Exegetical Science for the Interpretation of the Bible: Algorithms and Software for Quantitative Analysis of Christian Documents

Hajime Murai

Abstract. Systematic thought (such as Christian theology) has primarily been investigated using literature-based approaches, with texts that are usually more abstract and subjective in nature than scientific papers. However, as systematic ideas and thought influence all areas of human activity and thinking, the application of scientific methodologies such as bibliometrics, natural language processing, and other information technologies may provide a more objective understanding of systematic thought. This paper introduces four methods of quantitative analysis for the interpretation of the Bible in a scientific manner. The methods are citation analysis for interpreters' texts, vocabulary analysis for translations, variant text analysis for canonical texts, and an evaluation method for rhetorical structure. Furthermore, these algorithms are implemented for Java-based software.

Keywords: Bible, theology, interpretation, bibliometrics, NLP.

1 Introduction

As an aspect of higher cognitive functions, systematic thought has primarily been investigated using literature-based approaches, with texts that are usually more abstract and subjective in nature than scientific papers. However, as systematic ideas and thought influence all areas of human activity and thinking, the application of scientific methodologies such as bibliometrics, natural language processing, and other information technologies may provide a more objective understanding of systematic thought. By utilizing these new scientific methods,

Hajime Murai
Graduate School of Decision Science,
Tokyo Institute of Technology, Tokyo, Japan
e-mail: h_murai@valdes.titech.ac.jp

R. Lee (Ed.): *SNPD*, SCI 492, pp. 67–86.
DOI: 10.1007/978-3-319-00738-0_6

we can (a) ensure the objectivity and replication of results; (b) handle large-scale data precisely in a uniform manner [1].

I believe that it is possible to analyze the abstract thoughts and value systems embodied within a text corpus with such methods. In this paper, I focus on a Christian text corpus. Throughout history, traditional religions have exerted great influence on humanity. Most religions have certain canonical texts at their core, with the hermeneutics, or interpretations, of the canon also usually in text format. Thus, it is possible to represent key conceptualizations through the objective analysis of the canonical texts.

2 Approaches to Scientific Interpretation

As mentioned above, for the scientific analysis of thoughts, it is necessary that interpretations of the canon of target thoughts be analyzed scientifically. Unfortunately, it is currently impossible to achieve a scientific interpretation comparable to human interpretation, but it is possible to partially reproduce several human techniques of interpretation by utilizing scientific methods.

There are two quantitative approaches to interpreting the canon. The first is not a semantic interpretation of the canon itself, but an indirect approach to more clearly extract the details of interpretation. Although the second approach is direct, there are concerns that the resulting analysis is shallow because the canon itself does not always include as much information as is required in order to analyze the interpretation.

In the first, indirect approach, the relationship between the target text (the canon) and the texts that describe the interpretation of the target text (theologians' texts) is important. These relationships are called intertextuality. Citation analysis is an effective method for analyzing the relationships between texts. It clarifies the interpretation of some parts of the canon and the relationships between several parts of the texts. Therefore, this analysis is able to visualize the structure of interpreters' concepts. Citation analysis enables the scientific analysis of theological differences between theologians and between eras or sects [2]. Of course, it is also possible to extract the characteristics of theological interpretation by the quantitative analysis of distinctive and frequent vocabulary in the texts that describe the interpretation of the canon. It is also possible to extract the interpretation of the translator by comparing the correspondence between the original text and its translation into another language, because a translation is an interpretation of the original text [3].

In the second, direct approach, the extraction of characteristic words is fundamentally based on their frequency. Utilizing techniques such as TF-IDF, characteristic words can be extracted quantitatively. For characteristic vocabularies, it is possible to use co-occurrence [4] and dependency analyses to numerically clarify the usage tendencies of important words. Co-occurrence analysis is the study of word occurrences in common with the target word, and dependency analysis investigates the words in dependent relationships. The semantic analysis of words is fundamental to interpretation. These quantitative methods are equivalent to those used in the humanities, which are collectively termed concordance interpretation.

Another effective direct approach is to make comparisons between entire canonical texts or between small parts of canonical texts. This allows an analysis of the theological emphasis made when the canon was written [5] as well as a study of the process of creating the canon [6] on the basis of the quantitative comparison of several variant texts that are included in the canon. In addition, a comparison of parts of the text in the canon makes it possible to numerically verify the rhetorical structure that is constructed from the relationships among text parts [7].

In this paper, I would like to introduce four methods of quantitative analysis for the interpretation of religious thought in a scientific manner. The citation analysis of interpreters' texts and vocabulary analysis of translations can be considered as indirect approaches. For a direct approach, methods are introduced to extract canonical theology from differences in variant texts and to evaluate rhetorical structure. Moreover, I will discuss a software application for utilizing these algorithmic methods.

3 Co-citation Analysis of Religious Texts

3.1 Background

There are many theological differences between specific religious groups. As a result of these differences, interpretations of the canon can differ. If it is possible to scientifically extract these differences, the transition or mutual influence can be numerically analyzed.

Specifically, this method aims to automatically extract the main elements of a number of key conceptualizations from a religious text corpus and analyze their cluster construction using an objective and replicable methodology. This, in turn, will provide an objective basis for the examination of systematic thought [2].

3.2 Constructing Networks and Extracting Clusters

Here, we focus on the writings of St. Augustine and St. Thomas Aquinas, two influential Church Fathers, as well as those of Jean Calvin, Karl Barth, and Pope John Paul II. This enables us to extract essential teachings of Christian dogma through historical transitions and identify the individual characteristics of hermeneutics. Based on the patterns of Bible citations within their writings, we created networks for frequently cited sections of the Bible and extracted the main elements and clusters of these in order to compare a number of key conceptualizations. Clusters were extracted according to a threshold value of co-citation frequency. Table 1 gives the total number of citations and co-citations in each author's writings.

The resulting clustered network for Augustine is presented in Fig. 1. The nodes' alphabets and numbers are symbols that correspond to the Bible sections; dense parts are clusters. The differences in clusters extracted for each author are presented in Table 2.

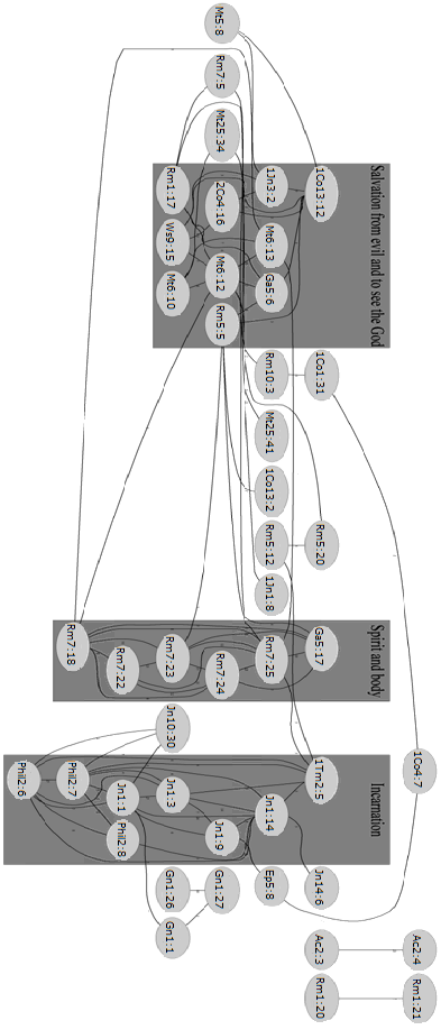

Fig. 1 Example of a Clustered Co-citation Network (Augustine)

Table 1 Citations and Co-citations

Author	Titles	Citations	Co-citations	Average citations per verse
Augustine	43	22674	215824	6.94
Thomas Aquinas	32	36015	800457	15.05
Jean Calvin	47	70324	2005864	13.51
Karl Barth	113	53288	2661090	23.67
John Paul II	1939	32166	643708	9.34

Table 2 Extracted Clusters

Author / Cluster	Augustine	Thomas Aquinas	Jean Calvin	Karl Barth	John Paul II
Incarnation	○	○		○	○
Salvation from evil	○		○		
Spirit and body	○			○	
Predestination			○		
Commandments			○	○	
Evangelization				○	○
Sola Fidei				○	
Suffering servant				○	
Creation					○
Judgment					○

3.3 Discussions

This analysis identified the core element of Christian thought to be incarnation, because almost all the famous theologians shared the same cluster about incarnation (which includes Jn1:14, Phil2:6, Phil2:7, Phil 2:8, Gal4:4). In addition, distinctions between individual theologians in terms of their sect (Protestant theologians share a cluster about the Commandments) and era (modern-age theologians share a cluster about evangelization) were identifiable.

As Christianity literally believes that Jesus Christ is the Messiah, the result indicating the core element of Christianity seems to be valid. Moreover, as Protestants resist the rules of the Catholic Church, it is reasonable that they might emphasize the Commandments of the Bible instead of those of the Catholic Church. Likewise, in the modern age of science and globalization, modern churches need to strengthen the concept of evangelization.

The co-citation analysis results seem to match the circumstances of each theologian. This method could be applicable to other theological corpora.

4 Extracting the Interpretive Characteristics of Translations

4.1 Background

Although there have been some studies that focus on background interpretations by comparing and analyzing translations, these have utilized the methodologies of the humanities, which are unsuitable for maintaining objectivity and for large-scale analysis. Utilizing information technologies, this paper proposes some methods for numerical comparisons and the extraction of background interpretations in translations.

Specifically, the first step is to estimate the correspondence between the original vocabulary and the translation of that vocabulary on the basis of quantitative data. The next step is to objectively and quantitatively extract the differences in translators' interpretations from the differences in corresponding vocabulary in each translation [3].

4.2 Extracting Corresponding Vocabularies

Various high-performance algorithms are available for extracting corresponding word pairs from original and translated texts. These algorithms emphasize precision rather the recall ratio, because there are many large size corpora available for modern languages. However, in the case of classic texts such as the Bible (in ancient Greek and Hebrew), there are not enough original texts for large-scale quantitative analysis.

First, a new algorithm is designed to identify word pairs between the original text and the translated version. The algorithm incorporates three features: a word-for-word correspondence hypothesis, a recalculation of mutual information after the elimination of identified pairs, and an asymptotic threshold reduction. Through the combination of these features, recall rates improve by 20% compared to conventional methods and it is possible to extract multiple words corresponding to each of those in the word pairs.

Three Japanese translations of the Bible (*Colloquial Japanese*, *New Japanese*, and *The New Interconfessional Translation*) were analyzed using the proposed method, and vocabulary pairs of ancient Greek and Japanese were extracted.

4.3 Creating Networks on the Basis of Correspondences

In the next stage, two types of network are created on the basis of word correspondences, and the characteristics of translated words are extracted by calculating centrality values.

The network creation steps are depicted in Figs. 2–4. These identify the vocabulary correspondences (bipartite graph in Fig. 2), the relationship between words in the original languages (Fig. 3), and the relationships between words in the translated languages (Fig. 4).

Next, a centrality analysis (closeness, betweenness, and Bonacich) was applied to the extracted networks. The network centers reflect the conceptual center of the texts, because the central words signify that some concepts were more frequently used as an integrating concept or hypernym.

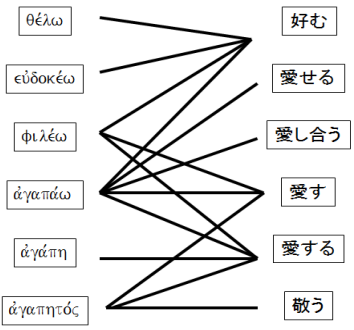

Fig. 2 Example of Corresponding Vocabularies

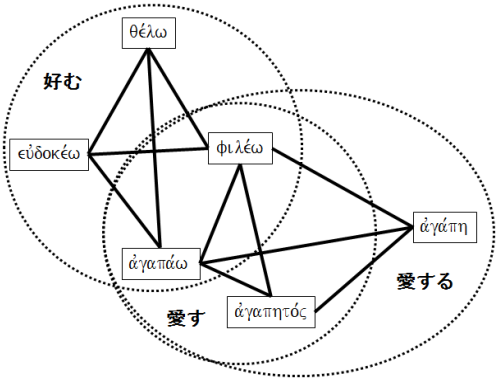

Fig. 3 Example of a Corresponding Network in the Original Language

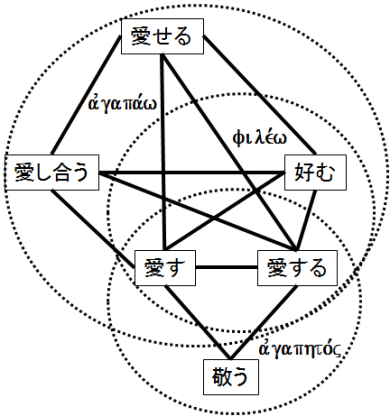

Fig. 4 Example of a Corresponding Network in the Translated Language

In addition to the relationships between original and translated languages, it is possible to make a network for the Bible that is composed of relationships between two original languages. The Old Testament was mainly written in Hebrew and the New Testament was mainly written in Greek. Therefore, the Bible describes the theology of the same God in two different languages. Because of this, modern Bible translations should interpret the conceptual theological relationships between Hebrew and Greek and translate them into one language. These relationships enable the correspondences between concepts in Hebrew and Greek to be analyzed using a modern translation as a medium.

As a case study, the relationship between the words *God* and *Lord* in the *New Revised Standard Version* (*NRSV*; Protestant translation) and *New American Bible* (*NAB*; Catholic translation) was analyzed. The results are depicted in Figs. 5 and 6, respectively.

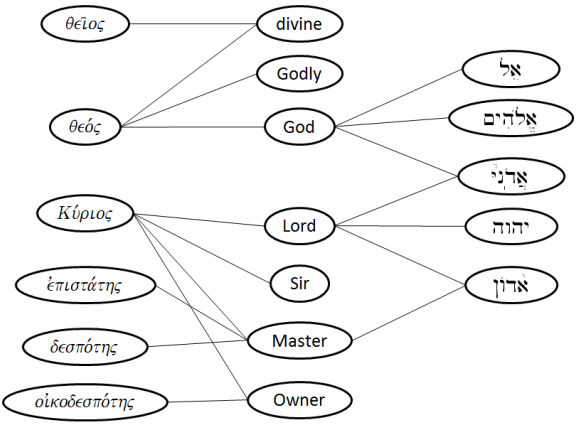

Fig. 5 Network of the conceptual relationship between *God* and *Lord* in *NRSV*

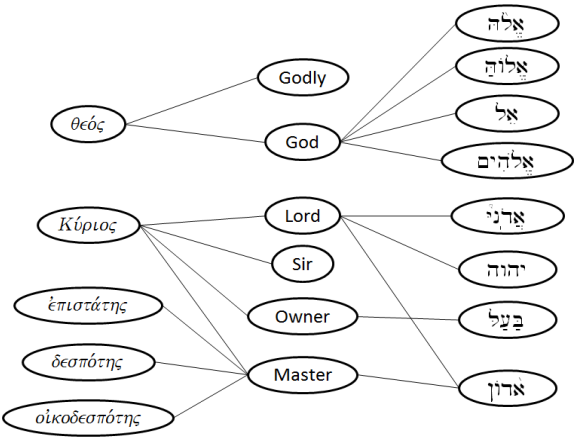

Fig. 6 Network of the conceptual relationship between *God* and *Lord* in *NAB*

4.4 Results and Discussion

The results from the vocabulary correspondence analysis of Japanese translations are summarized in Table 3. During translation, words that are important to a translator are translated carefully. Highly correspondent word pairs (in which concrete, mutual information is contained) between the original and translated text indicate characteristics of translation. Such highly correspondent word pairs were included in the results table.

The extracted results match the background of each translation. *Colloquial Japanese* is the oldest Japanese colloquial Bible. Fundamentalist Christians (equivalent to Evangelicals in the USA) were not satisfied with this version, and created the *New Japanese* version to emphasize the miracles and power of God. *The New Interconfessional Translation* was created to introduce a common Bible to the Catholic and Protestant churches, and therefore emphasizes peace among people of God.

Table 3 Characteristics of Translations

	Colloquial Japanese	**New Japanese**	**The New Interconfessional Translation**
Highly correspondent word	The Second Coming	Forgiveness and judgment on sins	Peace given by the Spirit, the Mission
Closeness centrality	Seeking for	*Fides quaerens intellectum*, the Mission	*Fides quaerens intellectum*
Betweenness centrality	Church as people who are called by God	Life is given by the Spirit	Coming of the Kingdom of God
Bonacich centrality	To know		

From the case study of conceptual relationships between two original languages, it can be inferred that *NRSV* interprets *God* and *Lord* as directly related concepts whereas *NAB* interprets them as fundamentally separated concepts. In other parts, word correspondences tend to be similar in the two translations. One Greek word was translated into several English words, and one English word was translated into several Hebrew words. Therefore, the level of detail in concepts about *God* and *Lord* is higher in Hebrew than in English or Greek.

From these results, it can be concluded that differences of interpretation between translations can be extracted by quantitative methods.

5 Synoptic Analysis of Religious Texts

5.1 Background

Undoubtedly, there are many cases where a group of people have sought to spread their message, and therefore developed a literature of *canonical documents*, but have encountered problems concerning the interpretation of the texts and the relationships between various individual documents. This kind of situation exists not only within Christianity, but also within other religions and schools of political thought. Such interpretative issues appear to have a direct influence on many matters in the modern world.

The central aim of this section is to develop a scientific information-technological method to analyze semantic differences that arise between multiple overlapping *canonical texts*. I believe that this method can be applied not just to the Bible, but also to the interpretation of systematic thinking embodied within collections of *canonical texts* in other spheres.

Specifically, this section introduces a method to analyze how central messages emerge from the existence of multiple overlapping *canonical texts*. This is applied to

the four traditional Gospels in the Bible, allowing a comparison with the Catechism of the Catholic Church. This gives a numerical illustration of precisely which messages Christianity has sought to convey with the selection of the four traditional Gospels [5].

5.2 *Creating Networks and Clusters*

The internal structure of each Gospel is divided into segments called pericopes. *Pericope* is an ancient Greek word meaning *cut-out*. Each pericope corresponds to a small segment of a biblical story that was transmitted orally.
In the Gospels, pericope units are numbered, such as No. 235. However, a particular pericope in one Gospel may correspond to multiple pericopes in another Gospel. This one-to-many relationship is due to the editing process, as each Gospel writer combined pericopes that he believed were related. Thus, if one author saw a connection between one pericope and several others, that particular pericope unit would be repeated in a number of sections within the Gospel. Accordingly, there are many pericopes in the four Gospels that have the same verses, because they were taken and edited from the same source pericope.

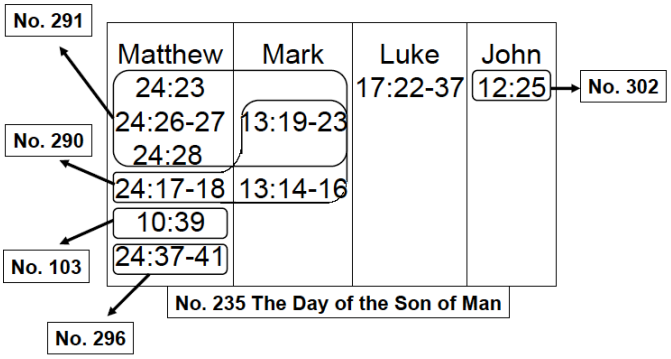

Fig. 7 Example of Pericope Relations

As Fig. 7 shows, pericopes containing verses in common with pericope No. 235 are Nos. 103, 290, 291, 296, and 302. This suggests that the writer of Matthew perceived some relationship among pericope Nos. 235, 103, 290, 291, and 296. Similarly, the writer of Mark imagined relationships between Nos. 290 and 291, whereas the writer of John made a link between pericope Nos. 235 and 302.

These pericope relationships can be converted into networks that regard pericopes as nodes and their relationships as edges. This study uses the *Synopsis of the Four Gospels* from the *Nestle-Aland Greek New Testament* (version 26) [8] as the data source of pericope relationships. This is believed to be the basis for various charts of pericope relations.

In order to identify the internal structure of the Gospels, the maximum connected subgraph was clustered and the core element was extracted. Four cores were extracted by combining node sharing cliques (Fig. 8 and Table 4). The following are the messages of the four cores:

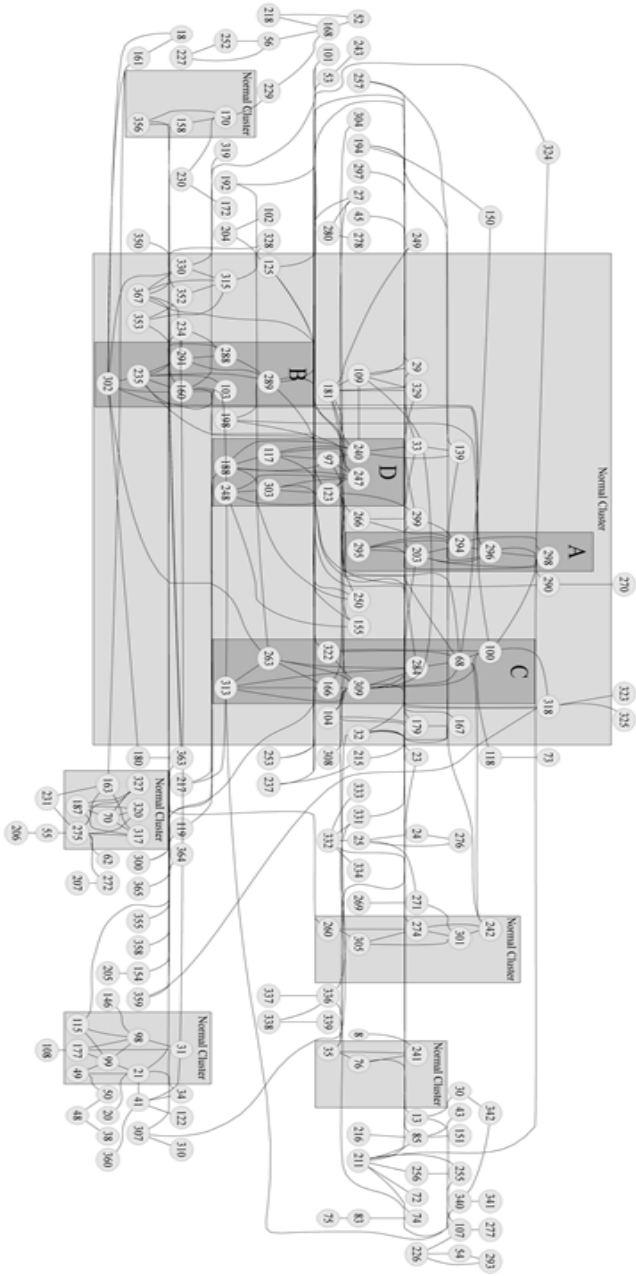

Fig. 8 Clustered Maximum Connected Partial Graph of the Four Gospels

Table 4 Contents of Clusters in the Network of the Four Gospels

Cluster	Pericope	Chapter and Verse of the Bible
A	203	Lk12:35–48
	294	Mk13:33–37
	295	Lk21:34–36
	296	Mt24:37–44
	298	Mt25:1–13
B	103	Mt10:37–39
	160	Mt16:24–28/Mk8:34–9:1/Lk9:23–27
	235	Lk17:22–37
	288	Mt24:3–8/Mk13:3–8/Lk21:7–11
	289	Mt24:9–14/Mk13:9–13/Lk21:12–19
	291	Mt24:23–28/Mk13:21–23
	302	Jn12:20–36
C	81	Lk6:37–42
	100	Mt10:17–25
	166	Mt18:1–5/Mk9:33–37/Lk9:46–48
	263	Mt20:20–28/Mk10:35–45
	284	Mt23:1–36/Mk12:37–40/Lk20:45–47
	309	Jn13:1–20
	313	Lk22:24–30
	322	Jn15:18–25
D	97	Mt9:32–34
	117	Mt12:22–30/Mk3:22–27
	188	Lk11:14–23
	240	Jn7:14–39
	247	Jn8:48-59

A) Preparation for the Day of Judgment because we do not know when it will come;
B) Foretelling persecution and recommending the path of discarding everything;
C) Teachings to the community of disciples;
D) Whether the miracles of Jesus were due to demons.

These teachings are believed to be the focus points of the old Church Fathers who canonized the New Testament.

In order to compare the results, the same analysis was applied to the *Catechism of the Catholic Church* [9]. Each item of the Catechism has a number, which is used as its ID. The ID numbers range from 1 to 2865. The relationships among the numbered items are complicated, and it is not unusual for one item to be related to several others. It is possible to construct a network by regarding items as nodes and relations as links, as for the pericopes in the Gospels.

As a result, ten clusters were extracted from the network of the Catechism's relationships (details of clusters are presented in Table 5). These clusters contained the following:

A) The Holy Spirit and the Sacraments;
B) The authorities of the Church;
C) The Virgin Mary;
D) The temptation of sin and malice; the miracles of the Christ;
E) Repentance, remittance, atonement;
F) Icons;
G) Human dignity in the figure of God;
H) Death;
I) Poverty;
J) Participation of laypeople in priesthood and prophecy.

Table 5 Contents of Clusters in Catechism

Cluster	Catechism ID Number
A	737, 788, 791, 798, 103, 1092, 1093, 1094, 1095, 1096, 1098, 1099, 1100, 1101, 1102, 1103, 1104, 1105, 1107, 1108, 1109, 1154
B	85, 86, 87, 88, 888, 889, 890, 891, 892, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040
C	143, 148, 153, 485, 489, 494, 506, 722, 723, 726, 963, 1814, 2087, 2609, 2617
D	394, 518, 519, 538, 540, 542, 546, 550, 560, 1115, 2119, 2816, 2849
E	980, 1424, 1431, 1451, 1455, 1456, 1459, 1473
F	476, 1159, 1160, 1161, 1162, 2129
G	225, 356, 1700, 1703, 2258
H	958, 1032, 1371, 1689
I	544, 2443, 2544, 2546
J	784, 871, 901, 1268

5.3 Discussion

The themes of the three most concentrated clusters in the Catechism (corresponding to A, B, and C) are *the Holy Spirit and the Sacraments*, *the authorities of the Church*, and *the Virgin Mary*. These themes are typical of the Catholic Church, but are not approved by many Protestant churches. Considering that the emphasized messages are typically different in Protestant churches, it is possible that the differences between Catholics and Protestants were especially considered and enhanced when the Catechism was edited. Another typically Catholic characteristic is the cluster concerning icons (F).

The message that disciples should abandon everything to follow Jesus is included in cluster I, which has a size of 4; however, this is a small part of the entire network.

Other than that, the problems of liturgy (A) and authority in the Church (B) are closed up. In the four Gospels, *serve each other* is an important message; however, the Catechism insists upon the authority of priests. One main characteristic of the four Gospels is an eschatological warning, but the Catechism does not emphasize this. Overall, the messages to the disciples became more suitable for religious organization and the eschatological messages were weakened.

6 Validation Methodology for Classic Rhetorical Structure

6.1 Background

Literary criticism is a promising field for interpreting the Bible precisely. This methodology analyzes the Bible as literature and examines its use of literary techniques. A marked literary characteristic of the Bible is its sophisticated structures, which comprise classic rhetorical structures such as chiasmus (in Fig. 9), concentric structures (in Tables 6 and 7), and parallelisms.

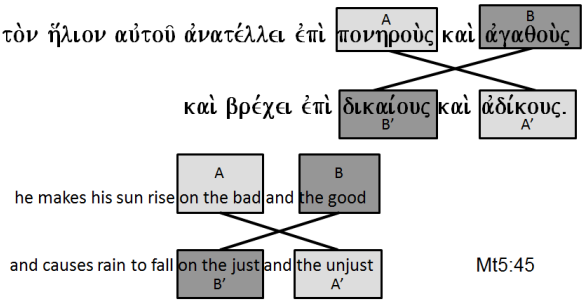

Fig. 9 Example of a Simple Chiasmus (Mt5:45)

Table 6 Example of Corresponding Pericopes in a Concentric Structure (Mk8:22–10:52)

	Part	Name of Pericope
A	8:22–26	Jesus cures a blind man at Bethsaida
B	8:27–30	Peter's declaration about Jesus
C	8:31–33	Jesus foretells his death and resurrection
D	8:34–38	Losing life for Jesus
E	9:1	A man who does not taste death
F	9:2–13	The transfiguration
G	9:14–29	The healing of a boy with a spirit
H	9:30–32	Jesus again foretells his death and resurrection
G'	9:33–50	Who is the greatest?
F'	10:1–12	Teaching about divorce
E'	10:13–16	Jesus blesses little children
D'	10:17–31	The rich man
C'	10:32–34	The third time Jesus foretells his death and resurrection
B'	10:35–45	The request of James and John
A'	10:46–52	The healing of blind Bartimaeus

Table 7 Example of a Corresponding Theme in a Concentric Structure (Mk8:22–10:52)

	Common Theme
A, A'	Healing the visually impaired
B, B'	Jesus is Messiah
C, C'	Foretelling death and resurrection
D, D'	Persecution and life
E, E'	Who enters the kingdom of the God
F, F'	Moses
G, G'	Evil spirit and child

There are several merits to identifying rhetorical structures in the Bible. It can clarify the divisions in a text; moreover, the correspondence of phrases signifies deeper interpretation. If the rhetorical structure is concentric, the main theme of that text is also clarified.

However, there are some problems regarding rhetorical structures. First, there is no clear definition of a valid correspondence. Some structures correspond by words or phrases, but more abstract themes may also be the element of correspondences. The length of the text unit is not uniform. Some structures are composed of phrases, whereas other structures are composed of pericopes. Therefore, a quantitative validation method for the rhetorical structure of the Bible is necessary [7].

6.2 Evaluation Algorithm for Rhetorical Structures

In this methodology, the relationships between each pericope in the rhetorical structure were first validated on the basis of the common occurrence of rare words and phrases. If corresponding pairs of pericopes more frequently include rare words and phrases, the probability of intentional arrangement is believed to be higher. Second, on the basis of the mean and standard deviation of a random combination of pericopes, the probability of accidental occurrences of common words and phrases in the test hypothesis is calculated. The common words and phrases are assumed to be normally distributed. The results are examined to determine whether the correspondences exceed the level of statistical significance.

Fig. 10 depicts an example calculation for the probability of random word pairs appearing when the rhetorical structure has nine pericopes and a particular word appears three times in that structure. The occurrence was counted in the form of a single word, a two-word phrase, and a five-word window; appearance thresholds of less than 10, 20, and 30% of pericopes were used for each form.

For the comparison, randomly divided pericopes were constructed from the same text. Two types of validation were also executed on the basis of the random division of pericopes. At first, the same combination of pericope patterns was applied to randomly divided pericopes. Next, a random combination of pericope patterns was applied to randomly divided pericopes. Thus, there are three estimates of validity:

A) Random combination and hypothetical pericope;
B) Hypothetical combination and random division;
C) Random combination and random division.

These three random situations were used to statistically validate the hypothesis of rhetorical structures. *Parallel Concentric Structures within the Bible* [10–12] was selected for the hypothesis. Tables 8–10 depict the results of the three types of validity estimation. The symbol ** signifies a 1% level of statistical significance, * signifies 5%, and + signifies 10%.

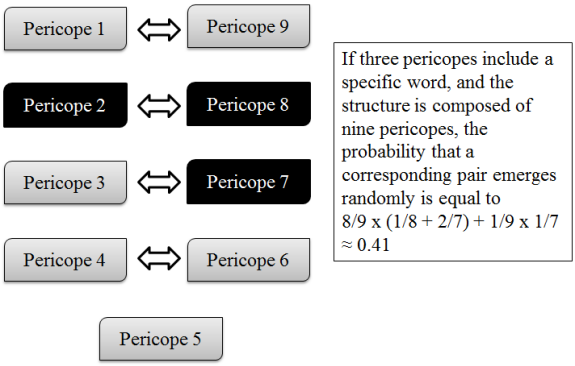

Fig. 10 Calculating the Probability of a Random Pair

Table 8 Statistical Evaluation of Rhetorical Structure 1 (Genesis-Deuteronomy)

		One word			Two-word phrase			Five-word window		
		10%	20%	30%	10%	20%	30%	10%	20%	30%
Genesis	**A**	*	**	**				**	**	**
	B	**		+						
	C	**		*						
Exodus	**A**			*	**	**	**			
	B		*	*	+	*	+	*	*	*
	C	+	**	*	*	*	*	**	*	*
Leviticus	**A**	+	*							
	B		*	+						
	C		**	*						
Numbers	**A**									
	B		**	+						
	C		**	*						
Deuteronomy	**A**	+	+							
	B					*	+		+	+
	C		+						+	*

Table 9 Statistical Evaluation of Rhetorical Structure 2 (Samuel-Ezekiel)

		One word			Two-word phrase			Five-word window		
		10%	20%	30%	10%	20%	30%	10%	20%	30%
Samuel 1, 2	A	*	*		**	*	**			
	B	*								
	C	*								
Kings 1, 2	A		**							
	B		*		*	*	**			+
	C		*		*	**	**			+
Isaiah	A	*			**	**	**			
	B						*			
	C				+		**			
Jeremiah	A	**	**	**	*					
	B		**							
	C		**			+			+	
Ezekiel	A								+	*
	B								*	+
	C							+	*	*

Table 10 Statistical Evaluation of Rhetorical Structure 2 (Matthew-Revelation)

		One word			Two-word phrase			Five-word window		
		10%	20%	30%	10%	20%	30%	10%	20%	30%
Matthew	A			*				*		
	B		+		+			*	**	**
	C							*	*	**
Mark	A	*	**	**						
	B	+	**	*						*
	C	+	**	*						
Luke	A	+	*							
	B									
	C	+								
John	A	*								
	B									
	C	+								
Acts	A									
	B	*								
	C	*								
Revelations	A	**	**	*	**	**	**	**	**	*
	B	*	*	**	**	**	**	**	**	**
	C	*	**	**	**	**	**	**	**	**

6.3 Results and Discussion

The results show that, in many texts of the Bible, the hypothesis is validated in terms of the relationships between words that occurred in less than 10% and 20% of the text, in the forms of either A or B or C.

A similar tendency in statistical significance among various books of the Bible seems to confirm that a unified rhetorical structure is included in these texts.

7 Software Application for Interpretation of the Bible

A project to develop a computer application to perform the above analyses is ongoing. The latest development version of a Java-based server–client model has been published [12].

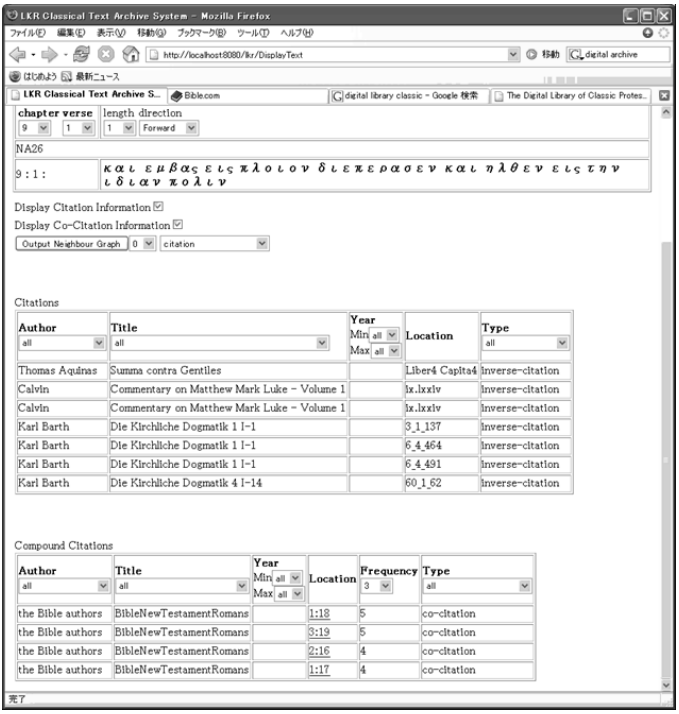

Fig. 11 Screenshot of the Interface for Browsing the Citation Database

The software contains an implementation of the algorithms described in this paper and data to support the interpretation. In addition, it includes general functions of Bible software, such as browsing and searching through several translations in parallel.

For the citation analysis function, a citation and reference database and browsing interface have been implemented (Fig. 11 shows a screenshot of the citations

and references browsing interface), as has a network analysis function for the citation network. Furthermore, a browsing interface for cited text is included so that researchers can interpret the Bible on the basis of information about the relationships between citations and references (due to copyright issues, this function is not available in the test version).

For the translation analysis, an asymptotic correspondence vocabulary presumption method has been implemented as the estimation algorithm. This function outputs the results of vocabulary estimation in CSV format by dividing data on the original text and the translated text into morphemes. The next version of the software will enable users to browse the results of vocabulary estimation for each translated text alongside the original and translated texts.

An algorithm for verifying the validity of the rhetorical structure has also been implemented. In addition, the software includes a database of rhetorical structure hypotheses, and users can browse rhetorical structures for each text location (Fig. 12 shows a screenshot of the rhetorical structure browsing interface).

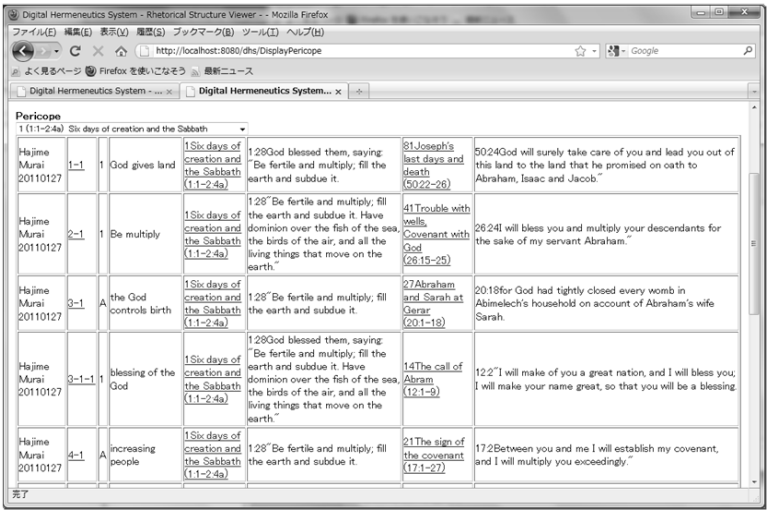

Fig. 12 Screenshot of the Interface for Browsing the Rhetorical Structure Database

8 Conclusions

This paper introduced four methods for the quantitative analysis of the interpretation of a canon of systematic thoughts. Although various quantitative methods can be used to analyze interpretations, it is difficult to analyze contextual information that is not described in texts (such as the circumstances of the author, historical facts, or cultural backgrounds). As in other fields where quantitative methods have not been applied, narratology and discourse analysis must also be used. To enable the scientific analysis of thought, it is necessary to resolve these problems and

enrich the algorithms for the quantitative analysis of interpretation. These new algorithms should also be easily available to many researchers. Therefore, computer software supporting interpretation should implement an interface that allows an integrated analysis and a flexible combination of the results of several other algorithms.

Acknowledgments. This work was supported by KAKENHI 22700256 (Grant-in-Aid for Young Scientists [B]).

References

1. Murai, H., Tokosumi, A.: Extracting concepts from religious knowledge resources and constructing classic analysis systems. In: Tokunaga, T., Ortega, A. (eds.) LKR 2008. LNCS (LNAI), vol. 4938, pp. 51–58. Springer, Heidelberg (2008)
2. Murai, H., Tokosumi, A.: Co-citation network analysis of religious text. Transactions of the Japanese Society for Artificial Intelligence 21(6), 473–481 (2006)
3. Murai, H.: Extracting the interpretive characteristics of translations based on the asymptotic correspondence vocabulary presumption method: Quantitative comparisons of Japanese translations of the Bible. Journal of Japan Society of Information and Knowledge 20(3), 293–310 (2010)
4. Landauer, T.K., Foltz, P.W., Laham, D.: Introduction to latent semantic analysis. Discourse Processes 25, 259–284 (1998)
5. Murai, H., Tokosumi, A.: Network analysis of the four Gospels and the Catechism of the Catholic Church. Journal of Advanced Computational Intelligence and Intelligent Informatics 11(7), 772–779 (2007)
6. Miyake, M., Akama, H., Sato, M., Nakagawa, M., Makoshi, N.: Tele-synopsis for biblical research: Development of NLP based synoptic software for text analysis as a mediator of educational technology and knowledge discovery. In: Conference on Educational Technology in Cultural Context (ETCC) in conjunction with ICALT, pp. 931–935 (2004)
7. Murai, H.: A study about validation methodology for classic rhetorical structure by utilizing digital archive. In: IPSJ SIG Computers and the Humanities Symposium 2011, vol. 8, pp. 211–218 (2011)
8. Kurt, A.: Synopsis of the four Gospels revised standard version. The United Bible Societies (1982)
9. Paul II, J.: Catechism of the Catholic Church (1992), http://www.vatican.va/archive/ccc/index.htm
10. Murai, H.: Parallel concentric structure within the Gospel of Mark. Journal of Catholic Theological Society of Japan 20, 65–95 (2009)
11. Murai, H.: The parallel concentric structures within Exodus. Society of Biblical Literature Annual Meeting (2010).
12. Murai, H.: Rhetorical Structure of the Bible (2012), http://www.valdes.titech.ac.jp/~h_murai/