

平成 30 年 5 月 10 日現在

機関番号：20103

研究種目：若手研究(B)

研究期間：2014～2017

課題番号：26730168

研究課題名(和文) 修辞・物語構造のデータベース化による人文テキストの計量的分析基盤の構築

研究課題名(英文) Development of a foundation for quantitative analysis for humanities text by utilizing database of rhetoric and story structure

研究代表者

村井 源 (Murai, Hajime)

公立はこだて未来大学・システム情報科学部・准教授

研究者番号：70452018

交付決定額(研究期間全体)：(直接経費) 2,500,000円

研究成果の概要(和文)：約300の星新一のショートショート作品を対象として物語の構造をコンピュータから分析可能な形で記述する手法を開発した。このために物語言動辞書・物語主体辞書・物語慣用表現辞書などを構築した。物語構造のデータ化の結果に基づいて各物語のオチを自動的に抽出するシステムの研究を行った。また新旧約聖書の修辞構造をデータ化するフォーマットを開発し1507種類7715対修辞構造のデータ化を行いその妥当性を統計的に検証した。

研究成果の概要(英文)：A methodology to describe story structures in the computer readable format was developed towards about 300 micro-fictions of Shin'ichi Hoshi. In order to that, Story Behavior Dictionary, Story Agent Dictionary and Story Idiomatic expression Dictionary were constructed. Based on story structure data of those results, automatic punchline extraction algorithm have been researched. Moreover, a data format for the literary structures of the Bible was developed and 7715 text pairs in 1507 literary structures were digitized. The validity of those structures was examined statistically.

研究分野：デジタルヒューマニティーズ, 計量文献学, 物語論

キーワード：物語構造 修辞構造 計量分析 ショートショート 聖書

## 1. 研究開始当初の背景

[1] テキスト電子化の進展による人文学的なテキストの計量的意味解釈の可能性

近年、情報処理技術と情報化社会の進展によりテキストの電子化が著しく進展している。ただし人文学的なテキストの分析においては、古典的な計量書誌学的な指標に基づく著者推定などが中心的であり、人文学で重要なテキストの内容自体の意味解釈に踏み込んだ分析は一般的ではない。この理由として、文学・詩・思想書などの様々な人文学的なテキストは、種々の修辞技法を凝らした文章表現や複雑な物語構造の多用、語られるテーマの抽象性などから取り扱いが難しいことが挙げられる。そのため、人文学的なテキストをコンピュータで分析可能にするには従来のテキストマイニングとは異なる分析手法や切り口が必要となる。人文学的なテキストにおいても計量的な指標に基づく意味解釈の手法を実現することができれば、検索や分類などの工学的な応用はもとより、計算機を用いた大規模で精密かつ客観的な解析によって人文学自体の発展にも大きく寄与することが可能と考えられる。

[2] 人文学的なテキストの計量的解釈に向けた現在までの取り組み

人文学的なテキストの分析や意味解釈を機械的に実現するためのアプローチとして、人間が行ってきた意味解釈行為をアルゴリズムとデータによってコンピュータ上へと実装することは有望な方法論であると考えられる。人文学的なテキストのうちで文学や物語形式の小説などの作品をそれ自体として解釈する従来の手法としては、大きく分けて表現の良し悪しについて語る文体論と話の展開の良し悪しについて語る物語論[1,2]がある。代表者はこれまで小説や思想書などの人文学的なテキストの分析に関して、計量書誌学および計量文体学的方法論に基づく構造の抽出や計量的な指標に基づく解釈などの手法の開発を進めてきた。しかし、計量文体学の対象は基本的に単語や語彙のレベルであり、人文学的な文体論が包含する様々な修辞技法はまだ計量的な分析が困難である。また、人文学的なテキストを理解する上でのもう一つの重要な要素である物語論に関しては、計量的な分析は一部を除いてほとんど手付かずの状態にある。修辞や物語の分析を計量化するためには計算機でも処理可能な明確な表記やデータ構造が必要であり、また自動的な分析を実現可能にするためには、分析の基準となる規範データも必要である。しかし現状では応用性の高いデータ構造も、分類アルゴリズム等に利用可能な学習や分析用の基本データもない状態である。人文学的なテキストの科学的な分析を発展させるためにはコンピュータ将棋で言えばすぐれた棋譜の電子化とデータベース化に相当するような基礎的な作業が必要不可欠な段階と言える。

## 2. 研究の目的

研究期間内に行うことは大きく、[1]修辞・物語の人文的な構造に適した機械的なデータ構造の分析と策定、[2]修辞・物語構造のデータベースシステム開発、[3]修辞・物語構造のデータベースを用いた計量的意味解釈手法の開発、の三つに分けられる。以下に各々の内容を説明する。

[1] 修辞・物語の人文的な構造に適した機械的なデータ構造の分析と策定

修辞技法を含んだ、より高度な文体論的意味解釈、および物語論的方法論のデータとアルゴリズムによる実装を実現するためには、第一段階として、修辞と物語をデータ化し計量的に扱うことを可能にする必要がある。また実際に分析を進めるためには修辞と物語を電子的に扱うためのデータベースの仕様の策定とデータの収集による実装が必要となる。代表者はすでに星新一のショートショート作品を電子テキスト化し、星新一作品に特徴的と言われる逆転や暗転などの物語の起伏とオチのパターンに着目した物語分析を行ってきた。また修辞構造のデータ化に関しても聖書における古典的修辞構造を対象にデータ構造の策定を進めてきている。本研究課題ではこれらの知見を踏まえ、修辞と物語の構造に関して、計量的な分析に適した形でのデータ構造の策定を行っていく。

[2] 修辞・物語構造のデータベースシステム開発

代表者は古典的思想テキストの修辞構造に関して意味解釈に利用可能なデータ構造とデータベースをすでに開発してきており、本研究においてはそれらを基礎として、物語構造のデータと修辞構造のデータを相互参照可能な形でデータベース実装を行う。データベースにおいては類似する構造や関連する構造の検索機能を備えて、従来は困難であった網羅的な修辞学・物語論の研究の基礎的な資料として利用可能な形での公開を目指す。

[3] 修辞・物語構造のデータベースを用いた計量的意味解釈手法の開発

データベース中のデータは人間が検索等で利用できるだけでなく、自動的な構造抽出、構造分析結果の妥当性評価、過去の作品との関連性、同時代での特徴や相違点などを機械的に分析可能とする。

人文学の研究者が意味解釈の参照用データとして利用するためのインターフェースを備えて、かつ計量的な分析の機能を付与することで、人間の側からも機械の側からも利用可能なデータベースとしての提供を目標とする。これらによって人間の専門家の洞察と機械的な精密さの融合を図る。

### 3. 研究の方法

[1]重複部分を持つ修辞・物語構造記述のためのリレーショナルデータベース方式  
 テキスト構造の記述ではタグによる入れ子構造が一般的であったが、入れ子構造ではテキスト中に様々な形で重複的に配置される多様な要素とその解釈を自由にデータ化できない。そのため、タグによる入れ子構造を廃し、テキストの中の位置指定に合わせて当該個所の修辞・物語的な機能をデータとして記述するリレーショナルデータベース形式を採用する。この方式により様々な技法が多層的に配置されたような高度なテキストに対しても取り扱いが可能となり、また研究者によって解釈の異なる修辞や物語の構造を柔軟に並行的にデータベース化し、利用者が比較考察することも実現できる。

[2]各物語要素への属性の多重的記述  
 物語構造を要素に分けて記述する手法にはプロップ[1]やグレマス[2]などの古典的な人文学的分析手法があるが、プロップは特定ジャンルに特化しており、グレマスは要素を6種類に単純化しており、汎用的な反面登場人物の行動の動機や感情状態を記述できない欠点を持つ。これらの伝統的手法では様々な形態の物語分析に必要な特徴が十分に含まれているかに関して疑問が残る。ただし一方で、物語に含まれるすべての要素の記述することは実質的に不可能である。そこで、現段階で必要性が明白な物語構造の属性からデータ化を初め、分析の発展に応じ随時必要となった属性を追加する形で行うことが現実的解法と考えられる。

代表者は星新一作品における特徴的な物語展開をデータ化し、その時代による変遷やテーマごとの物語展開の差異などを抽出してきている。これらの知見をもとに作成した、登場人物の言動を単位とする必要な物語機能の現段階でのリストを表1に示す。

表1 星新一ショートショートに基づく物語機能の属性による分類例

領域		極性	一人称	二人称	三人称
対象	情報	肯定	調査・発見	説明・質問	傍観
		否定	隠蔽	詐欺・嘘	無視
	事物	肯定	作成・発明	贈与	生成
		否定	破壊	奪取	崩壊
自身	存在	肯定	生誕	召喚	登場
		否定	死亡	追放	退出
	身体	肯定	成長	治癒	強化
		否定	衰退	危害	弱化
状況	移動	肯定	移動	送迎	輸送
		否定	静止	拘束	保管
	環境・関係	肯定	安全	友好	平安
		否定	危険	競争・喧嘩	動乱
意図	秩序・約束	肯定	遵守	契約	制定
		否定	違反	不履行・破棄	廃止
	指示・支配	肯定	依頼	許可	提案
		否定	制止	禁止	批判
	目標	肯定	努力	助力	祈り
		否定	放棄	妨害	呪い
評価	評価	肯定	自信	称賛・昇進	好評
		否定	自己否定	侮蔑・降格	不評

表1では物語の機能を登場人物の言動から規定して、言動に領域(種類)と極性(肯定・否定)と人称(主体的・客体的・傍観的)を属性として付加した場合での物語要素の機能の分類例を示している。他にも例えば各物語要素に対する登場人物の感情状態や、登場人物の言動に対する結果の表記を加えることで多様な側面から物語構造の分析が可能となる。

[3]多重的な属性間の計量的な比較と検討からの修辞・物語構造の抽出  
 修辞・物語構造のデータ化により、作者が想定したであろう様々な属性間の関係性を、計算機を活用して網羅的に検証し、多様な角度から類似度や関連性を抽出していくことが計量的テキスト分析の強みであると考えられる。図1は星新一の『契約者』の分析例であるが、物語の登場人物の類似の言動の対応関係が修辞構造(場面1と5, 2と4の対応による交差配列法)を示していたり、主客関係の逆転(場面4と5)による物語のオチの構造があることがデータから発見可能である。

人物	場面	1	2	3	4	5	6
悪魔		放棄 魔王/仕事 失敗	契約 男/命 成功		契約 男/命 成功	放棄 魔王/仕事 失敗	侮蔑 男 成功
	男			破棄 悪魔/契約 失敗		契約 悪魔/命 成功	
魔王		依頼 悪魔/仕事 開始					

表中では上段:言動、中段:対象、下段:言動の結果

図1 『契約者』での分析例

修辞・物語の大規模なデータ化によりこれらの物語・修辞構造を計量的に発見し、それらの歴史的・文化的なパターンなどの抽出や分析が可能になると期待される。

[4]広く受け入れられているテキストの活用  
 将来的に他の人文テキストとの比較等で有用となる普遍的な構造を効率よく抽出するため、社会的に広く受け入れられていてかつ短い起承転結のある物語構造や修辞構造に対して分析を行うことを検討している。候補の一つは聖書の中に含まれるたとえ話などの様々な短い物語(Pericope)である。聖書は世界中で広く長い期間にわたって読まれ続けており、人間に受け入れられやすい物語と修辞の構造を考えるうえで適切な題材と考えられる。もう一つの候補は星新一のショートショート作品である。星新一作品は物語構造の多様性と面白さに定評があり、作品数が多く、かつ分析の比較的容易な短編作品である。これらの広く受け入れられている比較的短い単位の分析から始め、関連作品や類似の比較対象へと分析対象を徐々に広げて行き、汎用的な物語構造の特定を目指していく。

[5]既存の修辞構造データベースシステムの活用と拡張

代表者はすでにテキストにおける修辞構造のデータベース化を行っており、また修辞構造の表示・検索用の Java アプリケーションも開発している。本研究課題においてはこれらの資産を活用し、修辞構造に合わせて物語構造をデータベース中に格納し、種々の処理を行い、結果の確認を行えるように拡張を行う予定である。既存のテキスト分析用アプリケーションを活用することで、より確実にデータベースソフトウェアの開発を行うことが可能である。

#### 4. 研究成果

[1] 物語構造のデータ化とオチ分析アルゴリズムの開発

星新一のショートショート作品を対象として物語の構造を各場面における登場主体（人物でない場合も含む）の言動の連続であるプロットとして記述する手法を開発した。プロットにおける言動の記述にあたっては、物語の起承転結に深く係る重要な言動をリストアップして類似のものを階層的に分類した物語言動辞書を構築し、他の物語や同じ物語の別の場面における類似度の比較検討が可能な形で物語構造の記述を実現した。

これらのフォーマットに基づき星新一作品中で悪魔物語、薬物語、宇宙物語に相当する約 300 話を分析し、コンピュータ可読の形式としてデータ化を行った。これに合わせて各物語のオチを特定しかつその種類を分類しオチと物語の構造の関係性を分析した。結果として星新一の物語における主要なオチは逆転構造であり、逆転の主な種類は主客逆転・言動逆転・動機逆転の三種類に大別できることが明らかになった。

データ化した物語構造およびオチの種類に対して、計量的な分析を行った結果明示的に記述された逆転型のオチに関してはプロット中の主体と言動の記述から機械的に特定が可能であることが判明した。実際にプログラムとしてオチの判定アルゴリズムを実装した結果約 6 割の精度で物語構造のデータからオチを特定し抽出することに成功した。

[2] 物語構造の自動抽出プログラムの試作  
物語構造のデータ化におけるフォーマットを用いて、自然言語処理技術によるテキストからの物語構造の自動的な抽出プログラムの試作を行った。物語からプロットの軸となる主体の言動のリストを抽出するため、一般的な物語における主体を示す語彙のリストアップと構造化を行い、物語における主体語彙辞書を作成した。また、物語言動辞書を拡張して、言動を表す動詞的な語彙だけでなく慣用句的な表現と、会話における語用論的機能を含んだ文末表現も盛り込んだ物語慣用表現辞書を構築した。また主体の省略表現を補うために代名詞の補完アルゴリズムを作

成した。これらの辞書およびアルゴリズムに基づいて、物語テキストを場面で分割し、物語中で行われた主体の言動に相当する部分を自動的に抽出する物語構造自動抽出プログラムの試作を行った。

[3] 修辞構造のデータフォーマットの作成とデータ化、修辞構造の妥当性の統計的判定手法の開発

新旧約聖書（新約 27 巻，旧約 39 巻）における修辞構造（交差配列法、並行法、集中構造）を機械可読のデータとして記述するためのフォーマットを作成し、1507 種類の修辞構造に対してそれを構成する 7715 対のテキスト個所のペアとその対応関係をデータ化した。またデータ化した結果を Web 上で公開した。これらのデータの妥当性を検討するために、低出現頻度語彙が修辞構造上ペアとなるテキスト個所間でどの程度出現するかを指標としてランダムに作成した修辞構造の場合と統計的に比較する手法を開発し、分析した修辞構造の妥当性を検証した。検証の結果得られた修辞構造はテキストペア中に共通に含まれる低出現頻度語彙の観点からは統計学的に有意水準であることが確認された。

#### <引用文献>

[1]:ウラジーミル・プロップ（北岡誠司,福田美智代訳）:「昔話の形態学」,水声社,1987.  
[2]:グレマス, A. J.（田島宏;鳥居正文訳）:「構造意味論 方法の探求」,紀伊國屋書店,1988.

#### 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 5 件)

1. Hajime Murai, "Quantitative Evaluation Method for a Literary Structure Hypothesis within the Bible", Journal of the Japanese Association for Digital Humanities, Vol. 3, accepted, 2018 査読あり.
2. Hajime Murai, "Prototype Algorithm for Estimating Agents and Behaviors in Plot Structures", the International Journal of Computational Linguistics Research, Vol. 8, No. 3, pp. 132-143, 2017 査読あり.
3. Hajime Murai, "Automatic Extraction of Reversal-Type Punch Lines in Shin'ichi Hoshi's Flash Fictions", Journal of the Japanese Association for Digital Humanities, Vol. 2, pp.31-47, 2017 査読あり.
4. Hajime Murai, "Database System for Archiving the Literary Structure of the Bible", Journal of the Japanese Association for Digital Humanities Vol. 1, No. 1, p.44-57, 2015 査読あり.

5. Hajime Murai, "Plot Analysis for Describing Punch Line Functions in Shinichi Hoshi's Microfiction", 2014 Workshop on Computational Models of Narrative, OpenAccess Series in Informatics, Vol. 41, pp. 121-129, 2014 査読あり.

〔学会発表〕(計 12 件)

1. Hajime Murai, "Situational Effects on Functional Word Frequencies within Conversational Sentences in Japanese Novels", Proceedings of JADH Annual Conference 2017, pp. 40-42, 2017.
2. Hajime Murai, "Characteristics of Utterances in Japanese Fiction-writing", IJCAI 2017, the 2nd. International Workshop on Language Sense on Computer, pp. 6-10, 2017.
3. 村井源, "言動分類による物語機能辞書の汎用化に向けて", 人文科学とコンピュータシンポジウム論文集, 情報処理学会シンポジウムシリーズ, Vol. 2017, No. 2, pp. 225-230, 2017.
4. 村井源, "日本語小説の会話文タグ付コーパスの開発に向けて", The 31th JSAI, 1D2-OS-29a-2 (PDF), 2017.
5. Hajime Murai, "Attributes of Agent Dictionary for Speaker Identification in Story Texts", Proceedings of JADH Annual Conference 2016, pp. 38-40, 2016.
6. Hajime Murai, "Extracting Agents and Behaviors from Stories for Automatic Plot Structure Identification", IJCAI 2016, International Workshop on Language Sense on Computer, pp. 1-8, 2016.
7. 村井源, "主体語彙辞書を用いた物語テキスト中の主体推定システムに向けて", じんもんこん 2016 論文集, pp. 209-214, 2016.
8. 村井源, "物語の構造抽出のための慣用的表現の処理", The 30th JSAI, 1K5-OS-06b-3 (PDF), 2016.
9. Hajime Murai, "Towards Automatic Extraction of Punch-lines: Plot Analysis of Shinichi Hoshi's Flash Fictions", Proceedings of JADH Annual Conference 2015, pp. 17-18, 2015.
10. Hajime Murai, "Identifying Synonymous Word Groups in the Synoptic Gospels: A Quantitative Analytical Approach", Digital Humanities 2015.
11. 村井源, "物語の構造抽出のための主体語彙辞書の作成", じんもんこん 2015 論文集, pp. 111-116, 2015.
12. Hajime Murai, "Extracting Factors of Small Stories from the Synoptic Gospels", Proceedings of JADH Annual Conference 2014, pp. 7, 2014.

〔図書〕(計 1 件)

村井源編集, 徃住彰文監修, 量から質に迫

る: 人間の複雑な感性をいかに「計る」か, 新曜社, pp. 1-240, 2014.

〔その他〕

ホームページ等

[http://www.bible.literarystructure.info/bible/bible\\_e.html](http://www.bible.literarystructure.info/bible/bible_e.html)

## 6. 研究組織

### (1) 研究代表者

村井 源 (MURAI, Hajime)

はこだて未来大学・システム情報科学部・  
准教授

研究者番号: 70452018