**MURAI — Identifying**

<2 figures + 2 tables as figures)

**Identifying Synonymous Word Groups in the Synoptic Gospels: A Quantitative Analytical Approach**

Murai, H.

This paper is concerned with comparative analysis of the Synoptic Gospels, or, more specifically, the *synoptic problem*. In order to examine the synoptic problem scientifically, average synonymous word groups were first extracted based on the interrelationship between words in the Greek New Testament and various translations thereof. In the next step, those synonymous word groups were used as parameters for factor analysis. Thirty-four factors were identified, most of which pertain to theological concepts. Factor scores were then used to identify the major themes of each Synoptic Gospel. The major theme of Matthew was identified as ‘justice’; of Mark, ‘miracles’ and ‘salvation’; and of Luke, ‘missionaries’ and ‘what disciples should do’.

\* \* \*

The so-called *synoptic problem* pertains to the complexity involved in editing the first three New Testament Gospels (the Synoptic Gospels of Matthew, Mark, and Luke), largely due to the similarity in content between them. The similarities and differences between these texts have been analyzed extensively in the field of Bible studies using traditional humanities’ methodologies, though descriptive statistics were utilized in some cases (Rosché, 1960). In such studies, the synoptic problem has been mainly divided into two themes: the first regards how the texts are edited, and the second regards the differences between the texts. In other words, there exists both diachronic (process of editing) and synchronic (differences in the texts) research into the synoptic problem in the field of Bible studies.

However, with the rapid development of information technologies, it has become easier to quantitatively analyze texts, and in the digital humanities field, the Bible has become one of the major research targets. Regarding the synoptic problem, for example, Miyake et al. (2005) were able to automatically extract similar parts of the Synoptic Gospels and created Natural Language Processing–based synoptic tables. In another project, Murai (2007) quantitatively extracted background theological focus points about the editorial process based on relationships between similar text parts. Until now, however, such quantitative research into the synoptic problem has largely focused on the diachronic problem.

**Research Aim**

In light of this research gap, it is necessary to analyze the synoptic problem synchronically using quantitative methodologies. Although superficial differences between the texts, such as word or phrase frequency, have been analyzed manually since the 19th century, no quantitative analysis has been conducted into the differences between theological concepts.

One of the major difficulties in analyzing theological concepts is the diversity of expressions and words used in the original Greek New Testament. The same theological concept is expressed using many different words. Therefore, the relationship between synonymous words should be identified in order to conduct intertextual analysis of theological concepts. Moreover, because several theological concepts are mixed in each pericope (small story) in the Bible, it is necessary to use some methodology to separate out these concepts.

For the reasons outlined above, theological concepts cannot be extracted directly simply by counting the words. Therefore, in this research, the first step is to develop synonymous word groups and then to develop highly related word groups in the Synoptic Gospels by factor analysis of frequently appearing words. The resultant groups of highly related words would include theological concepts that are composed of co-occurring words.

**Method and Results**

*Identifying Synonymous Word Groups*

The nuance of word meaning in the Bible has been frequently contended in the field of Bible studies. There are numerous hypotheses and interpretations regarding the meanings of the original Greek words, so it is difficult to propose a neutral thesaurus for the biblical Greek. Therefore, in this research, an average of several interpretations was used to extract synonymous word groups.

The average interpretation of word relationships between the original Greek and subsequent translations was extracted. In order to identify the relationship between the original text and translations, the Asymptotic Correspondence Vocabulary Presumption Method (ACVPM) was used (see Murai, 2010). This method enabled the extraction of multiple word relationships for each word in order to form a bipartite network of original and translational words (see left network of Figure 1). Based on this network, two original words are connected when they share a corresponding translational word; in this way, a network of original words can be created (see right network of Figure 1) that signifies the relationships between original Greek words (Murai, 2013). Moreover, this method enables to compare characteristics of translational languages and background ideologies (Murai, 2012).

A separate network of original words can be constructed from each translation of the Bible. If many different networks of original words have a common edge between some two original words, it suggests that the two words are very similar in meaning. Therefore, by combining networks from several translations, and by extracting strong edges, average synonymous word groups are extracted as partial connected graphs of that network (Figure 2).

In this research, four English translations (the New Jerusalem Bible, the New King James Version, the New Revised Standard Version, and the New American Bible) and four Japanese translations (the Colloquial Japanese, the New Japanese, the Franciscan Japanese, and the New Interconfessional Translation) of the New Testament were analyzed and edges were extracted when more than four translations had that edge in common. Using this method, 224 similar word groups were obtained.

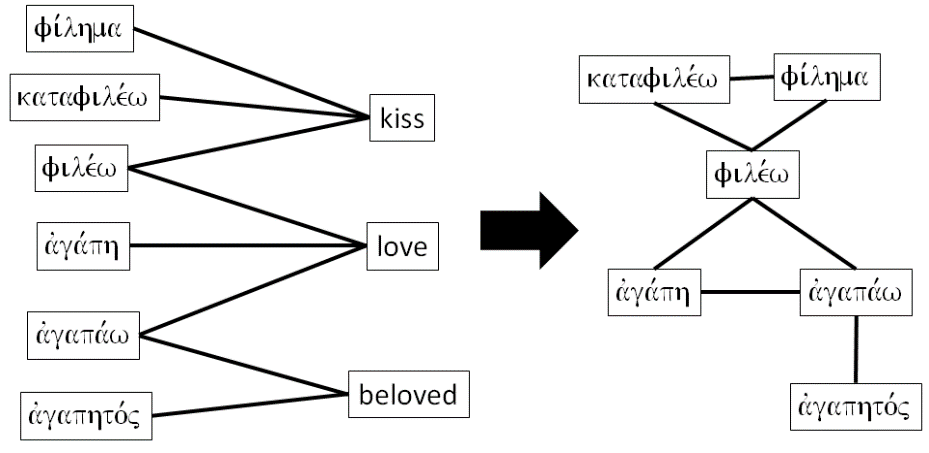


Figure 1. An example of extracting a network of original Greek words from a bipartite network of original and translational words.

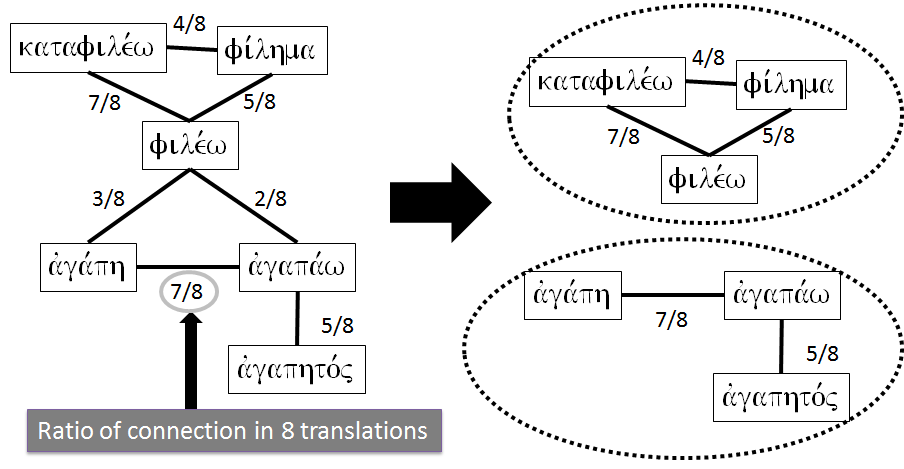


Figure 2. An example of creating combined networks and extracting stronger edges.

*Factor Analysis of the Synoptic Gospel*

In previous research, several methods have been used to extract concepts from texts—for example, LSA (Latent Semantic Analysis) or the Topic Model (Hoffman, 1999). Because LSA is used to analyze specific words, it is not suitable for whole text analysis. While the Topic Model enables the extraction of an arbitrary number of topics from target texts with a high degree of accuracy, it does not indicate the number of essential topics. Therefore, in this research, factor analysis, which allows indication of the number of essential topics, was used to extract factors composed of synonymous word groups from the Synoptic Gospels.

For this stage of the research, pericopes were adopted as the units of analysis. The Gospels of Matthew and Luke each contain 145 pericopes, while the Gospel of Mark contains 81 pericopes. From these, 175 synonymous word groups were selected, comprising those that appeared in more than 10% of all pericopes. The parameters used for factor analysis were the frequencies of those synonymous word groups in each pericope. The rotation method used was promax. According to the results of Parallel Analysis, 34 factors were identified, and the cumulative contribution ratio of those 34 factors was 0.785. The names of obtained factors and synonymous word groups with greater than 0.5 factor loadings are shown in Table 1. In Table 1, synonymous word groups are represented by one of the included Greek words.

The largest factors concern ‘the salvation of sinners’. These are followed by factors concerning ‘the beginning and the end’, ‘bread in the wilderness’, ‘journey to Jerusalem’, ‘hypocrisy of the Pharisees’, and ‘parables’. Apart from Factor 12 (‘numeral’), all identified factors were related to important theological concepts.

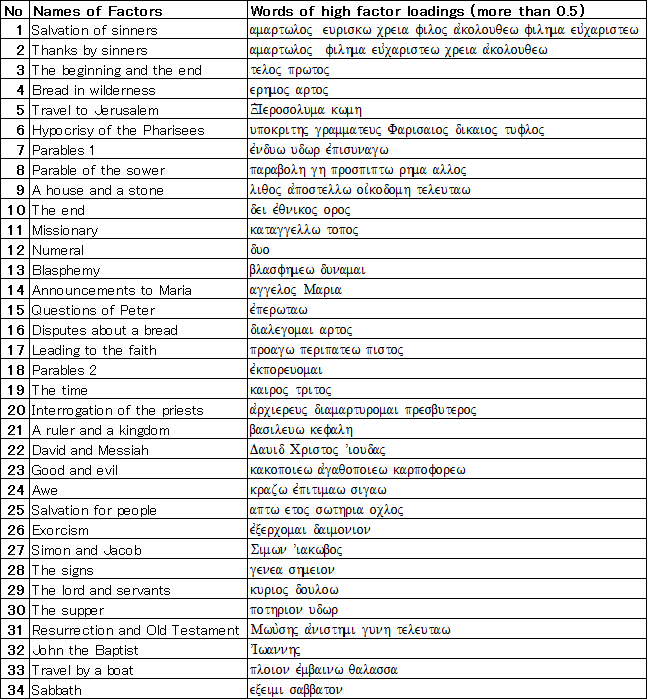


Table 1. Factor names and synonymous word groups.

In order to compare the major themes of the Synoptic Gospels, characteristic factors were identified based on factor scores. The average factor scores in each Gospel are shown in Table 2. In Matthew, Factors 6 (‘hypocrisy of the Pharisees’), and 23 (‘good and evil’) are found to be influential, which may allow us to consider the major theme of Matthew’s Gospel to be justice. In Mark, Factors 4, 8, 10, 13, 16, 17, 20, 21, 24, 25, 26, 27, 30, and 31 are influential, which may lead to the conclusion that Mark’s major themes are miracles (‘bread in the wilderness’, ‘exorcism’, ‘resurrection’, and ‘awe of people’) and salvation (‘parables of sower’, ‘leading to faith’, ‘salvation for the people’). In Luke, Factors 2, 5, 9, 11, 14, 19, and 29 are influential, so the major theme of Luke may relate to missionaries (‘travel to Jerusalem’, ‘missionary’) and what disciples should do (‘thanks by sinners’, ‘announcements to Maria’, ‘the lord and servant’).

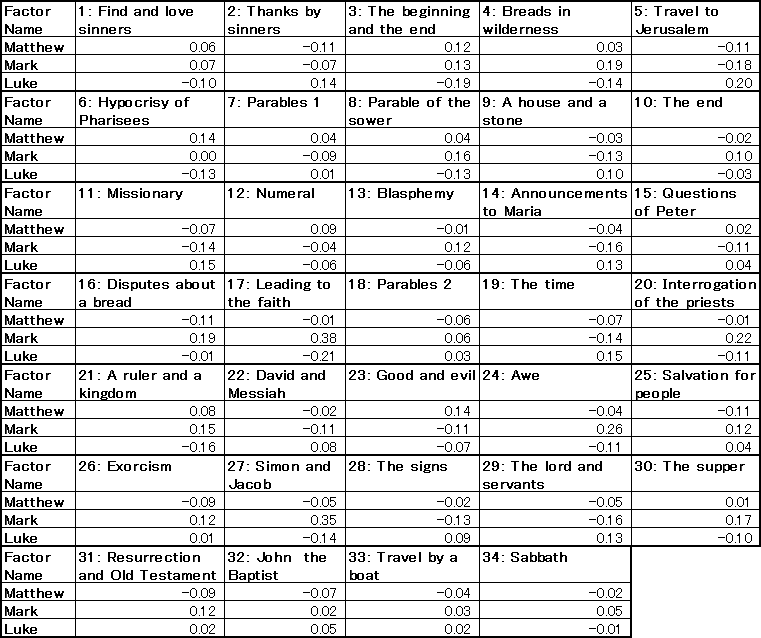


Table 2. Average factor scores in each Synoptic Gospel.

**Conclusion and Future Work**

In this research, average synonymous word groups were extracted using a quantitative method. These word groups were used to extract theological concepts from the Synoptic Gospels by utilizing factor analysis in a falsifiable way. Moreover, the theological differences between the three Synoptic Gospels were identified based on the average factor scores in each text. Unlike conventional exegetic methods, these results are reproducible, and third-party evaluation is possible because they are based on scientific methodology.

For future work, detailed analysis and consideration of obtained factors should be conducted. In addition, other books of the Bible should be analyzed and compared in order to understand the theological concepts of the Gospels and the Bible.

**References**

**Hofmann, T.** (1999). Probabilistic Latent Semantic Analysis. *SIGIR 99 Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley, CA, 15–19 August 1999, pp. 50–57.

**Miyake, M., Akama, H., Nakagawa, M. and Makoshi, N.** (2005). The Computed Synoptic Table-Tele-Synopsis for Biblical Research, *Proceedings of ACH/ALLC 2005*, Victoria, BC, 15–18 June 2005, pp. 152–54.

**Murai, H.** (2010). 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.

**Murai, H.** (2012). Semantic Networks between Texts in Different Language Based on the Asymptotic Correspondence Vocabulary Presumption Method. *Information Processing Society of Japan Symposium Series,* **2012**(7): 61–68.

**Murai, H.** (2013). Exegetical Science for the Interpretation of the Bible: Algorithms and Software for Quantitative Analysis of Christian Documents. *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing Studies in Computational Intelligence*, Vol. 492. Springer International Publishing, pp. 67–86.

**Murai, H. and Tokosumi, A.** (2007). Network Analysis of the Four Gospels and the Catechism of the Catholic Church’. *Journal of Advanced Computational Intelligence and Intelligent Informatics,* **11**(7): 772–79.

**Rosché, T. R.** (1960). The Words of Jesus and the Future of the ‘Q’ Hypothesis. *Journal of Biblical Literature,* **79**(3): 210–20.