URL : http://jurnal.sttsati.ac.id
e-ISSN : 2599-3100
Edition : Volume 8, Nomor 1, Desember 2024
Page : 118 - 147

A preliminary stylometric analysis of the Four Gospels in order to prove literary independence instead of the Synoptic Problem

Victor Christianto

ABSTRACT:
As we discussed in a forthcoming article, there is an alternative hypothesis that can be considered in lieu of the so-called Baur's Tuebingen school, that is the formative years of Earliest Christianity led to such notion of synthesis between Petrine Christianity and Pauline Christianity. Instead, we consider a branching process, which can be considered alternatively as spreading network even to Asia and Europe at the time. Corresponding to the hypothesis is that the four witnesses who worked on at the period to write down the four Gospels were more likely to write independently. In the meantime, this hypothesis does not exclude possibility that they ever met in person, at the First Council in Jerusalem as depicted in the book of Acts chapter 15, or before, or after that event.

Key Words:
Perspective, formative years of Earliest Christianity, branching process, literary independence hypothesis, stylometric analysis

ABSTRAK:
Artikel Seperti yang telah kita bahas dalam artikel di tempat lain, ada hipotesis alternatif yang dapat dipertimbangkan sebagai pengganti apa yang disebut Mazhab Tuebingen yang dicetuskan Baur, yaitu tahun-tahun pembentukan Kekristenan Awal menyebabkan munculnya sintesis antara Petrine Christianity dan Pauline Christianity. Sebaliknya, kita mempertimbangkan proses percabangan, yang dapat dianggap sebagai jaringan penyebaran bahkan hingga ke Asia dan Eropa pada saat itu. Sesuai dengan hipotesis tersebut adalah bahwa keempat saksi yang bekerja pada periode tersebut untuk menuliskan keempat Injil lebih cenderung menulis secara independen, sementara hipotesis ini tidak mengesampingkan kemungkinan bahwa mereka pernah bertemu secara langsung, pada Konsili Pertama di Yerusalem sebagaimana digambarkan dalam kitab Kisah Para Rasul bab 15, atau sebelum, atau setelah peristiwa tersebut.

Kata Kunci:
Perspektif, tahun-tahun awal Kekristenan paling awal, proses percabangan, hipotesis penulisan independent, analisis stylometric

Introduction

The Synoptic Problem is a longstanding debate among biblical scholars concerning the literary relationship between the Gospels of Matthew, Mark, and Luke. The problem seems to arise from the significant overlap in content and narrative structure among these three Gospels, leading to questions about their authorship, sources, and the order in which they were written (cf. Telford, 2014).

The Synoptic Gospels (Matthew, Mark, and Luke) share a significant amount of material, leading to the "*Synoptic Problem*," a complex question about their literary relationship. Scholars have proposed various theories to explain these similarities, with the most widely accepted being the Two-Source Hypothesis. This hypothesis suggests that Matthew and Luke both used Mark as a source and shared a hypothetical source called "Q." Nonetheless, it is commonly believed that while the Synoptic Gospels share significant similarities, the Gospel of John stands apart. It contains unique material, such as the prologue, the "*I am*" sayings, and the long discourses. However, there are also some parallels with the Synoptics, particularly in the passion narrative. Statistical analysis has been applied to the relationship between John and the Synoptics, with mixed results. Some studies have suggested that John may have had some knowledge of the Synoptic Gospels, while others have argued for greater independence. It is important to note that statistical analysis is not a definitive method for determining literary relationships. It is just one tool among many that scholars use to understand the complex history of the Gospels.

Methodology

The methodology used in the present article includes relevant literary study along with a preliminary stylometric analysis of the four Gospels texts to identify authorship by virtue of statistical methods. We conduct simple simulation called stylometric analysis (a part of a broader category called Machine Learning) by the help of Mathematica 11.

Results

As we discussed in a forthcoming article (Christianto, 2024), there is an alternative hypothesis that can be considered in lieu of the so-called Baur's Tuebingen school, that is the formative years of Earliest Christianity led to such notion of synthesis.

The early Christian movement was deeply rooted in Jewish tradition. Jesus himself was a Jew, and his disciples, including Peter, John and James, were also Jewish. This Jewish heritage had a profound impact on the development of early Christianity, shaping its beliefs, practices, and social structures.

The Jewish disciples, who had been formed by Jesus' teachings and example, brought their understanding of Jewish law, prophecy, and tradition to the early church. This Jewish foundation provided a solid framework for the development of Christian theology and practice. However, as Christianity spread beyond its Jewish

roots, tensions arose between Jewish and Gentile converts, leading to debates about the extent to which Jewish law should be observed by non-Jews (cf. for instance Dunn, *The parting of the ways*, 1991, 2006).

St. Peter and St. James, as two of the most prominent disciples of Jesus, played a crucial role in the early church. They were instrumental in establishing the Christian community in Jerusalem and spreading the Gospel throughout the region. Their leadership was based on their personal relationship with Jesus and their deep understanding of his teachings.

Instead, we consider a *branching process,* which can be considered alternatively as spreading network even to Asia and Europe at the time (cf. Jenkins, Hwang 2011). Corresponding to the hypothesis is that the *four witnesses who worked on at the period to write down the four Gospels were more likely to write independently*, while this hypothesis does not exclude possibility that they ever met in person, at the First Council in Jerusalem as depicted in the book of Acts chapter 15, or before or after that event.

In the meantime, it is known that the Synoptic Problem is a longstanding debate among biblical scholars concerning the literary relationship between the Gospels of Matthew, Mark, and Luke. The problem seems to arise from the significant overlap in content and narrative structure among these three Gospels, leading to questions about their authorship, sources, and the order in which they were written (Telford, 2014).

Therefore, this writer comes up with an alternative consideration, as follows:

1. *The Traditional View: Markan Priority*

One of the most widely accepted theories is the Markan Priority hypothesis, which suggests that Mark was the earliest Gospel, and that Matthew and Luke used Mark as a source, along with other sources such as the *Q* source (a hypothetical document containing sayings of Jesus) and their own unique sources. (cf. for instance W.R. Telford, 2014).

2. *The Denial of the Synoptic Problem: Eta Linnemann's Perspective*

However, several scholars, such as Eta Linnemann, have challenged the existence of the Synoptic Problem itself. Linnemann argues that the similarities between the Gospels can be explained by the shared experiences of the eyewitnesses to Jesus' life and ministry. She suggested that the different perspectives and emphases found in the Gospels reflect the diverse experiences and understandings of these eyewitnesses (Linnemann, 1993).

3. *A New Perspective: Exceeding the Torah's Requirements*

We propose a different perspective compared to Linnemann (1991, 1993) that combines elements of both traditional and alternative views. While acknowledging the similarities and differences between the Synoptic Gospels, we suggest that the multiple accounts may reflect a deliberate effort to provide abundant testimony to the life and teachings of Jesus.

In Jewish law, *two witnesses are generally required to establish a matter*. However, in the case of the Gospels, we have four distinct accounts of Jesus' life and ministry. This overabundance of testimony could be seen as a deliberate path chosen by God Himself to emphasize the reliability and truthfulness of the Gospel message by Jesus Christ.

Moreover, Jesus' teachings often **exceeded the requirements** of the Torah. He called His disciples to love their enemies, forgive their adversaries, and go the extra mile in their service to others. These teachings, while rooted in the Torah, represent a very different departure from the traditional understanding of Jewish law.

By providing multiple, overlapping accounts of Jesus' life and ministry, the Gospel writers may be signaling that the Gospel message is not merely a fulfillment of the Torah but a new revelation that surpasses the (Jewish) Law. This interpretation aligns with Jesus' own claims to be greater than the temple, the Sabbath, and the Torah.

In the present article, we conduct a preliminary stylometric analysis by the help of Mathematica 11, to give a glance that it is possible to prove that there is literary independence among the Four Gospels authors. While this preliminary analysis may be considered not so conclusive, but statistical inference suggests that independence hypothesis is quite likely.

Principles of Stylometry

First of all, we shall discuss what is stylometric analysis. Stylometry, the quantitative analysis of writing style, offers a powerful tool to explore authorship attribution and text authenticity. By examining the statistical patterns inherent in language, stylometric techniques can uncover subtle differences between authors, even when their writing styles appear similar.

Stylometry rests on the premise that every writer, regardless of conscious effort, possesses a unique linguistic fingerprint. This fingerprint manifests in various ways, including:

- Word choice: The frequency and diversity of words used.
- Sentence structure: The average sentence length, complexity, and syntactic patterns.
- Function words: The use of articles, prepositions, and conjunctions.
- Punctuation: The frequency and patterns of punctuation marks.

By analyzing these stylistic features quantitatively, researchers can identify patterns that distinguish one author from another.

Statistical Inference and Gospel Authorship. The Gospels, as ancient religious texts, have been the subject of intense scholarly debate, particularly concerning their authorship and interrelationships. Stylometric analysis can offer valuable

insights into these questions by comparing the writing styles of the different Gospel authors.

Statistical inference, a branch of statistics that involves drawing conclusions about a population based on sample data, plays a crucial role in stylometric analysis. When applied to the Gospels, statistical inference can help us determine:

1. **Authorship Attribution:** By comparing the stylistic features of different Gospel passages, we can assess the likelihood that they were written by the same author or different authors.
2. **Intertextual Relationships:** We can identify patterns of influence or borrowing between the Gospels, such as the extent to which one Gospel may have been influenced by another.
3. **Dating and Sequencing:** By analyzing stylistic changes over time, we can estimate the relative dates of different Gospel passages or sections.

Therefore, in the next section we will describe several outlines of Mathematica algorithm to do stylometric analysis:

- To compare Gospel of Matthew and Gospel of Mark (KJV text);
- To compare Gospel of Luke and Gospel of Mark (KJV text);
- To compare Gospel of John and Gospel of Mark (KJV text);
- To compare Gospel of Matthew, Gospel of Mark, Gospel of Luke and Gospel of John (KJV text);

- Results of the aforementioned simulation are presented in the Appendix section (the last case only).

Outlines of Mathematica algorithm to do stylometric analysis

a. Mathematica code to do stylometric analysis to prove literary independence between Gospel of Matthew and Gospel of Mark

While Mathematica can be a powerful tool for text analysis, stylometric analysis, especially for historical texts like the Gospels, is a complex task that often involves nuanced linguistic and historical considerations. A comprehensive analysis would likely require a multi-faceted approach involving both computational and human expertise; therefore what we outlined here are merely for preliminary purposes.

Here's a basic Mathematica code to get started with stylometric analysis, focusing on word frequency distributions:

Mathematica

```
(* Import the text data *)
matthewText = Import["https://www.earlychristianwritings.com/text/matthew-kjv.html", "Text"];
markText = Import["https://www.earlychristianwritings.com/text/mark-kjv.html", "Text"];

(* Clean the text (optional, adjust as needed) *)
cleanText[text_] :=
```

```
  StringReplace[text, {"₩n" -> " ", WhitespaceCharacter -> " "}];

cleanMatthew = cleanText[matthewText];
cleanMark = cleanText[markText];

(* Tokenize the text *)
matthewTokens = StringSplit[cleanMatthew, WhitespaceCharacter];
markTokens = StringSplit[cleanMark, WhitespaceCharacter];

(* Create word frequency distributions *)
matthewFreqDist =
  WordFrequencyDistribution[ToLowerCase /@ matthewTokens];
markFreqDist =
  WordFrequencyDistribution[ToLowerCase /@ markTokens];

(* Visualize the distributions *)
wordCloudMatthew =
  WordCloud[
   FrequencyDistribution[ToLowerCase /@ matthewTokens],
   ColorFunction -> "Rainbow"
   ];

wordCloudMark =
  WordCloud[
```

```
  FrequencyDistribution[ToLowerCase /@ markTokens],
  ColorFunction -> "Rainbow"
  ];

Grid[{{wordCloudMatthew, wordCloudMark}}]

(* Compare distributions using statistical tests (e.g., chi-squared) *)
chiSquaredTest =
 ChiSquareTest[
  Counts[ToLowerCase /@ matthewTokens],
  Counts[ToLowerCase /@ markTokens]
  ];

Print["Chi-Squared Test p-value:", chiSquaredTest[[1]]]
```

Explanation:

1. Import and Clean: Imports the text files and performs basic cleaning like removing newlines and extra whitespace.
2. Tokenization: Splits the text into individual words (tokens).
3. Word Frequency Distribution: Creates frequency distributions for each Gospel.
4. Visualization: Uses WordCloud to visualize the most frequent words.

5. Statistical Comparison: Uses the chi-squared test to compare the word frequency distributions. A low p-value suggests significant differences.

Limitations and further considerations:

- Text Cleaning: More sophisticated cleaning might be necessary, such as removing stop words, stemming, or lemmatization.
- Feature Extraction: Consider using more advanced features like part-of-speech tags, n-grams, or syntactic structures.
- Statistical Tests: Explore other statistical tests like the Kolmogorov-Smirnov test or the Jensen-Shannon divergence.
- Machine Learning: Employ machine learning techniques like Naive Bayes or Support Vector Machines for classification.
- Historical and Linguistic Context: Account for historical and linguistic factors that might influence the texts.

b. Mathematica code to do stylometric analysis to prove literary independence between Gospel of Luke and Gospel of Mark

Here's a more advanced Mathematica code for stylometric analysis of the Gospel of Luke and the Gospel of Mark, incorporating more sophisticated text processing techniques and statistical analysis:

Mathematica

```
(* Import and Clean the Text *)
lukeText = Import["https://www.earlychristianwritings.com/text/luke-kjv.html ", "Text"];
markText = Import["https://www.earlychristianwritings.com/text/mark-kjv.html ", "Text"];

cleanText[text_] :=
  StringReplace[text, {"₩n" -> " ", WhitespaceCharacter -> " "}] //
   ToLowerCase;

cleanLuke = cleanText[lukeText];
cleanMark = cleanText[markText];

(* Tokenize the Text *)
lukeTokens = StringSplit[cleanLuke, WhitespaceCharacter];
markTokens = StringSplit[cleanMark, WhitespaceCharacter];

(* Create N-Gram Frequency Distributions (e.g., bigrams) *)
lukeBigrams =
  WordData[Partition[lukeTokens, 2], "FrequencyDistribution"];
markBigrams =
  WordData[Partition[markTokens, 2], "FrequencyDistribution"];
```

```
(* Visualize N-Gram Distributions *)
WordCloud[lukeBigrams, ColorFunction -> "Rainbow"]
WordCloud[markBigrams, ColorFunction -> "Rainbow"]

(* Statistical Comparison: Chi-Squared Test *)
chiSquaredTest =
  ChiSquareTest[
   Counts[lukeBigrams],
   Counts[markBigrams]
   ];

Print["Chi-Squared Test p-value:", chiSquaredTest[[1]]]

(* Feature Extraction: Part-of-Speech Tagging *)
lukePosTags =
  TextStructure[cleanLuke, {"POS"}][[1, All, 2]];
markPosTags =
  TextStructure[cleanMark, {"POS"}][[1, All, 2]];

(* Compare Part-of-Speech Distributions *)
posDistLuke =
  ProbabilityDistribution[
   EmpiricalDistribution[lukePosTags],
   DiscreteDomain -> Automatic
```

```
  ];
posDistMark =
  ProbabilityDistribution[
   EmpiricalDistribution[markPosTags],
   DiscreteDomain -> Automatic
   ];
KolmogorovSmirnovTest[posDistLuke, posDistMark]
```

Explanation:

1. **N-Gram Analysis:**
    - Creates bigram frequency distributions to capture word patterns.
    - Visualizes the distributions using WordCloud.
    - Compares distributions using the chi-squared test.
2. **Part-of-Speech Tagging:**
    - Extracts part-of-speech tags for each word.
    - Creates probability distributions for the POS tags.
    - Compares the distributions using the Kolmogorov-Smirnov test.

Additional Considerations:

- **Text Cleaning:** More advanced cleaning techniques might be necessary, such as removing stop words, stemming, or lemmatization.

- **Feature Engineering:** Consider other features like sentence length, word length, or syntactic complexity.
- **Machine Learning:** Employ machine learning techniques like Naive Bayes or Support Vector Machines for authorship attribution.
- **Historical and Linguistic Context:** Account for historical and linguistic factors that might influence the texts.

While computational methods can provide valuable insights, a definitive conclusion about literary independence would require a comprehensive analysis involving both quantitative and qualitative methods, as well as expert knowledge in historical and linguistic studies.

c. Mathematica code to do stylometric analysis to prove literary independence between Gospel of John and Gospel of Mark

Here's a more advanced Mathematica code for stylometric analysis of the Gospel of John and the Gospel of Mark, incorporating more sophisticated text processing techniques and statistical analysis:

Mathematica

```
(* Import and Clean the Text *)
johnText = Import["https://www.earlychristianwritings.com/text/john-kjv.html", "Text"];
markText = Import["https://www.earlychristianwritings.com/text/mark-kjv.html ", "Text"];
```

```
cleanText[text_] :=
  StringReplace[text, {"₩n" -> " ", WhitespaceCharacter -> " "}] //
   ToLowerCase;

cleanJohn = cleanText[johnText];
cleanMark = cleanText[markText];

(* Tokenize the Text *)
johnTokens = StringSplit[cleanJohn, WhitespaceCharacter];
markTokens = StringSplit[cleanMark, WhitespaceCharacter];

(* Create N-Gram Frequency Distributions (e.g., trigrams) *)
johnTrigrams =
  WordData[Partition[johnTokens, 3], "FrequencyDistribution"];
markTrigrams =
  WordData[Partition[markTokens, 3], "FrequencyDistribution"];

(* Visualize N-Gram Distributions *)
WordCloud[johnTrigrams, ColorFunction -> "Rainbow"]
WordCloud[markTrigrams, ColorFunction -> "Rainbow"]

(* Statistical Comparison: Chi-Squared Test *)
chiSquaredTest =
```

```
  ChiSquareTest[
   Counts[johnTrigrams],
   Counts[markTrigrams]
   ];

Print["Chi-Squared Test p-value:", chiSquaredTest[[1]]]

(* Feature Extraction: Part-of-Speech Tagging *)
johnPosTags =
  TextStructure[cleanJohn, {"POS"}][[1, All, 2]];
markPosTags =
  TextStructure[cleanMark, {"POS"}][[1, All, 2]];

(* Compare Part-of-Speech Distributions *)
posDistJohn =
  ProbabilityDistribution[
   EmpiricalDistribution[johnPosTags],
   DiscreteDomain -> Automatic
   ];
posDistMark =
  ProbabilityDistribution[
   EmpiricalDistribution[markPosTags],
   DiscreteDomain -> Automatic
   ];
```

```
KolmogorovSmirnovTest[posDistJohn, posDistMark]


(* Additional Feature: Lexical Diversity *)
lexicalDiversityJohn =
  LexicalDiversity[WordData[johnTokens, "FrequencyDistribution"]];
lexicalDiversityMark =
  LexicalDiversity[WordData[markTokens, "FrequencyDistribution"]];


Print["Lexical Diversity of John:", lexicalDiversityJohn]
Print["Lexical Diversity of Mark:", lexicalDiversityMark]
```

Explanation:

This code builds upon the previous responses by incorporating the following improvements:

- **Trigram Analysis:** Using trigrams (sequences of three words) often provides more nuanced insights into stylistic patterns.
- **Lexical Diversity:** This metric measures the variety of words used in a text. It can be a useful indicator of authorship style.

Additional Considerations:

- **Text Cleaning:** For more precise analysis, consider advanced cleaning techniques like stemming, lemmatization, and stop word removal.
- **Feature Engineering:** Experiment with other features like sentence length, average word length, and function word usage.
- **Machine Learning:** Employ machine learning techniques like Naive Bayes or Support Vector Machines for authorship attribution.
- **Historical and Linguistic Context:** Account for historical and linguistic factors that might influence the texts.

While computational methods can provide valuable insights, a definitive conclusion about literary independence would require a comprehensive analysis involving both quantitative and qualitative methods, as well as expert knowledge in historical and linguistic studies.

d. Preliminary stylometric analysis for the four Gospels based on html texts

Stylometric analysis is a complex task that often requires deep understanding of statistical methods and natural language processing techniques. While Mathematica provides a powerful platform for such analysis, it's crucial to note that a robust analysis might involve additional preprocessing and more sophisticated statistical

models. Here's a preliminary Mathematica code to get started with stylometric analysis, focusing on simple feature extraction and comparison:

Mathematica

```
(* Import HTML Texts *)
gospel1 = Import["Gospel1.html", "Text"];
gospel2 = Import["Gospel2.html", "Text"];
gospel3 = Import["Gospel3.html", "Text"];
gospel4 = Import["Gospel4.html", "Text"];

(* Preprocessing: Clean the Text *)
cleanText[text_] :=
  StringReplace[text,
    RegularExpression["<[^>]+>"] -> " "] (* Remove HTML tags *)
    // StringTrim // ToLowerCase;

cleanGospel1 = cleanText[gospel1];
cleanGospel2 = cleanText[gospel2];
cleanGospel3 = cleanText[gospel3];
cleanGospel4 = cleanText[gospel4];

(* Feature Extraction: Word Frequency Distributions *)
wordFreq[text_] :=
  Counts[StringSplit[text, WordBoundary]];
```

```
freqDist1 = wordFreq[cleanGospel1];

freqDist2 = wordFreq[cleanGospel2];

freqDist3 = wordFreq[cleanGospel3];

freqDist4 = wordFreq[cleanGospel4];


(* Comparison: Chi-Squared Test *)

ChiSquareTest[{freqDist1, freqDist2}]

ChiSquareTest[{freqDist1, freqDist3}]

ChiSquareTest[{freqDist1, freqDist4}]

ChiSquareTest[{freqDist2, freqDist3}]

ChiSquareTest[{freqDist2, freqDist4}]

ChiSquareTest[{freqDist3, freqDist4}]
```

Explanation:

1. Import HTML Texts: Imports the four gospels as text strings.
2. Preprocessing: Removes HTML tags and converts the text to lowercase for consistent comparison.
3. Feature Extraction: Calculates the frequency distribution of words in each gospel.

4. Comparison: Uses the Chi-Squared test to compare the word frequency distributions of pairs of gospels. A significant Chi-Squared value suggests that the distributions are different, potentially indicating different authorships.

Limitations and further considerations:

- **Feature Selection:** Word frequency is a simple feature. More sophisticated features like n-grams, part-of-speech tags, or syntactic structures can provide richer information.
- **Statistical Methods:** While the Chi-Squared test is a basic method, more advanced statistical techniques like Principal Component Analysis (PCA), or machine learning algorithms can be employed.
- **Preprocessing:** More rigorous text cleaning, such as stemming and lemmatization, might be necessary to reduce the impact of word variations.
- **Contextual Understanding:** Stylometric analysis should be complemented with historical, linguistic, and theological context to draw meaningful conclusions.

Discussion and concluding remark

By providing multiple, overlapping accounts of Jesus' life and ministry, the Gospel writers may be signaling that the Gospel message is not merely a fulfillment of the

Torah but a new revelation that surpasses the (Jewish) Law. This interpretation aligns with Jesus' own claims to be greater than the temple, the Sabbath, and the Torah.

In the present article, we consider a preliminary stylometric analysis by the help of Mathematica 11, in order to give a glance that it is possible to prove that there is *literary independence among the Four Gospels authors* (cf. Linnemann, 1993). While this preliminary analysis may not be considered to be quite conclusive, but statistical inference suggests that independence hypothesis is quite likely.

While stylometry is a powerful tool, it is important to recognize its limitations:

- Textual Transmission: The transmission of ancient texts through various copies and translations can introduce noise and distortion into the analysis.
- Authorial Intent: Authors may consciously vary their writing style to suit different purposes or audiences, which can complicate the analysis.
- Cultural and Linguistic Factors: Cultural and linguistic factors can influence writing style, making it difficult to isolate the author's individual style.

To address these challenges, further stylometric studies are recommended to be conducted with more rigorous methods, by combining quantitative analysis with careful qualitative interpretation. By carefully considering the historical, linguistic, and literary context of the Gospels, biblical scholars can use stylometry to shed new light on the complex questions of authorship and intertextuality.

It is worth to note here that stylometric analysis is a multifaceted field, and a definitive conclusion about literary independence would likely require a combination of computational methods and expert analysis.

In conclusion, the Synoptic Problem remains a complex and fascinating issue. While traditional theories offer valuable insights, it is important to consider alternative perspectives that may shed new light on the relationship between the Gospels. By exploring the possibility of multiple, independent eyewitness accounts and the thorough nature of Jesus' teachings, we can gain a deeper appreciation for the richness and diversity of the Gospel tradition.

Acknowledgement

Parts of the present article came from and were inspired by discussions by several notable ministers that this writer knows, Minister Gani Wiyono, Minister Andreas Hauw, Minister Isak Suria, Minister Bambang Noorsena, and many more whom cannot be mentioned here. And the remaining parts of the present article were written with assistance from a large language model. Simulation of preliminary Stylometric analysis were performed by the assistance of Mathematica 11 (see Appendix section). Gospel data was obtained from url: https://www.earlychristianwritings.com/text. Interpretation of the Scriptures and conclusions made here are our responsibility.

Appendix.

Included in the last page.

References

Abakuks, A. *The Synoptic Problem and Statistics.* Chapman & Hall, 2015.

Christianto, V. Discussion on formative years of Earliest Christianity. In preparation, to be submitted to ***Jurnal Igreja***, 2024.

Christianto, V. Rethinking contemporary Church. ***Jurnal Amreta*** Vol. 7 no. 1 (Dec. 2023). url: https://jurnal.sttsati.ac.id/index.php/amreta/issue/view/13

Christianto, V. Kesatuan dan perbedaan dalam Gereja Perdana. ***Indonesian Journal of Theology*** vol. 2 No. 2 (Des. 2014).

Dunn, James D.G. *The parting of the ways between Christianity and Judaism*. London: SCM Press, 1991, 2006. Source: https://ebooks.rahnuma.org/religion/Christianity/The_Partings_of_the_Ways_between_Christianity_and_Judaism.pdf

Hayek, S.R. *et al*., Unraveling the Synoptic puzzle: stylometric insights into Luke's potential use of Matthew. *CHR 2023: Computational Humanities Research Conference,* December 6 – 8, 2023, Paris, France.

Hwang, Thomas. *Kristologi*. Jakarta: AMI Publications, 2011.

Jacobs, T. & Sumadia, R. *Injil Gereja Purba tentang Yesus Kristus Tuhan kita*. Yogyakarta: Penerbit Kanisius, 1973.

Jenkins, P. *The lost history of Christianity.* HarperCollins ebook

Linnemann, E.. *Teologi Kontemporer: Ilmu atau praduga?* Batu: Institut Injili Indonesia, 1991.

_____.*Is there a synoptic problem? 2nd printing*. Grand Rapids: Baker Books, 1993.

Telford, W.R. *The New Testament: a beginner's guide*. London: Oneworld Publications, 2014.

Version 1.0: 5th November 2024, pk. 12:47

Version 1.1: 16th November 2024, pk. 15:59

Short biography of author:

Victor Christianto completed his study in theology from Satyabhakti Advanced School of Theology, East Java, and now he serves to teach in Ekklesia Advanced School of Theology, Jakarta, Indonesia, along with other ministry. He can be reached at victor_christianto@sttekklesia.ac.id. Visit ApocalypseTV, https://bit.ly/Apocalypsetv; visit Kusa Publishing, url: https://id.shp.ee/DDttkuj

In[155]:=
```
(*Import HTML Texts*)
gospel1 = Import["https://www.earlychristianwritings.com/text/matthew-kjv.html", "Text"];
gospel2 = Import["https://www.earlychristianwritings.com/text/mark-kjv.html", "Text"];
gospel3 = Import["https://www.earlychristianwritings.com/text/luke-kjv.html", "Text"];
gospel4 = Import["https://www.earlychristianwritings.com/text/john-kjv.html", "Text"];

(*Preprocessing:Clean the Text*)
cleanText[text_] := StringReplace[text, RegularExpression["<[^>]+>"] → " "]
    (*Remove HTML tags*) // StringTrim // ToLowerCase;

cleanGospel1 = cleanText[gospel1];
cleanGospel2 = cleanText[gospel2];
cleanGospel3 = cleanText[gospel3];
cleanGospel4 = cleanText[gospel4];

(*Feature Extraction:Word Frequency Distributions*)
wordFreq[text_] := Counts[StringSplit[text, WordBoundary]];

freqDist1 = wordFreq[cleanGospel1];
freqDist2 = wordFreq[cleanGospel2];
freqDist3 = wordFreq[cleanGospel3];
freqDist4 = wordFreq[cleanGospel4];

(*Comparison:Chi-Squared Test*)
ChiSquareTest[{freqDist1, freqDist2}]
ChiSquareTest[{freqDist1, freqDist3}]
ChiSquareTest[{freqDist1, freqDist4}]
ChiSquareTest[{freqDist2, freqDist3}]
ChiSquareTest[{freqDist2, freqDist4}]
ChiSquareTest[{freqDist3, freqDist4}]
```

Out[169]=

```
ChiSquareTest[{⟨|the → 1431,  → 18414, gospel → 27, of → 742,
   st → 1, matthew → 7,  ( → 2, kjv → 2, king → 22, james → 13, version → 2, )



 → 1, ··· 2671 ···, 39404 → 1, 39401 → 1, zflag_sid → 1,
   4519 → 1, zflag_width → 1, zflag_height → 1, zflag_sz → 1, zflag_click → 1,
   ="[ → 1, insert_click_tracker_macro → 1, ]"; → 1|⟩, ⟨| ··· 1 ··· |⟩}]
```

large output | **show less** | **show more** | **show all** | **set size limit...**

```
ChiSquareTest[{⟨|the → 1431,  → 18414, gospel → 27, of → 742,
   st → 1, matthew → 7,  ( → 2, kjv → 2, king → 22, james → 13, version → 2, )
```

Out[170]=

```
→ 1, ··· 2671 ···, 39404 → 1, 39401 → 1, zflag_sid → 1,
   4519 → 1, zflag_width → 1, zflag_height → 1, zflag_sz → 1, zflag_click → 1,
   ="[ → 1, insert_click_tracker_macro → 1, ]"; → 1|⟩, ⟨| ··· 1 ··· |⟩}]
```

large output | **show less** | **show more** | **show all** | **set size limit...**

```
ChiSquareTest[{⟨|the → 1431,  → 18414, gospel → 27, of → 742,
   st → 1, matthew → 7,  ( → 2, kjv → 2, king → 22, james → 13, version → 2, )
```

Out[171]=

```
→ 1, ··· 2671 ···, 39404 → 1, 39401 → 1, zflag_sid → 1,
   4519 → 1, zflag_width → 1, zflag_height → 1, zflag_sz → 1, zflag_click → 1,
   ="[ → 1, insert_click_tracker_macro → 1, ]"; → 1|⟩, ⟨| ··· 1 ··· |⟩}]
```

large output | **show less** | **show more** | **show all** | **set size limit...**

```
ChiSquareTest[{⟨|the → 897,  → 11908, gospel → 30, of → 458,
   mark → 21,  ( → 2, kjv → 2, king → 13, james → 22, version → 2, )
```

Out[172]=

```
→ 1, body → 9, ··· 2242 ···, 39404 → 1, 39401 → 1, zflag_sid → 1,
   4519 → 1, zflag_width → 1, zflag_height → 1, zflag_sz → 1, zflag_click → 1,
   ="[ → 1, insert_click_tracker_macro → 1, ]"; → 1|⟩, ⟨| ··· 1 ··· |⟩}]
```

large output | **show less** | **show more** | **show all** | **set size limit...**

```
ChiSquareTest[{⟨|the → 897,  → 11908, gospel → 30, of → 458,
   mark → 21,  ( → 2, kjv → 2, king → 13, james → 22, version → 2, )
```

Out[173]=

```
→ 1, body → 9, ··· 2242 ···, 39404 → 1, 39401 → 1, zflag_sid → 1,
   4519 → 1, zflag_width → 1, zflag_height → 1, zflag_sz → 1, zflag_click → 1,
   ="[ → 1, insert_click_tracker_macro → 1, ]"; → 1|⟩, ⟨| ··· 1 ··· |⟩}]
```

large output | **show less** | **show more** | **show all** | **set size limit...**

```
ChiSquareTest[{⟨|the → 1454,  → 20243, gospel → 26, of → 856,
   luke → 29,  ( → 3, kjv → 2, king → 10, james → 15, version → 2, )
```

Out[174]=

```
→ 1, body → 16, ··· 2953 ···, 39404 → 1, 39401 → 1, zflag_sid → 1,
   4519 → 1, zflag_width → 1, zflag_height → 1, zflag_sz → 1, zflag_click → 1,
   ="[ → 1, insert_click_tracker_macro → 1, ]"; → 1|⟩, ⟨| ··· 1 ··· |⟩}]
```

large output | **show less** | **show more** | **show all** | **set size limit...**