 2011, p. 38, 2011/9/14 Osaka University.
  12. 村井源, 川島隆徳, 工藤彰, “映画と演劇の批評文における記述対象の分類”, 日本認知科学会第28回大会, pp. 314-319, 2011/9/23, 東京大学.
  13. 工藤彰, 村井源, 往住彰文, “村上春樹12長篇の構文解析を用いた登場人物と行動の関係性”, 日本認知科学会第28回大会, pp. 685-692, 2011/9/23, 東京大学.
  14. 村井源, 川島隆徳, 工藤彰, “映画と演劇の批評文における記述対象の計量的比較”, 情報知識学会第19回年次大会, Vol. 21, No. 2, pp. 279-284, 2011/5/29, 香川大学.
  15. 工藤彰, 村井源, 往住彰文, “小説の進行を特徴づける指標としての動詞生起頻度”, 情報知識学会第19回年次大会, Vol. 21, No. 2, pp. 285-290, 2011/5/29, 香川大学.
  16. 村井源, “デジタルアーカイブへのIntra-textualityの導入に向けて”, 人文科学とコンピューターシンポジウム, ISPJ Symposium Series Vol. 2010, No. 15, pp. 131-136, 2010/12/12, 東京工業大学.
  17. 工藤彰, 村井源, 往住彰文, “村上春樹の計量的変遷と共時的フィクションの語彙形成”, 人文科学とコンピューターシンポジウム, ISPJ Symposium Series Vol. 2010, No. 15, pp. 97-102, 2010/12/11, 東京工業大学.
  18. Hajime Murai, “The Parallel Concentric Structures within Exodus”, Society of Biblical Literature Annual Meeting, 2010/11/20, Atlanta.
  19. 村井源, 往住彰文, “文芸批評の計量解析による批評行為の背景的特徴の抽出”, 情報知識学会第18回年次大会, Vol. 20, No. 2, pp. 117-122, 2010/5/16, 東京大学.
  20. 佐藤知恵, 村井源, 往住彰文, “星新一ショートショート文学の物語パターン抽出”, 情報知識学会第18回年次大会, Vol. 20, No. 2, pp. 123-128, 2010/5/16, 東京大学.
  21. 工藤彰, 村井源, 往住彰文, “計量分析による村上春樹文学の語彙構成と歴史的変遷”, 情報知識学会第18回年次大会, Vol. 20, No. 2, pp. 135-140, 2010/5/16, 東京大学.
- 〔その他〕  
ホームページ等  
[http://www.valdes.titech.ac.jp/~h\\_murai/bible/bible.html](http://www.valdes.titech.ac.jp/~h_murai/bible/bible.html)
6. 研究組織  
(1) 研究代表者  
所属機関：東京工業大学  
部局：大学院社会理工学研究科  
職名：助教  
村井源 (Murai Hajime)
- 研究者番号：70452018