

## 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

村井 源 三宅 真紀 赤間啓之 中川正宣 住住彰文  
東京工業大学 社会理工学研究科

歴史を通じ多大な影響を人類に与えてきた古典的テキストも、近年の情報処理技術の発達に伴いデジタルアーカイブ化が進んでいる。しかしその多くは蓄積・検索・表示の効率化のみにとどまっており、デジタル化の恩恵を十分に活かしているとは言い難い。本稿では古典テキストのテキスト間参照情報も含めて蓄積を行い、情報処理技術を活かした解析機能も含めて一つのデジタルアーカイブシステムとする手法を提案し、聖書とその関連テキストについて実際にシステムの構築を行った。

## Digital Archives of Classic Texts Using Text Referential Information

Hajime Murai Maki Miyake Akama Hiroyuki Masanori Nakagawa Akifumi Tokosumi

Graduate School of Decision Science and Technology  
Tokyo Institute of Technology

With the recent advances in information processing technology, there has also been progress in the creation of digital archives handling classic texts that have had a major impact on the human race throughout history. However, most digital archives focus solely on improving efficiency in terms of data storage, retrieval and display, and fail to fully realize the potential of digitalization. This paper proposes a digital archive system that integrates the storage of classic texts together with reference text data and includes analytical functions drawing on information processing technology. The paper also introduces the construction of a real system for the Bible and related texts.

### 1. まえがき

情報技術の進歩に伴い、古典文献の分野においてもさまざまなデジタル化プロジェクトが進行している。そしてその多くは、コンピュータ化によって旧来の利用法である、蓄積・検索・表示の効率化をはかっている。その一方で古典文献自体の意味解析は、情報技術の成果を利用しつつも、人手によって行われている。

本研究は旧来のデジタルアーカイブシステムの一般的な機能に合わせて、アーカイブに含められたテキスト間の関係性から発出する意味・概念の解析機能も提供する統合的システムを提案する。具体的にはあるテキストの元となるテキスト群を解明する過去にさかのぼる方向と、その解釈を行うテキストにより未来へと広がる方向の両面からの意味解析を統合的に行う。単に関連するテキストを集めるのではなく、その相互の関係性を数値的に解析することでテキストの統合的理解を深めることが期待される。

デジタルアーカイブ化する対象としては聖書を選択し、その関連テキストと引用データを用いた。

### 2. 間テキスト性とデジタルアーカイブ

#### 2.1 間テキスト性

間テキスト性(intertextuality)とはジュリア・クリステヴァ(Julia Kristeva)[1]等により提唱された概念で、あるテキストは関連する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。

この間テキスト性が顕著に表れる場合としては、

過去に書かれたテキストの節やフレーズの一部分の意味を汲み取り、明示的あるいは非明示的な形で行われる引用や(文書解釈)、さらには部分のみでなく他テキスト自体を何らかの意図・解釈をもって編集する(文書成立)ことなどが考えられる。過去のテキストからの影響を全く受けずにあるテキストが成立するとは考えられないため、一般にテキストは間テキスト性を持つとされている。

#### 2.2 古典テキストと間テキスト性

歴史を通じ多大な影響を人類に与えてきた古典的テキストにおいては、これらの間テキスト性を解析することは重要な問題となる。

時間軸を過去にさかのぼる方向としてはある古典テキストがどのようなテキスト群の影響を特に強く受けて構築されたのか、それはどのような作成過程を経ているのか、逆に未来に向けた方向としては、このテキストがどのようなテキストに影響を与えたのか、またどのように理解されてどのような文脈で引用されるか、これら間テキスト性に関わる種々の関係性を解き明かすことで、テキストの理解をより深めることが可能となると考えられる。

#### 2.3 デジタルアーカイブとの親和性

情報技術の進歩に伴い近年発達してきたデジタルアーカイブの特徴として大量のテキストを集積し、高速に様々な情報処理を施すことが可能な点がある。このため古典での間テキスト性に基づく成立過程と文書解釈の解析を電子化可能なアルゴリズムで表記できれば、人手のみでは困難な大規模で正確な解析

が可能となり、新たな関係性の発見や従来の仮説のデータに基づく、より客観的論証を行うことができると考えられる。

これによりデジタルアーカイブは単にテキストを集めただけではなく、テキストの相互関連の中での重層的な意味を提示する統合的システムになりうると期待される。

#### 2.4 旧来のデジタルアーカイブ

旧来の多くの古典的テキストを収録したデジタルアーカイブプロジェクトにおいては、デジタル化の主目的が蓄積・検索・表示の効率化であって、それに対し間テキスト性はあまり重視されていない。語の出現頻度や出現位置を計算し表示するデジタルコンコルダンス機能や、本文テキストの単語と辞書を連動させる機能がわずかに見られる程度である。

多くのプロジェクトが存在する聖書関連でも、上にあげた機能以外ではデジタル化する以前の引用表をそのまま掲載する程度のものである[2][3]。また聖書解学者・聖書研究者が用いるために開発された聖書ソフトウェアにおいても、旧来の多数種類の翻訳版データの蓄積とそれらの並行表示、原典が書かれた言語であるギリシア・ヘブライ語の辞書との連動や単語検索の容易化が中心である[4][5]。

#### 2.5 提案するデジタルアーカイブシステム

本稿において提案するのは、ある古典テキストを中心として、それと関係性の強い一群のテキストのデジタルデータに合わせ、それらの相互関連情報と解析機能が一体化したシステム構成である(図1)。

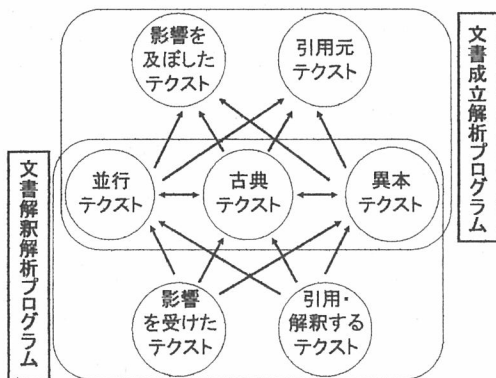


図1: 提案するデジタルアーカイブモデル

ある古典テキストの成立に影響を与えるのはそれが引用する文書などであるが、それらが直接的には入手不可能な場合も少なくない。そのため成立過程を解析する場合には、同様に成立した並行テキスト群や、同内容でありながらも多少の差異を含む異本テキスト群の比較検討が有効な手段となる。

また古典テキストがどのように解釈されているかを解析する場合にも、解釈を記したテキストのみを対象とするのではなく並行テキストや異本テキストも含めた全体としての関連テキスト群に対する解析を行うことで、ある古典テキストが記された当時の文脈に対する相対的な解釈のありようを捉えることが可能となる。

### 3. テキスト成立における間テキスト性

まずテキストの成立における間テキスト性(編集過程)を考える。テキストを記した著者はゼロから新概念を構築するわけではなく、すでにある知識や物語、思想等を背景にして創作を行うと考えられる。これらのテキスト成立の背景となる情報は必ずしも文書化されているとは限らないが、文書化がされている場合には注目しているテキストと背景テキストの間には間テキスト性が成立する。

ただし背景テキストがあったはずだが残存しない場合や、あるいは複数テキストが並立して存在し、これらの内どれが背景テキストであるかという依存関係が明確ではない状況も古典テキストにおいては多数考えられる。

このような状況においては、注目するテキストに対して同様の状況で成立した並行テキスト、同内容を異なる形で伝える異本テキスト等との間テキスト性を数値的・定量的に解析することで背景テキストをより客観的に特定できると考えられる。

本研究では聖書解釈学での編集史批判(Redaction criticism)を用い、テキストの過去にさかのぼる間テキスト性を解析する手法を実装する。

#### 3.1 編集史的研究: 共観福音書問題

新約聖書中でイエスの直接的言動を伝える4つの福音書があるが、その中でマルコ(以降式中ではMKと表記)、マタイ(同MTと表記)、ルカ(同LKと表記)の3文書は、互いに密接な類縁関係があり、並行する対観表の形にあらわすことができるため「共観福音書」(Synopsis)と呼ばれている。そして、これらの類似性から、相互にどのような文献的な依存関係があるかという、いわゆる「共観福音書問題」が長年議論されてきた。この問題に対し編集史的研究は、3文書の相互関係のみならず、他の福音書間との相互関係を明確にした。また、これらの間テキスト性から立てられた「二資料説」が、現在の聖書学においては、ほぼ定説化している[6]。

#### 3.2 オリジナル文書抽出モデル

われわれは、この「共観福音書問題」に対して、仮説検証(オリジナル文書抽出)計量モデルを考案し(図2)、文書の特徴を質的ばかりでなく量的に表現することを試みた。

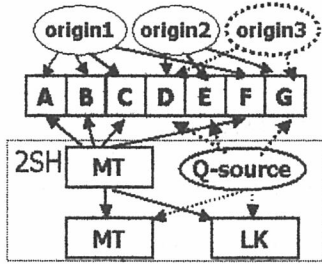


図 2：ニ資料説&オリジナル文書抽出モデル

計量モデルには、以下のような7つのカテゴリで共観福音書(MT,MK,LK)を表し、共観福音書に出現する単語の頻度数を、7つの文書カテゴリに分配した。

$$\text{Union}(MT,MK,LK) = \text{Union}(A,B,C,D,E,F,G)$$

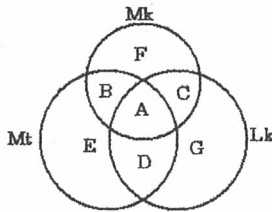


図 3：7つのカテゴリ

図 3 に、7つの分配カテゴリをベン図で表す。ここで、3書共通部分(A)、マタイ・マルコ共通部分(B)、マルコ・ルカ共通部分(C)、マタイ・ルカ共通部分(D)、と、それらの共通部分を除いたマタイ(E)、マルコ(F)、ルカ(G)部分である。

このカテゴリを用いて、「ニ資料説」のオリジナル文書(Mk, Q)を表現することが可能である。

$$\begin{aligned} Mk &= \text{Union}(A,B,C,D) \\ Q &= \text{Union}(D,E,G) \end{aligned}$$

このように、文書の成立上で用いたと考えられている資料が、分類した一つのカテゴリ、あるいは複数のカテゴリの和によって説明される。したがって、カテゴリの間の相関関係を計量的に明らかにし、その相関に基づいてカテゴリ間のクラスタリングを示すことによって、仮説を検証することが可能である。

そこで、それぞれのカテゴリに所属する単語に着目し、それらの出現頻度をもとに計算されたカテゴリ間の相関係数が、そのカテゴリ間の計量的相関関係を表していると仮定し、その相関係数を用いて因子分析を行った。

その結果、聖書学で立てられた定説とは異なる成立過程を示唆する結果に至った[7]。

他の多くの古典文献においても、唯一絶対の原典が特定されておらず、複数の類似文書間での編集史が問題となるケースが考えられる。本研究ではこの計量モデルに基づいた文書解析アルゴリズムを文書成立過程解析ツールとしてアーカイブに導入する。

#### 4. テキスト解釈における間テキスト性

次にある古典テキストから後世へと広がる解釈による間テキスト性を考える。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り、当該古典テキストのみではなく、それらの解釈をも合わせてアーカイブ化することで、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になると考えられる。

特に宗教の正典等に代表される思想書は、たとえ話や比喻などの間接的で高度な解釈を必要とする表現に富んでおり、そもそも唯一の解釈を限定することが困難である。このため一つの文書に対して多数の解釈が並立しており、それぞれのグループの教義を記した文書を集積することでこれらの相違と関係性もまた解析できるものとする。

本研究では聖書と、聖書の解釈によって成立するキリスト教教義テキストの聖書引用をデータとしてアーカイブ化し、引用・共引用データの解析により各箇所解釈の相違を示す手法を実装する。

##### 4.1 教義文書の間テキスト性

キリスト教教義は、正典とする聖書(旧約聖書・新約聖書)を引用し解釈することで主に構成されているため、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これら教義文書間でも引用は頻繁に行われる(図4)。

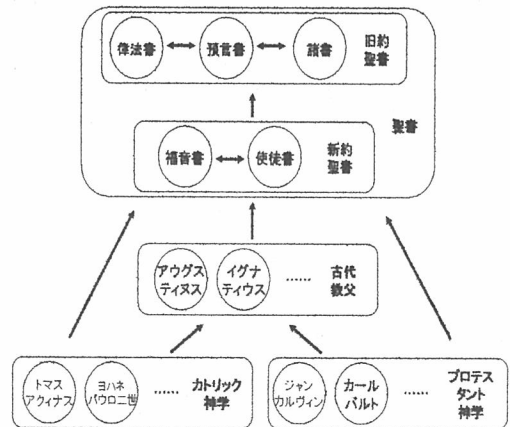


図 4：聖書と教義文書の引用関係

また、これら個々の引用と解釈が全体として体系化されるため、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[8]。

これら密接なかかわりを持つ文書群の間テキスト性は明示的な参照情報である引用箇所情報を用いてその主な部分を示すことができると考える。

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。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、その相違点を比較することで教義グループごとの教義体系の相違を抽出することが考えられる。

#### 4.2 共引用パターンによる教義解析

聖書と教義文書の間を関係付ける密接な引用関係に対して、引用・共引用の解析から解釈の相違を抽出する手法を考案した(図5)。これは聖書の各節をノードとし、そのうちで共引用関係にある節間をリンクで結ぶことで、それぞれの解釈者にとっての聖書各節の関係をネットワーク表現する手法である。

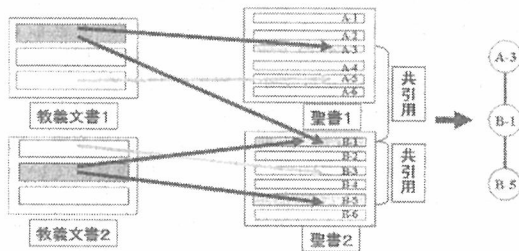


図5：聖書共引用からのネットワーク構成モデル

このアルゴリズムを用いカトリック教会の各時代を代表する神学者を三名選びそれぞれの共引用ネットワークを構成し、それらの相違点の検証により、カトリックという思想集団での聖書解釈の共通性と時代による変遷を示すことが可能である(図6)[9]。

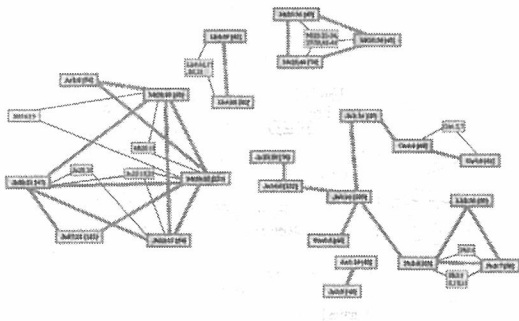


図6：聖書共引用関係から抽出したネットワーク図

これらの引用データにプロテスタント神学者らの引用データも合わせ解析アルゴリズムと共に本論文のアーカイブシステムへの実装を行う。

## 5. システム構成

古典テキストとその関連文書におけるテキスト間参照情報を解析する機能を持つデジタルアーカイブシステムを、古典テキストデータとテキスト間参照情報およびそれぞれの解析アルゴリズムを統合したインターフェースとして構築する。テキストは聖書関連に限らず汎用的に利用可能なXMLスキーマに沿って記述する。これらのテキストとデータを共通として各々の解析プログラムが動作する形式を取る。

### 5.1 収録データ

聖書データとしてはギリシャ語版新約聖書(Nestle-Aland 27版)を中心として、ドイツ語版ルター訳聖書、日本語版では新共同約聖書に合わせ、フランス語版訳聖書(新約部分のみ)、ラテン語ではウルガタ訳聖書(Nova Vulgate Bibliorum Sacrorum Editio)のほかにギリシャ語版旧約聖書(セプチュアギンタ)の収録を現在検討中である。

周辺文書としてはカテキズム(Catechism of the Catholic Church)、教会法(The Code of Canon Law)、第二バチカン公会議文書、近代の(レオ13世以降)教皇文書等の主にカトリック文書をデータ化している。

また聖書引用データとしてはアウグスティヌス(354~430)、トマス・アキナス(1225~1274)、ジャン・カルヴァン(1509~1564)、カール・バルト(1886~1968)、ヨハネ・パウロ二世(1920~2005)の著作からとった引用箇所データを電子データ化している(表1)。

|           | 全引用数   | 全共引用数   |
|-----------|--------|---------|
| アウグスティヌス  | 22674  | 215824  |
| トマス・アキナス  | 36015  | 800457  |
| ヨハネ・パウロ二世 | 32166  | 643708  |
| ジャン・カルヴァン | 70324  | 2005864 |
| カール・バルト   | 53256  | 2653522 |
| 合計        | 214435 | 6319375 |

表1：聖書引用・共引用データ

### 5.2 共観福音書検索システム

共観福音書のギリシャ語テキスト[10]の共観表示&検索機能、および単語の出現頻度分布の計算機能を持った「異本解析支援システム：Tele-Synopsis」を開発した。また、テキストの主要単語の近傍で共起する単語を取得する、「共起情報取得システム：Tele-COEX」を開発した。

### 5.3 聖書引用解析システム

キリスト教教義文書からの聖書引用箇所をデータとして、聖書本文に合わせて聖書引用箇所の逆引き

表示と引用元の教義文書の表示を行う機能を持つ「聖書引用表示システム」を開発した(図7)。

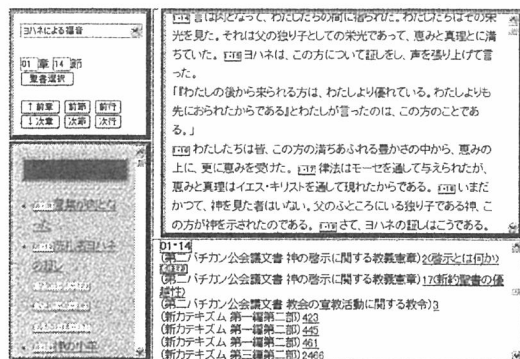


図7：聖書引用逆引き表示機能実行画面

また、聖書引用データから共引用データを構築し、引用データと合わせてネットワーク図化するためのデータを出力する「聖書引用共引用解析システム」を開発した。

## 6. 今後の方向性

### 6.1 アーカイブシステムの拡張

Web上の利用を視野に入れJSPとサーブレットを用いたJAVAアプリケーションとしてシステム全体を再構築する。データ形式およびインターフェースもより多様な古典テキストに対応するため汎用性を高める。

また、解析アルゴリズムは現在あるものに合わせて利用者がプラグインを後から任意に追加可能な形で提供する。最終的には仕様を公開しプログラム本体はGPLとして提供する予定である。

### 6.2 ネットワーク解析と表示機能

現在、聖書箇所共引用をネットワーク図化するためのデータ出力まで行っているが、これに合わせ聖書における単語共起情報もネットワーク図化するためのアルゴリズムを実装し、解析とデータの作成のみではなくネットワーク図としての表示まで一貫して行うためTouchGraph[11]、Graphviz[12]等の導入を検討している(図8)。

### 6.3 オントロジーの作成

本研究の一環として研究中の、独自のネットワーククラスタリング手法を用い、テキストの意味構造に踏み込んだ意味ネットワークから聖書オントロジーの構築を行う予定である。また聖書原典以外の時代に書かれたテキストや翻訳などの多様なテキストを使用することで、通時的・文化的な背景を明示するようなオントロジーの構築を目指す。

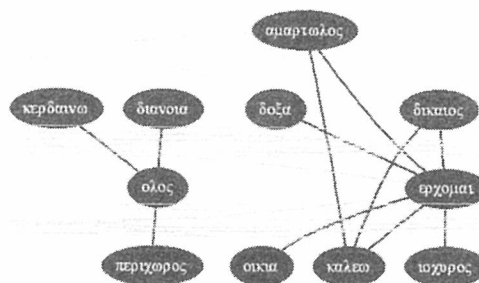


図8：開発中の共起単語情報ネットワーク化機能

## 7. 謝辞

本研究は、21世紀COEプログラム「大規模知識資源の体系化と活用基盤構築」の言語・文献資源分野に関する研究の一環として行われたものである。

## 8. 参考文献

- [1] Julia Kristeva, *Le Texte du roman*, 1970, 『テクストとしての小説』谷口勇訳, 国文社, 1985.
- [2] New Advent, <http://www.newadvent.org/>.
- [3] CCEL, <http://www.ccel.org/>.
- [4] Bible Works, <http://www.bibleworks.com/>.
- [5] Accordance, <http://www.accordancebible.com/>.
- [6] Conzelmann, H. & Lindemann, A., *Interpreting The New Testament*, trans. by Siegfried S. Schatzmann, Hendrickson Publishers, pp. 45-53, 1988.
- [7] 三宅真紀, 赤間啓之, 中川正宣: 聖書ソフトウェアの開発と因子分析に基づく福音書の特徴分析—因子得点を中心にして, 文理シナジー学会誌, Vol. 8, No. 3, pp. 199-207, 2004.
- [8] John Paul II, *Catechism of the Catholic Church*, 1992.
- [9] Hajime Murai & Akifumi Tokosumi, A Network Representation of Hermeneutics Based on Co-citation Analysis, *WSEAS Transactions on Information Science and Applications*, Vol. 11, No. 6, pp. 1513-1517, 2004.
- [10] Nestle-Aland, *Novum Testamentum Graece 26th edition*, German Bible Society Stuttgart, 1979.
- [11] TouchGraph, <http://www.touchgraph.com/>
- [12] Graphviz, <http://www.graphviz.org/>