

## デジタルアーカイブへの intra-textuality の導入に向けて

村井源

東京工業大学 大学院社会理工学研究科

人文的テキストでの意味理解を機械的にサポートするためには、単語や単語と評価語の組のみを対象とするのだけでは十分ではなく、フレーズや文・段落などの、より大きなテキストの単位での意味関係のデータ処理も必要となる。本論文では、このようなより大きな単位での意味処理の実現に向けた基盤構築のために、デジタルアーカイブへ intra-textuality (テキスト内での小箇所間の関係性)の導入を提案する。まずケーススタディとして、旧新約聖書で頻繁に用いられる古典的修辞構造を対象としてデータの電子化を行い、データベース上での構成を検討した。将来的には既存のデジタルアーカイブシステムと統合し、計量的な意味分析のためのツールとして公開予定である。

### Towards for introducing intra-textuality in a digital archive system

Hajime Murai

Graduate School of Decision Science and Technology  
Tokyo Institute of Technology

In order to support semantic analysis of humanities text mechanically, it is not enough to analyze words and evaluation words pair, but it is needed to process larger unit such as phrases, sentences and paragraphs. This paper proposes to introduce intra-textuality within digital archive system. The future goal of that is to develop basis for realizing semantic analysis of larger unit. At first, classical rhetorical structures, which are frequently used in Old and New Testament, are digitized as a case study. A format for database is also investigated. In the future, those will be united within existing database system and will be open as a tool for numerically semantic analysis.

#### 1. はじめに

古典文献に関する学問では、大量のテキスト資料等を必要とするため、旧来より文献の蓄積・保存・整理および検索などが大きな問題となってきた。これらの問題に対して近年急速に発達しつつある情報技術は有効であり、古典文献とその周辺分野でも種々の電子化プロジェクトが進行している。電子化プロジェクトによって構築されつつある各種のデジタルアーカイブシステムはテキストの検索や比較（コンコルダンス等）さまざまな情報処理技術を取り込む形で進歩してきている。

このような中で筆者は、テキストのアーカイブ化のみでなく、情報の理解（テキストの場合解釈）のサポートまでを視野に入れた電子解釈システム（Digital Hermeneutic System: 以下 DHS）を提案し[1]、より汎用的なテキスト処理プラットフォームを模索してきた。

機械的な電子解釈のサポートを実現する上では、各種の言語処理アルゴリズムの柔軟で総合的な適応が必要であるが、従来の多くの分析アルゴリズムにおいては語彙のレベル、あるいは単語とその評価の極性の問題を中心に、学術論文・特許文献・大規模 Web コーパスなどを対象とした研究が進められてきている。しかし、人文的テキストでの意味理解を機械的にサポートするためには、単語や単語と評価語の組のみを対象とするのだけでは十分ではなく、フレーズや文・段落などの、より大きなテキストの単位での意味関係のデータ処理も必要となる。

本論文では、このようなより大きな単位での意味処理の実現に向けての基盤構築のために、デジタルアーカイブへの intra-textuality の導入を提案する。

#### 2. テキスト間の意味処理とアーカイブ

##### 2.1. 古典文献と inter-textuality

現在古典文献と呼ばれるようなテキストは、幾多の古いテキストの中で、歴史を通じて大きな影響を及ぼしてきたテキストが読解や研究等の対象として残ってきたものである。そのため、ある古典文献の周辺には、それに影響を受けた複数のテキストが存在している。また、古典テキストは後世に解釈され直される事で新たな意味を付与されることが少なからずある。このため、当該古典テキストのみではなく、古典テキストを解釈したテキストも合せて解釈することで、テキストが歴史の中でどのように捉えられてきたか、総体としてどのような概念形成に寄与してきたのか、を解析することが可能となる。

このような複数のテキスト間の関係性から生じる意味は inter-textuality (間テキスト性)とも呼ばれる。間テキスト性はジュリア・クリステヴァ (Julia Kristeva) [2]等に提唱された概念で、あるテキストは関連する他のテキスト群と相互に影響を及ぼしあう関係にあることをいう。これらは主に引用（解釈）や編集（文書成立）という形であらわれてくるが一般にテキストは他テキストとの間に間テキスト性を持つと考えられている。

## 2.2. 引用解釈と inter-textuality

古典文献の中でも、特に宗教の正典テキスト等に代表される思想書は、たとえ話や比喩などの間接的で高度な解釈を必要とする複雑な表現に富んでいる。そのため、あるフレーズの意味を唯一の解釈に限定すること自体が困難である。唯一の解釈が特定困難な結果として、一つの文書に対して、多数のグループが作成した多数の解釈が並立する状況が築かれる。これらの並立する解釈がそれぞれテキスト化を行い古典テキストの周辺に複雑な inter-textuality の関係性が構築される。そのため、このような各グループの教義を記した文書を集積することで、グループごとの教義の相違と関係性も解析できると考えられる。

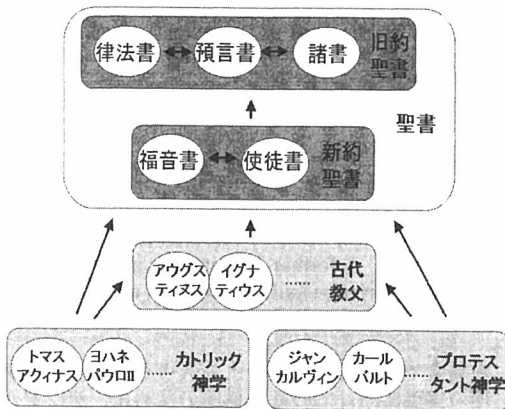


図1 聖書と教義文書の引用関係

たとえば、現代社会にも大きな影響を与えており、また多くのデジタルアーカイブプロジェクトの対象ともなってきた、キリスト教の聖書と関連文書群の関係性は、正典テキストである聖書の引用・解釈を中心として構成されている。このため、キリスト教教義を記した文書中では古代より多数の聖書引用が行われる。また、伝統的聖書解釈を重んじ正統的な教義であることを主張する意図から、教義文書間でも引用は頻繁に行われている(エラー! 参照元が見つかりません。)。キリスト教神学全体を俯瞰しようとするような著作では、これらの個々の引用と解釈を全体として体系化するために、ある箇所の特定の解釈を他箇所の解釈と結びつけることで、全体としての教義体系が構築されている。

図1のように、古典文献の解釈においては、時間的・空間的な隔たりのある関連周辺テキストとの間の関係性(引用・参照)により新たな意味(解釈)が派生しうる。

これらの密接なかかわりを持つ文書群内での inter-textuality は明示的な参照情報である引用箇所情報等を用いてその主要部分を示すことができると考えられる。

また聖書と教義文書のようにそれら全体が一つの体系的な概念・思想を指し示すものである場合には、文書群の関係性から抽出した解釈の体系が思想自体の構造を示唆すると考えられる。図2で示すようにこれらの関係性をネットワーク化することで、種々の計量的ネットワーク分析技法の適用が可能となる。

アーカイブへの inter-textuality 情報の蓄積と、計量的分析技法の適用により、文書群のテキスト参照情報による関係性を教義グループや時代ごとに抽出し、その相違点を数値的に比較することで教義グループごとの教義体系の相違の抽出も可能である[3]。

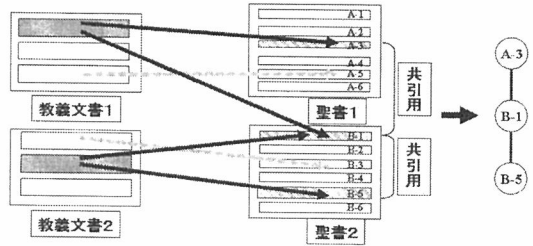


図2 聖書共引用からのネットワーク構成モデル

## 2.3. 異本テキストと inter-textuality

直接的な引用関係以外にも古典文献には多くの inter-textuality が存在する。その一つが編集過程における inter-textuality である。例えば、新約聖書における冒頭の三福音書は、類似点が多いことから共観福音書と呼ばれている(共に並べて観ることに由来)が、これらの成立過程には図2のような影響関係が想定されている。



図2 共観福音書成立過程(四資料説)

図2のように想定される仮定を経た結果として、三共観福音書間のテキスト中には、個々の小部分間での複雑な相違がみられる。聖書学ではテキスト内の小区分を、ギリシア語で切れ端(a cutting-out)を意味する pericope と呼ぶが、編集過程を経ることで各共観福音書の pericope には例えば図3に示すような関係性が見出される。図中の No.の後に続く番号は、“Synopsis of the four Gospels”[4]の共観福音

書の分析結果に基づく pericope 番号である。また：でつながれた数字は聖書で用いられる、聖書の箇所を示す略記号である。

このような異本間の関係性もまた一種の、inter-textuality のであると考えられる。異本間での編集結果の相違は、編集者の編集意図の相違を反映しており、図3のようなテキストの相違を pericope 間の関係性としてデータ化してアーカイブに含め、計量的に分析を行うことで、各テキストにおける編集上の中心的関心を抽出することも可能である[5]。

これらの知見に基づき筆者はデジタルアーカイブシステムに対して、周辺テキストとの間の関係性をデータとして、inter-textuality (間テキスト性) をアーカイブに導入することでより深い人間的な意味解釈のサポートが可能になることを提案してきた(図4) [6]。

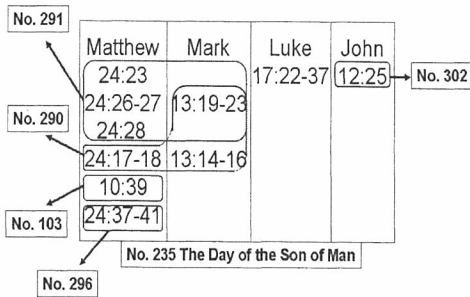


図3 pericope 間の関係の例

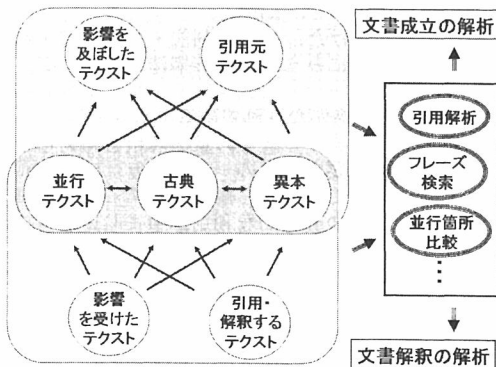


図4 inter-textuality に基づくアーカイブシステム

デジタルアーカイブの特徴として大量のテキストを集め同時に処理可能な点があるが、古典の研究では成立過程と文書解釈は大きな問題であり、inter-textuality の解析は不可欠である。このため関連するテキストを収集して大規模テキストコーパスを構築し、数値的に間テキスト性を解析できれば、単なるアーカイブではなく、テキストの相互関連の中での

意味分析を支援する統合的システムを構築しようと期待される。

### 3. アーカイブへの Intra-textuality の導入

#### 3.1. Inter-textuality と Intra-textuality

テキスト間の関係性から生じる意味は、計量可能なデータに基づく自動的意味分析に適しているが、テキストとテキストの関係性は単純ではなく、多様なあり方が想定しうる。一般的には inter-textuality とした場合複数の独立したテキスト間の関係性から生じる意味の問題を扱う場合が多い。しかし、特定テキスト内のある小部分から他の小部分への関係性によって間テキスト性が生じる場合も少なくない。このようなテキスト内での間テキスト性を inter-textuality と区別して intra-textuality と呼ぶことが提唱されている[7]。

Intra-textuality の具体的な形状としては、テキストの内的なデザイン、構造、分割などが意図的に整然と配置されている物などが想定されている。これらは現代のテキストにも当然見出される物であるが、修辞学の発達したギリシヤ・ローマ時代から、様々な修辞的技法の実践としてテキスト内の関係構造の精緻化が行われていた。テキスト構造の修辞技法で広く知られているところでは、聖書学の分野で近年精力的に研究が進められている交差配列法・並行法・集中構造などを挙げられよう[8]。

交差配列法 (chiasmus) とは、ABB'A' などと記号で表現されることが多いが、テキストのある部分が異なりながらも対応している小部分の関係性による入れ子になっている構造である。もっとも単純な形の交差配列法の例を図5に示す。図中での対応する箇所を線でつなぐと X (ギリシヤ文字のカイ) の字になることから、カイアスムス・キアスムス (chiasmus) と呼ばれる。

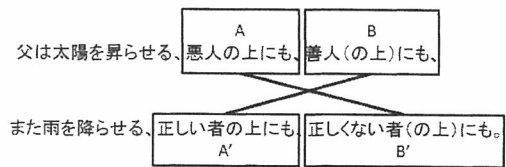


図5 交差配列法の例 (Mt5:45)

図5のように小さい構造としては単語やフレーズの単位で修辞的構造が作られるが、より大きな文の単位で作られる交差配列法も存在する。また、対応関係も AB 二種類のみではなく、より多層の入れ子構造の物が多数みられる。図6に多層で文単位での交差配列法の例を示す。

文よりもさらに大きな単位での構造として、先に述べた聖書学でのテキストの小区分である pericope 単位の交差配列法も多数存在する。pericope は長さ

が必ずしも一定していないが、数文から数段落程度の物語の切れ目で区切られた単位である。

次に集中構造(concentric structure)であるが、基本的な構造としては交差配列法と同じであるが中心に核となる部分が存在するような構造がこのように呼ばれる。中心に核となる部分が存在することで、テキスト中でどの部分が重要なかを明示的に示す技法であると考えられている。

マルコ福音書中にある pericope 単位での集中構造の例を図 7 に示す。図中の数字は pericope の範囲を示す聖書略記号である。この場合、中心に来るのは 9:30-32 にある二度目の受難の予告であり、当該箇所全体での中心的なテーマが受難の予告にあるということが解釈可能である[9]。

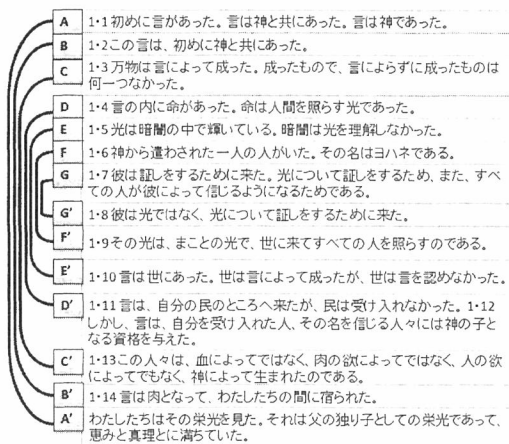


図 6 多層的交差配列法の例 (Jn1:1-14, 新共同訳)



図 7 pericope 単位の集中構造の例 (Mk8:22-10:52)

最後に並行法(parallelismus)であるが、名称の通りテキストのある二つの小部分が並行的な関係にある

修辞構造を指す言葉である。図 8 に文単位での並行法の例を示す。

これらの修辞構造は元来テキスト化以前の口頭での伝承の際の記憶の補助的役割を担っていたと考えられるが、修辞学の発達や、テキスト自身が種々の編集段階を経る中で、テキストの意味解釈の上でも重要な役割を担う者になってきたと考えられている。上記のような修辞構造の分析・特定により、

- ・ テキスト内の整合性の確認
  - ・ テキストの核となるメッセージの発見
  - ・ 対応によってメタファー等の解釈が明確化
  - ・ 語義やテキスト範囲の確定
- などが可能になると考えられる[10]。

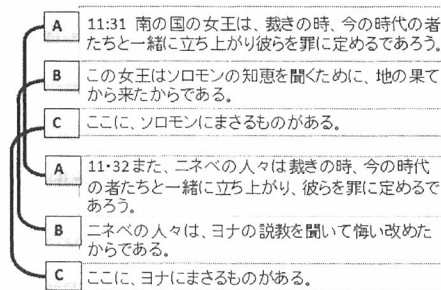


図 8 文単位の並行法の例 (Lk11:31-32)

また、修辞構造は特定のテキスト箇所に対し一つだけ適応されるのではなく、ある箇所がテキスト中で複数の別個の修辞構造の一部分として用いられることが多数発見されている[9]。

本論文で取り上げた、交差配列法・集中構造・並行法は聖書で顕著にみられる修辞構造ではあるが、他のギリシャ・ローマの古典文献や中世の文献等においても多くの研究がなされている。

修辞構造は、古典テキストの意味解釈を考える上で無視できない重要な要素の一つであり、デジタルアーカイブにおいてもこれら修辞構造データの蓄積および意味解釈への計量的な利用が望ましいと考えられる。

本論文では intra-textuality の中でも古典テキストでの構造的修辞技法のデジタルデータ化とその計量的な利用の実現を、聖書をケーススタディとして検討していく。

### 3.2. Intra-textuality のデータ表現

デジタルアーカイブに実装する場合、inter-textuality を表現するデータとしてはテキスト間の引用・参照関係が比較的实现が容易である。また引用・参照関係のアーカイブ化により、引用分析の手法を用いたテキスト解釈の概念構造を計量的に抽出することも可能であることが示されている[3,5]。

一方 intra-textuality をアーカイブ上に実装する場合には、これらがテキスト内の構造であるがゆえにテキスト内の区別の表記と合わせてデータ構造化

を行う必要がある。テキストの逐語的修辭構造の記述に関しては、Rhetorical Structure Theory に基づくテキスト記述が主に自動テキスト生成などの分野で用いられている[11]。しかし Rhetorical Structure Theory ではテキスト要素間の対応関係が基本的には隣接する一対一対応でしか記述できないという強い制限があり、多層的で広範囲にわたる古典的な修辭構造の記述には適さない。

そこで古典的な修辭構造の形式的な特徴と最終的なリレーショナルデータベースへの実装を鑑みて、以下のようなデータの記述形式を提案する。

まず考慮すべき点として、特定のテキスト箇所に対する修辭構造が複数独立に存在しうる点がある。このため、個々の修辭構造を別個に、かつ他の修辭構造の記述に依存せずに記述できる必要がある。

また、修辭構造の単位はフレーズ、文、pericope や場合によっては本単位での構造も考慮する必要がある[10]。このため各修辭構造の構成要素において柔軟なテキスト範囲の指定が可能である必要がある。

また、これらの修辭構造においては、指定された範囲内のどの部分（単語・フレーズ・文構造 etc.）が対応する範囲のどの部分と、どのような形で対応しているか（同じフレーズを含む、テーマが同じ、反意語で対応 etc.）などの情報も含め記述することが望ましい。

また、一つの修辭構造における対応する箇所の数は一定していないが、リレーショナルデータベースの特性上、要素の数が可変長となることは正規化の観点から望ましくない。そのため、修辭構造全体を一つのデータとして登録するのではなく、修辭構造の中に含まれる対応箇所の対を一つのデータとして登録することを考える。このため、データ中にどの修辭構造のどの位置の対応であるかを含め記述する。

以上を考慮して、修辭構造のアーカイブ化のためのリレーショナルデータベースの構造を以下のように提案する（表 1）。

表 1 修辭構造の基本データ形式

要素名	概要	形式
ID	修辭構造の識別用 ID	文字列
構造位置	構造上での対応の位置	文字列
特定者	構造を特定した解釈者	文字列
対応箇所 1	テキスト上の位置	文字列
対応箇所 2	テキスト上の位置	文字列
対応内容 1	箇所 1 の対応内容	文字列
対応内容 2	箇所 2 の対応内容	文字列
対応関係	対応の関係性、共通項	文字列

これらの修辭構造のデータ化においては、人文諸学共通の問題として解釈者により結果が異なることを考慮して誰による構造の提案かも併せてデータ化

を行ってゆく。データ入力フォームと合わせて Web 上で公開することにより多様な解釈を並行的に閲覧可能なプラットフォームになることを目指している。

#### 4. Intra-textuality の対象データ

聖書の具体的な修辭構造については多くの研究がなされており、聖書中の特定の箇所についての構造の提案も複数挙げられる場合が少なくない。構造の単位もフレーズレベルから本まで多種多様である。

本研究では、これらの様々な構造の提案を電子化していく予定であるが、手始めとして聖書の各書に体系的に pericope 単位で構築された多層集中構造[12]を対象として電子データ化を行い、すでに構築したアーカイブシステム上への導入を検討する。

現在電子化を完了したデータを表 2, 3 に示す。対象は新・旧約聖書であり、これらのテキスト中の交差配列法と集中構造を電子化している。最終的にはこれらのデータを DHS システムと統合し、Web 上での閲覧編集が可能なオープンシステムとして公開予定である。

表 2 電子化済み修辭構造データ（旧約）

書名	pericope 数	修辭構造数	対応データ数
創世記	81	48	357
出エジプト記	49	31	165
レビ記	25	39	96
民数記	49	31	165
申命記	49	31	165
ヨシュア記	31	31	119
士師記	31	31	119
サムエル記	85	60	416
列王記	85	60	416
イザヤ書	105	60	522
エレミヤ書	81	48	357
エゼキエル書	49	31	165
ホセア書	21	16	53
ヨエル書	9	6	11
アモス書	21	16	53
オバデヤ書	7	1	3
ヨナ書	5	3	4
ミカ書	13	16	32
ナホム書	5	3	4
ハバクク書	9	6	11
ゼファニヤ書	7	6	10
ゼカリヤ書	25	39	96
ハガイ書	5	3	4
マラキ書	9	6	11
ヨブ記	21	16	53
雅歌	17	15	32
ルツ記	9	6	11
コヘレトの言葉	25	39	96
エステル記	15	16	35
ダニエル書	17	15	32
エズラ記・ネヘミヤ記	25	39	96
合計	985	768	3709

表 3 電子化済み修辞構造データ (新約)

	pericope 数	修辞構 造数	対応デー タ数
マタイによる福音	144	102	860
マルコによる福音	85	60	416
ルカによる福音	144	102	860
ヨハネによる福音	81	48	357
使徒言行録	81	48	357
ローマの信徒への手紙	31	31	119
コリントの信徒への手紙一	25	39	96
コリントの信徒への手紙二	17	15	32
ガラテヤの信徒への手紙	13	12	24
エフェソの信徒への手紙	13	12	24
フィリピの信徒への手紙	9	6	11
コロサイの信徒への手紙	9	6	11
テサロニケの信徒への手紙一	9	6	11
テサロニケの信徒への手紙二	5	3	4
テモテへの手紙一	13	12	24
テモテへの手紙二	9	6	11
テトスへの手紙	5	3	4
フィレモンへの手紙	9	6	11
ヘブライ人への手紙	17	15	32
ヤコブの手紙	9	6	11
ペトロの手紙一	9	6	11
ペトロの手紙二	9	6	11
ヨハネの手紙一	13	12	24
ヨハネの手紙二	5	1	2
ヨハネの手紙三	5	1	2
ユダの手紙	7	1	3
ヨハネの黙示録	81	48	357
合計	857	613	3685

## 5. 今後の課題

現在データ入力を進め、DHS 上での実装に向けて準備を進めている。DHS における実装では単に修辞構造データが閲覧できるのみではなく、入力された修辞構造データに基づく意味解釈支援アルゴリズムの実装を検討している。具体的には、まず修辞構造自体の妥当性を検証するための、テキスト中の複数の特定箇所間の語彙・フレーズレベルでの比較用のアルゴリズムの実装が考えられる。修辞構造で対応するテキスト箇所間の多くは、その箇所に特徴的な語彙やフレーズが対応する形であり、TF・IDF 等の計量的な指標を用いて対応構造の妥当性が数値的に示しうると期待される。修辞構造の妥当性を評価可能になることで、従来は困難であった、複数の研究者によって提示される修辞構造間の、より客観的な比較や検討が可能になると期待される。

また、これらの対応する語彙・フレーズの関係性から、修辞構造全体としてのテーマ、中心的主張を抽出し、全テキスト中の意味関係の中での修辞構造の果たす役割を、ネットワークなどを用いて視覚的に表現することも考えられる。このためには、対応関係抽出によって、構造の核となる文・フレーズ・単語を抽出し、それらの全体構造での位置や、他の構造との関係性などを計量的に評価する必要がある。修辞構造のアーカイブ化の暁には、古典的な修辞構造のみならず汎用的なテキスト内構造である intra-textuality の記述と分析をも可能なソフトウェアとしてオープンソースでの公開を予定している。

## 謝辞

本研究は科研費「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」(22700256) および「知識共有のための価値指向型オントロジーの多分野多言語化」(20300074) の助成を受けた。

## 参考文献

- [1] 村井源, 徃住彰文, “古典文献用 Digital Hermeneutics System の提案と実装”, IPSJ SIG Computers and the Humanities Symposium 2007, Vol. 2007, No. 15, pp. 101-106, 2007.
- [2] Le Texte du roman, Julia Kristeva, 1970, 『テキストとしての小説』谷口勇訳, 国文社, 1985.
- [3] 村井源, 徃住彰文, “Co-citation Network による宗教思想文書の解析”, 人工知能学会論文誌, Vol.21, No.6, pp. 473-481, 2006.
- [4] Aland, Kurt, “Synopsis of the Four Gospels”, revised standard version, New York, the United Bible Societies, 1982.
- [5] 村井源, 徃住彰文, “正典テキスト群から編集的中心メッセージを抽出するネットワーク解析法”, 情報知識学会誌, Vol. 17, No. 3, pp.149-163, 2007.
- [6] 村井源, 三宅真紀, 赤間啓之, 中川正宣, 徃住彰文, “テキスト間参照情報を考慮した古典文献デジタルアーカイブ”, IPSJ SIG Computers and the Humanities Symposium 2005, Vol. 2005, No. 21, pp. 21-25, 2005.
- [7] Alison Sharrock, Helen Morales, “Intratextuality: Greek and Roman Textual Relations”, Oxford Univ Pr on Demand, 2001.
- [8] 森彬, “新・聖書の集中構造”, ヨルダン社, 2001.
- [9] 村井源, “マルコ福音書の多層集中構造”, 日本カトリック神学会誌, Vol. 20, pp. 65-95, 2009.
- [10] 森彬, “ルカ福音書の集中構造”, キリスト新聞社, 2007.
- [11] Maite Taboada, William C. Mann, “Applications of Rhetorical Structure Theory”, Discourse Studies, vol. 8, no. 4, pp. 567-588, 2006.
- [12] 村井源, “聖書における多層集中構造の妥当性の検討”, 日本カトリック神学会 第 21 回 学術大会, 2009.