



## Regular Article

Zdeňka Špiclová\*, Vojtěch Kaše

# Distant Reading of the Gospel of Thomas and the Gospel of John: Reflection of Methodological Aspects of the Use of Digital Technologies in the Research of Biblical Texts

<https://doi.org/10.1515/opth-2020-0111>

received January 9, 2020; accepted April 21, 2020

**Abstract:** The aim of this study is to demonstrate the applicability of selected methods of the so-called distant reading from the area of digital humanities for the interpretation of early Christian texts, specifically for approaching similarities and differences between the Gospel of Thomas and the Gospel of John. We use the term “distant reading” for the methods that allow us to explore, analyze, and visualize digitized textual data while using the tools from the area of data mining, natural language processing, or corpus linguistics. We want to explore whether methods from the field of digital humanities can allow for a sophisticated, quantifiable, and replicable comparison of the corpora of early Christian movements and thereby help to uncover the basic features of their theology and thus be a suitable complement to traditional exegesis and interpretation achieved by close reading.

**Keywords:** Gospel of Thomas, Gospel of John, digital humanities, distant reading, quantitative text analysis, document distances, frequency distribution, word co-occurrences

## 1 Introduction

This study<sup>1</sup> attempts to contribute to the ongoing scholarly debate concerning the relationship between the Gospel of Thomas and the Gospel of John: While one group of scholars emphasizes the similarities between the two, others highlight the differences. Therefore, the first aim of this study is to use selected methods of quantitative text analysis (QTA), which are based on standardized tools and algorithms from the fields of computational linguistics, information retrieval, and digital humanities, to explore whether these methods can help us to decide which of these two views – commonly based on more traditional exegetical interpretations – is better grounded in evidence.

On a more general level, we also attempt to demonstrate the applicability of these types of methods for the study of early Christian history in general. Over the last several years, these methods have become widely popular in the digital humanities and are referred to under the umbrella term “distant reading,” especially in respect to the macroanalysis of English literature.<sup>2</sup> However, it has been further suggested that distant reading might represent a fruitful methodology for the history of religions as well.<sup>3</sup> Following this research pathway, we want to explore to what extent these methods allow for a sophisticated,

---

1 This study was funded by the Grant System of the University of West Bohemia in Pilsen within the project SGS-2018-022.

2 Underwood, “A Genealogy of Distant Reading,” 1–12; Moretti, *Distant Reading*.

3 Slingerland, “The Distant Reading of Religious Texts.”

---

\* **Corresponding author: Zdeňka Špiclová**, Department of Philosophy, Faculty of Arts, University of West Bohemia, Pilsen, Czech Republic, e-mail: spiclovz@kfi.zcu.cz

**Vojtěch Kaše:** Department of Philosophy, Faculty of Arts, University of West Bohemia, Pilsen, Czech Republic; Department of History and Classical Studies, School of Culture and Society, Aarhus University, Aarhus, Denmark, e-mail: kase@kfi.zcu.cz

quantifiable, and replicable comparison of the two texts and their respective theologies. Going even further, we even believe that these differences might reflect differences in theologies of the early Christian movements associated with these texts.<sup>4</sup> In general, we propose that these methods represent a suitable complement to traditional exegesis and interpretation achieved by close reading of the texts. While our case study is based on a very small amount of documents, we envision that the very same methodology could be employed to analyze much larger corpora.

After the introductory section, which offers an overview of the recent debate concerning the relationship between the Gospel of John and the Gospel of Thomas, we proceed with a section on QTA consisting of three subsections. In Analysis 1, we first apply a rather general method to gain a rough comparison of the vocabulary of Jesus' sayings in the canonical<sup>5</sup> Gospels and the Gospel of Thomas. As an outcome of this analysis, we produce a series of plots visualizing distances between the documents. Subsequently, in Analysis 2, we focus more deeply on the most frequent terms in these sets of Jesus' sayings, which enables us to obtain more detailed insights concerning what is responsible for the differences between the documents. Finally, in Analysis 3, we utilize methods from distributional semantics to capture the differences in the meaning of a set of target words from these documents.

## 2 Historical analysis

### 2.1 The relationship between the Gospel of Thomas and the Gospel of John

In the following, we focus on the similarities and differences between the Gospel of Thomas and the Gospel of John that have been noted by a number of scholars. After the stage in which the Gospel of Thomas was compared mainly with the Synoptics, most researchers began to discover the relationship between the Gospel of Thomas and the Gospel of John. They noticed the usefulness in their comparison, as the Gospels share similar motifs and parallels. The question remains as to whether the common motifs were a sign of their similarity or the conflict between these two texts.

We can distinguish several attitudes<sup>6</sup> adopted by researchers regarding the relationship between the Gospel of John and the Gospel of Thomas.<sup>7</sup> The first group of researchers assumes that the Gospels of

---

<sup>4</sup> In this context, it is worth mentioning that there are scholars who question whether it is possible to reconstruct identifiable communities behind individual Gospels. They suggest that the consensus about the original intended audiences of the Gospels (the so-called Markan, Matthean, Lukan, and Johannine communities) has come about in New Testament scholarship without any substantial argument. This issue is mainly discussed in the publication *The Gospels for All Christians: Rethinking the Gospel Audiences*, edited by Richard Bauckham, the main proponent of this view. In his perspective, it is first of all necessary to realize that the Gospels are not Paul's epistles, i.e., it is not unproblematic to assume the specific addressees behind them. He emphasizes that mobility and communication in the first-century Roman world were exceptionally developed and many members of the early Christian churches regularly traveled. Moreover, we have concrete evidence for close contacts between churches in the period around which the Gospels were written. According to Bauckham, we cannot assume that it is possible to reconstruct the community from the text, because we certainly cannot take for granted that a Gospel was written in only one community. Bauckham even suggests that it is entirely possible that a Gospel was written over a period during which its author was a resident for a time in each of two or more very different communities. Therefore, the historical context of the Gospels is not the community of the author, but the early Christian movement in the late first century as a whole. For more see Bauckham, "For Whom Were Gospels Written?" 9–48; and Klink, *The Sheep of the Fold: The Audience and Origin of the Gospel of John*.

<sup>5</sup> We use the term "canonical" only as a quick reference term and have no ambition to suggest an approach to these texts, which is different to other texts originating in the same historical and cultural context (e.g., the Gospel of Thomas).

<sup>6</sup> According to Perrin and Skinner, we can distinguish four basic attitudes among researchers. See Perrin and Skinner, "Recent Trends in Gospel of Thomas Research," 77–81.

<sup>7</sup> For survey of research, see also Dunderberg, "John and Thomas in Conflict?" 361–362; Dunderberg, "Thomas' I-Sayings and the Gospel of John," 33–64 and Dunderberg, *The Beloved Disciple in Conflict?*, 14–46.

Thomas and John are based on common sources, regardless of whether this shared material was written or based upon an oral tradition.<sup>8</sup> This position was held by Raymond Brown,<sup>9</sup> who suggests that these two Gospels are based on Gnostic or Gnostic-like sources which functioned as an intermediary, and also by Quispel,<sup>10</sup> who attributes the parallels between the Gospels to their common source – collections of Jesus' sayings. The second group of researchers assumes that one Gospel depends on the other and differs in which one they consider to be earlier and which later. For example, Sell<sup>11</sup> and Marcovich<sup>12</sup> assume that the Gospel of Thomas is directly dependent on the Gospel of John. Koester,<sup>13</sup> on the other hand, suggests that some of the Nag Hammadi writings, including the Gospel of Thomas, Dialogue of the Savior and the Apocryphon of James, represent earlier stages of the development of Jesus' sayings – from a collection of sayings to fully developed dialogues and discourses in the Gospel of John. In addition, Koester assumes that the author of the Gospel of John reinterprets these sayings to reject the soteriology proposed by the Gospel of Thomas.<sup>14</sup> The third group of scholars perceives the relationship between the two Gospels to be even closer. They assume that these two Gospels have the same theological and social basis and even suggest that they were produced by the same community. This position is held very strongly by Mirkovic,<sup>15</sup> who concentrates on similarities between John and Thomas, suggesting that the similarities originate from the same *Sitz im Leben* – the Syrian syncretistic-wisdom tradition of wandering ascetics in the first century<sup>16</sup> – but also by Davies,<sup>17</sup> who assumes that the Gospel of Thomas is a text created by the same Christian community, which then drew up the Gospel of John. The thesis about the common origin of these two texts, which is considered to be responsible for similarities in theological terminology emphasized by these scholars, is also supported by Fieger.<sup>18</sup>

The fourth and fifth groups of researchers emphasize differences between the two Gospels and differ primarily in which of the two Gospels they consider to have originated earlier. Scholars who emphasize the dependence of the Gospel of Thomas on the Gospel of John especially include Popkes,<sup>19</sup> who observes that the use of the same metaphors in both Gospels is very different and considers the Gospel of Thomas to be a relecture of the Johannine text. The view emphasizing Thomasine priority and considering the Gospel of John as a polemical answer to Thomas' theology is maintained by Gregory J. Riley,<sup>20</sup> who focuses on the divergence in the concept of resurrection, or April DeConick,<sup>21</sup> who emphasizes the mystical character of Thomas' Gospel, against which the author of the Gospel of John takes his stance. This position is held also by Elaine Pagels,<sup>22</sup> who sees the most crucial moment in the different interpretations of primordial divine light from Genesis 1.

For the sake of clarity, in the next sections we focus especially on scholars maintaining strongly opposing views. Their arguments clearly illustrate the radical poles of both views. The following section

---

**8** Zöckler, Patterson and Koester, for example, hold this position. Zöckler, *Jesu Lehren in Thomasevangelium*, 101–102; 253–254. Patterson, *The Gospel of Thomas and Jesus*. Koester, *Introduction to the New Testament*, 178–180.

**9** Brown, "The Gospel of Thomas and St. John's Gospel," 175–177.

**10** Quispel, "Qumran, John and Jewish Christianity," 137–155

**11** Sell identifies these eight sayings as influenced by the Gospel of John: Prologue; GTh 8; GTh 13; GTh 28; GTh 38; GTh 43; GTh 91; GTh 92. Sell, "Johannine Traditions in Logion 61 of the Gospel of Thomas," 24–37.

**12** Marcovich suggests that GTh 11 refers to GJn 1:7; GJn 2:9; GJn 6:31–58; GJn 12:36. Marcovich, "Textual Criticism on the Gospel of Thomas," 53–74.

**13** Koester, "Gnostic Writings as Witnesses for the Development of the Sayings Tradition," 238–261.

**14** Koester, *Ancient Christian Gospels*, 113–128.

**15** Mirkovic, "Johannine Sayings," 1–25.

**16** *Ibid.*, 4, 23–24.

**17** Davies, "The Christology and Protology of the Gospel of Thomas," 663–82; Davies, *The Gospel of Thomas and Christian Wisdom*, 100–16.

**18** Fieger, *Das Thomasevangelium*, 8.

**19** Popkes, "Ich bin das Licht," 663–73.

**20** Riley, *Resurrection Reconsidered*, 100–56.

**21** DeConick, *Voices of the Mystics*.

**22** Pagels, *Beyond Belief*, 30–73 and Pagels, "Exegesis of Genesis 1," 477–96.

briefly introduces the main arguments for both perspectives, which should give us an overview of the whole spectrum of relevant scholarly opinions concerning the debate.

## 2.2 Similarities

The view emphasizing similarities and parallels between the two Gospels is primarily based on assumptions that the parallels and affinities between the Gospel of John and the Gospel of Thomas are the result of some common traditions. Proponents of this view tend to assume that the Gospel of Thomas and the Gospel of John originated around the same time<sup>23</sup> and had the same cultural background.<sup>24</sup>

Gilles Quispel assumes that the similarities that we can recognize in these two Gospels are based on a shared source of Jesus' sayings and, moreover, that the Gospel of John knows about the distinctive Palestinian traditions represented by Thomas.<sup>25</sup>

Stevan L. Davies goes further in his argument, claiming that the same community stands behind these two Gospels – in his view, the Johannine community first created the Gospel of Thomas and then the Gospel of John, because the Gospel of Thomas is a collection of sayings, an early stage of the Johannine community tradition.<sup>26</sup> According to Davies, both Gospels are derived from the tendency of early Christianity to apply to Jesus the terms and concepts of the Wisdom tradition.<sup>27</sup> Davies writes:

[...] I shall presume that when John and Thomas share christological conceptions and vocabulary derived from Sophiology, the similarities in derivation, conception, and vocabulary indicate a similarity of theological orientation, which is modified but not eliminated by the divergence of the literary forms and particular emphasis of the two documents. Thomas' theology is not identical to that of John, but neither is it "entirely" different.<sup>28</sup>

Davies especially emphasizes similarity in the usage of sapiential motifs<sup>29</sup> such as light,<sup>30</sup> a dualism of light and darkness,<sup>31</sup> Jesus as a teacher coming from above, i.e., from the Father,<sup>32</sup> a symbolism of drinking and eating,<sup>33</sup> a motif of failing in seeking and finding,<sup>34</sup> an ambivalent attitude toward the world,<sup>35</sup> a return to the beginning – to the original state of creation<sup>36</sup> – and a present eschatology.<sup>37</sup>

The similarities between the two Gospels are also emphasized by Mirkovic who does not agree that these two Gospels are literally dependent on one another.<sup>38</sup> Mirkovic assumes that these parallels originate from the same *Sitz im Leben*, which is in his view the popular wisdom of wandering ascetics – holy men

<sup>23</sup> They tend to date both Gospels around the end of the first century. Robinson, "The Johannine Trajectory," 232–68; Mirkovic, "Johannine Sayings," 1–4; Dunderberg, *The Beloved Disciple in Conflict?* 5, 114.

<sup>24</sup> Regardless of whether researchers emphasize parallels or differences between the two Gospels, there is no consensus on where to seek out their origins. Some scholars tend to place both Gospels in Syria, especially Riley's *Resurrection Reconsidered*, 177 or Mirkovic's "Johannine Sayings," 1–4; however, others emphasize the roots of the Gospel of John in the Palestinian context, see Quispel, "Qumran, John and Jewish Christianity," 144–6.

<sup>25</sup> Quispel, "Qumran, John and Jewish Christianity," 139–146.

<sup>26</sup> Davies, *The Gospel of Thomas and Christian Wisdom*, 115–116.

<sup>27</sup> Nevertheless, Davies himself admits that we can observe this tendency in other texts, such as the Gospel of Matthew, the Letter to the Colossians, Ephesians and First Corinthians. Davies, *The Gospel of Thomas and Christian Wisdom*, 100–6.

<sup>28</sup> Davies, *The Gospel of Thomas and Christian Wisdom*, 107.

<sup>29</sup> Davies, *The Gospel of Thomas and Christian Wisdom*, 106–113, based mostly on Brown's analysis.

<sup>30</sup> GJn 1:4; GJn 8:12; GJn 9:5; GTh 77; GTh 83.

<sup>31</sup> GJn 1:5; GJn 8:12; GTh 24; GTh 61.

<sup>32</sup> GJn 1:14; GJn 3:13; GJn 4:13–14; GJn 6:34–40; GJn 6:44–51; GTh 25; GTh 28; GTh 31–36; GTh 38; GTh 99.

<sup>33</sup> GJn 4:13–14; GJn 6:34–40; GJn 6:44–51; GTh 13; GTh 28; GTh 108; GTh 111.

<sup>34</sup> GJn 7:34; GJn 8:21; GJn 13:33; GJn 14:9; GTh 28; GTh 43; GTh 59; GTh 92.

<sup>35</sup> GJn 3:16; GTh 110.

<sup>36</sup> GTh 18; GTh 49; GTh 50.

<sup>37</sup> GJn 8:52; GJn 11:26; GTh 1; GTh 11; GTh 111.

<sup>38</sup> Mirkovic, "Johannine Sayings," 4.

and women in the first-century Syria. Moreover, Mirkovic assumes the theological proximity of the two Gospels, which, according to him, lies primarily in the fact that both communities believed in Jesus and considered him to be a self-conscious sage whose salvific words for us are true, flawless, and decisive. This concept comes from the aforementioned oral tradition of wandering holy men and women. The main task of their followers was also to keep, preserve and interpret the sayings of the master. Along these lines, this remarkable self-consciousness of Jesus in John and Thomas has its origin in the Stoic notion of “cataleptic impressions” – the concept in which a sage is able to discern truth from falsehood because he has a peculiar power of revealing his object and to teach so to speak without any need for external confirmation or logical proof.<sup>39</sup> This stoic concept, according to Mirkovic, is shared by both Gospels, which we see in the way they portray Jesus.

To conclude this section on parallels and similarities, it might be useful to make a short overview of the often-discussed shared features between the two texts, namely themes and metaphors. Concerning their theology, both Gospels share a negative attitude toward the world (GTh 28; GTh 42; GTh 56; GTh 80; GTh 81; GTh 110; GTh 111; GJn 1:10–11; GJn 7:7; GJn 8:23–26; GJn 9:39; GJn 12:25; GJn 12:31–32; GJn 14:18–31; GJn 15:18–19; GJn 16:33; GJn 17:9; GJn 17:14–19; GJn 18:36), emphasize the importance of the beginning (GTh 18; GTh 19; GJn 1:1–2; GJn 17:5), and experiment with the dualism of light and darkness (GTh 24; GTh 61; GJn 1:5; GJn 3:19–21; GJn 8:12; GJn 9:4; GJn 11:9–10; GJn 12:35). It seems that both Gospels also share the same view on Jesus and use the same metaphors associated with him – Jesus is the light (GTh 24; GTh 33; GTh 50; GTh 61; GTh 77; GTh 83; GJn 1:4–5; GJn 1:8–9; GJn 8:12; GJn 9:5; GJn 12:35–36; GJn 12:46), the living water (GTh 13; GTh 28; GTh 108; GJn 4:13–14), and he received everything from the Father (GTh 61; GJn 3:35; GJn 13:3; GJn 17:2; GJn 17:6–12). He is regarded as preexistent and is associated with the beginning and the origin of all things (GTh 18; GTh 19; GTh 77; GJn 1:3; GJn 8:58; GJn 17:5; GJn 17:24). Further, in both Gospels, Jesus is depicted as a true teacher who speaks with remarkable assurance; he needs no external proof for his testimony. The discipleship is a matter of election (GTh 49; GTh 50; GJn 6:70; GJn 13:18; GJn 15:16; GJn 15:19) and is about becoming like Jesus (GTh 106; GTh 108) – it means listening to his words (GTh 13; GTh 19; GJn 15:7) and keeping them (GTh 79; GJn 8:31). Also, according to both Gospels, those who keep his words will not taste death (GTh 1; GTh 111; GJn 8:51). In both Gospels, it is further emphasized that Jesus has departed, which makes his words even more important, as there will come a time when they will seek him and will not find him (GTh 38; GTh 59; GJn 7:34; GJn 16:16; GJn 16:28). John and Thomas both claim that to understand Jesus and his words means to achieve salvation and immortality (GTh 1; GTh 19; GTh 111; GJn 5:24; GJn 8:51–52; GJn 11:25–26; GJn 15,7; GJn 17) and in both Gospels Jesus’ incarnation is placed in contrast with human ignorance (GTh 28; GTh 31; GTh 43; GTh 92; GJn 1:9–11; GJn 8:12–30; GJn 8:31–59; GJn 14; GJn 16:12). The Gospel of Thomas and Gospel of John also work with a model of present/realized eschatology (GTh 3; GTh 51; GTh 113; GJn 3:36; GJn 5:24; GJn 11:25–26).

## 2.3 Differences

The second widespread position emphasizes the fact that, despite the abovementioned similarities, there are profound differences in theology and in the understanding of the role of Jesus. Without doubt, these scholars also recognize the points of contact between John and Thomas, especially in contrast to the Synoptics: Matthew, Mark, and Luke place emphasis on Jesus’ warning of the coming “end of time” and identify Jesus as God’s human agent or servant. On the contrary, John and Thomas emphasize that Jesus actually directs his disciples toward the beginning. They link Jesus to the divine light, which came into being in the beginning according to the account of creation, and understand him as God in human form. Thomas and John both say that this primordial light connects Jesus with the entire universe and with divinity, because all things were made through him, in John’s terminology through the “Logos.”<sup>40</sup>

<sup>39</sup> Ibid., 17–18.

<sup>40</sup> Pagels, *Beyond Belief*, 36–40.

However, according to these scholars, while the vocabulary and metaphors are identical, they are used to express a completely different understanding of the role of Jesus.

Gregory J. Riley claims that both communities were based in Syria and interacted closely.<sup>41</sup> However, Riley stresses the controversy between Johannine and Thomasine traditions, which is primarily in their view of the resurrection. In Riley's opinion, John promotes the fleshly resurrection of Jesus, while Thomas' view tends to deny the concept of physical resurrection.<sup>42</sup> He emphasizes the controversial character of John's depiction of the encounter with the risen Jesus. For the author of John's Gospel, the Doubting Thomas serves to depict the antihero, and his humiliation is also a literary construct through which John rejects the theology of the Thomasine movement.<sup>43</sup>

Another important figure in the research of the Gospel of Thomas, April DeConick, based her analysis on the traditio-rhetorical model. She also perceives the character of Thomas as a dramatic construct of the author of the Gospel of John; however, in her view, what divides them most is a Thomasine inclination to mysticism. According to the Gospel of John, the misunderstanding of the Gospel of Thomas lies in Thomas' claim that one must seek the path to Jesus, the path to ascend into heaven, and a visionary experience – *visio Dei*.<sup>44</sup> Thomas' conception, on the other hand, is very close to what we know from Jewish and later Christian mysticism, i.e., the image of God is hidden in each of us, regardless of whether the person is aware of it or not. Everyone has a divine spark within themselves, because they were created by God. This is precisely the kind of mysticism that the author of John's Gospel blames on the Thomasine community, arguing that salvation, and thus the coming of the Kingdom of God, can only be achieved through faith in Jesus. According to DeConick, the Johannine author created a "faith mysticism" in order to oppose the "visionary mysticism" and "mystical ascent soteriology" that can be found in the Gospel of Thomas.<sup>45</sup>

Elaine Pagels also holds a strong position that emphasizes the conflict between the two Gospels. She assumes that the Gospel of John was written to oppose certain Christology and, more precisely, to oppose a certain view of Jesus held by the Gospel of Thomas. Although the Gospel of John also speaks of Jesus as the light, the main message of the Gospel of Thomas is that this divine light – which comes from God and is most clearly manifested in Jesus – shines not only in him but potentially in anyone who is willing to seek this light. According to Pagels, "Thomas' Gospel encourages the hearer not so much to believe in Jesus, as John requires, as to seek for God through one's own divinely given capacity, since all are created in the image of God."<sup>46</sup> For Thomas, everyone in creation receives an innate capacity to know God – because we are all created in the image of the primordial light. Thus, one may come to recognize oneself and Jesus as identical twins.<sup>47</sup> The Gospel of John strictly opposes this view and claims that the image of God resides only in Jesus, because he is the incarnate Christ, Redeemer, true Logos, and God revealed in a human form.<sup>48</sup> According to Pagels, the author of the Gospel of John most likely knew the Gospel of Thomas and wrote his Gospel as an open controversy against this direction of thought and against all other Jewish and pagan mystical tendencies toward this optimistic vision of divine capacity or divine spark in each of us. The author of the Gospel of John decided to write his own Gospel to underscore that it is Jesus – and only Jesus – who is this primordial light and embodies God's word, and therefore speaks with divine authority. For this author, Jesus is unique, and only by believing in Jesus can we find the divine truth. John agrees with the Synoptic view that Jesus is called upon by God, that Jesus is the Messiah, rabbi, and prophet, but his teaching goes much further. The author of the Gospel of John appears to imply that the Synoptics did not fully understand who Jesus really was. He is more radical and dares to say that not only is Jesus chosen by God, he is even God himself – God who revealed himself to us in the form of Jesus. John declares that Jesus existed even before Abraham was born.<sup>49</sup> This is the crucial difference between John

<sup>41</sup> Riley, *Resurrection Reconsidered*, 177.

<sup>42</sup> *Ibid.*, 155–6.

<sup>43</sup> *Ibid.*, 100–26.

<sup>44</sup> DeConick, *Voices of the Mystics*, 85.

<sup>45</sup> *Ibid.*, 86–108; 109–32.

<sup>46</sup> Pagels, *Beyond Belief*, 34.

<sup>47</sup> The motif of the twin is very strong not only in the Gospel of Thomas but above all in the Acts of Thomas.

<sup>48</sup> Pagels, "Exegesis of Genesis 1," 481.



and Thomas as also expressed in the episode about Doubting Thomas.<sup>50</sup> From this perspective, this episode is an open polemic against Thomasine Christianity. Thomas is portrayed here as a disciple who does not believe and does not understand. However, after Jesus' revelation, he is surprisingly and perhaps sarcastically the only one who proclaims Jesus as Lord and even as God.<sup>51</sup>

Nevertheless, the view that we are actually dealing with a conflict between two texts here has been criticized by some researchers. For example, Skinner,<sup>52</sup> drawing on his analysis of characters appearing in the Gospel of John, claims that “the instances of Thomas’s misunderstanding identified by Riley, DeConick, and Pagels are actually part of a much larger presentation of characters in the Fourth Gospel. Thus, the supposition of John’s anti-Thomas polemic runs into difficulties unless it can be demonstrated that John’s similar treatment of Peter, Nicodemus, the Samaritan woman, etc. are also accounted for by positing numerous polemics aimed at individuals and/or communities within early Christianity.”<sup>53</sup> Dunderberg also holds a restrained position, suggesting that there should be more indicators and convincing signs of literary dependence to postulate either a closer relationship or conflict between the two Gospels. In his view, these two Gospels were probably part of more general discussions about certain similar topics and took different stances on them. Dunderberg also assumes that they did not know each other’s opinions.<sup>54</sup> Moreover, like Skinner, he also notes that Thomas is depicted negatively, but so are most other figures in the Gospel of John.<sup>55</sup>

With this in mind, it seems that despite the similarities that appear in both Gospels and the fact that we cannot say with certainty that the Gospels were formulated in conflict, we are dealing with a different theological message here. The author of the John’s Gospel probably knew and complemented the Synoptic Gospels; at the same time, he drove the view of Jesus into a radical belief that Jesus is a preexistent Logos, divine light, and incarnate God, and presents this concept through Jesus’ self-revealing monologues. Jesus explicitly designates himself several times as the light: Ἐγὼ εἰμι τὸ φῶς τοῦ κόσμου (I am the light of the world; GJn 8:12); ὅταν ἐν τῷ κόσμῳ ᾶ, φῶς εἰμι τοῦ κόσμου (While I shall be in the world, I am the light of the world; GJn 9:5); ἐγὼ φῶς εἰς τὸν κόσμον ἐλήλυθα, ἵνα πᾶς ὁ πιστεύων εἰς ἐμὲ ἐν τῇ σκοτίᾳ μὴ μείνῃ (I have come into the world as a light, so that everyone believing in me in darkness not should abide; GJn 12:46). For John, the motif of Jesus’ equality with God is very important, as it is already strongly expressed in the Johannine prologue: Ἐν ἀρχῇ ἦν ὁ Λόγος, καὶ ὁ Λόγος ἦν πρὸς τὸν Θεόν, καὶ Θεὸς ἦν ὁ Λόγος (In the beginning was the Word/Logos, and the Word/Logos was with God, and God was the Word/Logos; GJn 1:1) but also further in the gospel: ἐν ἐμοὶ ὁ Πατὴρ καὶ ἐγὼ ἐν τῷ Πατρὶ (In me (is) the Father and I (am) in the Father; GJn 10:38). The unity of the Father and the Son is also extensively presented in chapter 17, in Jesus’ prayer for the disciples: καὶ νῦν δόξασόν με σύ, Πάτερ, παρὰ σεαυτῷ τῇ δόξῃ ἣ εἶχον πρὸ τοῦ τὸν κόσμον εἶναι παρὰ σοί (And now glorify me, you, Father, in your presence with the glory I had before the world began with you; GJn 17:5); ἵνα πάντες ἔν ᾧσιν, καθὼς σύ, Πατὴρ, ἐν ἐμοὶ καὶ ἐγὼ ἐν σοί (That all one may be one as you, Father, (are) in me and I (am) in you; GJn 17:21); ἵνα ᾧσιν ἔν καθὼς ἡμεῖς ἔν (So that they may be one as we (are) one; GJn 17:22). This unity can never be achieved by disciples and believers, so the role of Jesus is emphasized as unique and irreplaceable. Thomas, who on the other hand perceives Jesus as a great sage, also uses the symbol of light and emphasizes the importance of his words, but assumes that the inner capacity to find divine light is in each of us and this is not an exclusive feature of Jesus. Jesus is the light: ἀνοκ πε π. ογοειν παει ετ. zi xw. ογ τηρ. ογ (I am the Light that is over all; GTh 77), but this light is potentially in each of us: ογη ογοειν φροπ μ. φογη ηη. ογ. ρμ. ογοειν λγω ς. ρ. ογοειν ε. π. κοσμος (There is light existing inside a man of light, and he makes light/shines to the whole world; GTh 24), for we

49 GJn 8:58.

50 GJn 20:24–29.

51 Pagels, *Beyond Belief*, 57–67.

52 Skinner, *John and Thomas – Gospels in Conflict?* 42–77.

53 Perrin and Skinner, “Recent Trends in Gospel of Thomas Research,” 81.

54 Dunderberg, “John and Thomas in Conflict?” 364–380; Dunderberg, *The Beloved Disciple in Conflict?* 1–13, 68–115, 199–208; Dunderberg, “Thomas’ I-Sayings and the Gospel of John,” 33–64.

55 Dunderberg, *The Beloved Disciple in Conflict?* 47–67.

have come out of this light: ε. γ. φαν. χοο. σ ηη. τη χε ητα. τετη. φωπε εβολ των χοο. σ ηα. γ χε ητα. η. ει εβολ σμ η. ουοειη η. μα εντ. λ. η. ουοειη. φωπε μμαγ εβολ σι τOOT. φ ουαατ. φ (If they say to you: “Where do you come from?” (then) say to them: We have come from the light, the place where the light has come into being by itself; GTh 50) and can become light again, be reunited and refilled with it: σOTAN ε. φ. φα. φωπε ε. φ. φηφ φ. ηα. μουσ ουοειη (If someone becomes [like] (God/the one who is equal), he will become full of light; GTh 61). Jesus’ role is not unique; anyone who follows Jesus and his words can be like him: ηεχε ις χε ηετα. σφ εβολ ση τα. ταπρο φ. ηα. φωπε η. τα. σε ληοκ σφ †. ηα. φωπε ε. ητοφ ηε λυφ ηεο. ηη ηα. ουοησ ερο. φ (Jesus says: “Whoever will drink from my mouth will become like me. I myself will become he, and what is hidden will be revealed to him”; GTh 108).

### 3 Quantitative text analysis

In the following text, we approach the historical research problem outlined above by means of QTA, i.e., utilizing computer algorithms and tools used in computational linguistics and computer science to quantitatively analyze human natural language data.<sup>56</sup> In the digital humanities, these methods are sometimes referred to under the umbrella term “distant reading,” drawing on the idea that the “close reading” of individual texts and passages that characterize traditional humanities scholarship might be fruitfully enriched by the employment of computational tools to “zoom out” from individual texts and passages in an attempt to discover more general patterns perceivable only from a larger distance.<sup>57</sup> The patterns of special interest are those concerning usage of certain words, either alone or in context, approached either within one or across many documents under scrutiny.

To proceed with our QTA, we first describe the textual data we have used in all our analyses and how we obtained them. Subsequently, we introduce three analyses employing these data to resolve particular research questions. Following good practice in quantitative scholarship, each analysis is divided into subsections of methods and results. We begin with the most general method, i.e., quantitatively comparing vocabulary employed by individual documents to calculate distances between these documents. Subsequently, we look more closely at these differences via an analysis of word frequencies. Finally, we employ a variant of word embeddings to trace differences in meaning of certain words across the documents.

#### 3.1 Data

At the core of our interest lie the similarities and differences in theology of the Gospel of John and the Gospel of Thomas. The first obstacle in approaching this problem while using QTA is associated with the fact that these two texts are fully extant only in different languages, namely ancient Greek and Coptic. Therefore, the texts cannot be compared directly in primary languages. To deal with this problem, a part of our analyses will be based upon English translations of the Gospels, forming a bridge between the two languages. However, wherever it is meaningful and possible, we use both Coptic and Greek. Furthermore, we also validate the results obtained for the Gospel of John by comparing them with results acquired by applying the same methods to other canonical Gospels, both in Greek and English.

For the English part of our analysis, we used the Revised Standard Version translation of the canonical Gospels<sup>58</sup> and Thomas O. Lambdin’s translation of the Gospel of Thomas.<sup>59</sup> We chose these versions also

<sup>56</sup> Jurafsky and Martin, *Speech and Language Processing*.

<sup>57</sup> Underwood, “A Genealogy of Distant Reading,” 1–12; Moretti, *Distant Reading*.



because they have been used in an online synopsis of the five Gospels produced in an HTML markup by John W. Marshall.<sup>60</sup> This way we were able to rely upon a single online resource. We extracted these texts from this resource using Python programming language, which we also used for all the subsequent analytical steps described in this article.<sup>61</sup>

In the case of the Greek text of the canonical Gospels, we used a morphologically tagged and lemmatized version of Eberhard Nestle's 1904 edition of the Greek New Testament, which has been made available online in a tabular format by biblicalhumanities.org via GitHub.<sup>62</sup>

For the Coptic version of the Gospel of Thomas, we turned to a machine-readable interlinear translation produced by Milan Konvička and made available online as a part of the Marcion 1.6.2 software.<sup>63</sup>

While approaching our task, we faced another obstacle – the fact that the Gospel of John and the Gospel of Thomas each represent a different genre, i.e., the first is framed as a narrative and the second as a set of Jesus' sayings. To overcome this obstacle while making our analyses of similarities and differences in the theology of these two Gospels more meaningful, we extracted from all texts only the words attributed to Jesus, i.e., the so-called Jesus' sayings. All analyses introduced below in Greek, Coptic, and English are based only upon this subset of the texts.

As a final step in preprocessing the data, we proceeded with lemmatization and word type filtering. Lemmatization is a process producing a dictionary form of a word (or: lemma), which can be conducted either manually or by means of automatic computational tools. It is commonly done simultaneously with morphological analysis of the word in the text, classifying whether it is a noun, adjective, etc. In the case of the Greek New Testament, this task was rather straightforward, since we could rely upon lemmatization and morphological analysis being a part of the edition we worked with. Thus, for each of Jesus' sayings in the Greek versions of the canonical Gospels, we could easily filter the data only for certain word types, namely nouns, adjectives, and verbs, and extract only the lemmata for these types.

In the case of the Coptic version of the Gospel of Thomas, instead of lemmata, we were able to use ID codes associated with each lexeme,<sup>64</sup> referring to online grammar and dictionary.

In the case of English, we had to make the lemmatization on our own, and for this purpose we used an automatic lemmatization tool in the Python NLTK library.<sup>65</sup> We further filtered words within this lemmatized dataset by automatically removing all stopwords using the same Python library.

As a result of these cleaning and preprocessing procedures, we obtained three highly comparable datasets. Since the three languages employ the word types differently and since we used different methods for filtering them, different language versions are slightly different in size. In the case of Jesus' sayings within canonical Gospels, there are 13,780 words in total in their English versions compared to 14,508 words in total in their Greek versions, which equals 0.94982. In the case of 415 sayings identified within the Gospel of John, we have a ratio of 2,644:2,817 words, which equals 0.93859. The cleaned and preprocessed version of Jesus' sayings in the English translation of the Gospel of Thomas consists of 1,654 words, while the Coptic version of the same text consists of 1,751 words, which equals a ratio of 0.9446. These words form 241 and 273 sentences, respectively,<sup>66</sup> with an average length of 6.86307 and 6.41392 words, respectively. Again, this is highly comparable to Jesus' sayings in the Gospel of John, with an average length of 6.37108 words.

<sup>58</sup> *The Holy Bible: Revised Standard Version*. National Council of Churches of Christ, 1973.

<sup>59</sup> Robinson, *The Nag Hammadi library*.

<sup>60</sup> <http://sites.utoronto.ca/religion/synopsis/>.

<sup>61</sup> All Python scripts in the form of Jupyter notebooks are accessible via GitHub: <https://github.com/kasev/DirGoT>.

<sup>62</sup> <https://github.com/biblicalhumanities/Nestle1904/tree/master/morph>.

<sup>63</sup> <http://marcion.sourceforge.net/nag-hammadi-library/gospel-of-thomas-en.html#transl>.

<sup>64</sup> For example, ID code "1–57" refers to εΙΩΤΗ (father).

<sup>65</sup> Bird, Klein, and Loper, *Natural language processing*.

<sup>66</sup> The Coptic version has slightly more sentences, which is probably caused by a difference in sentence splitting chosen by the editors. However, this difference should not devalue our overall results.

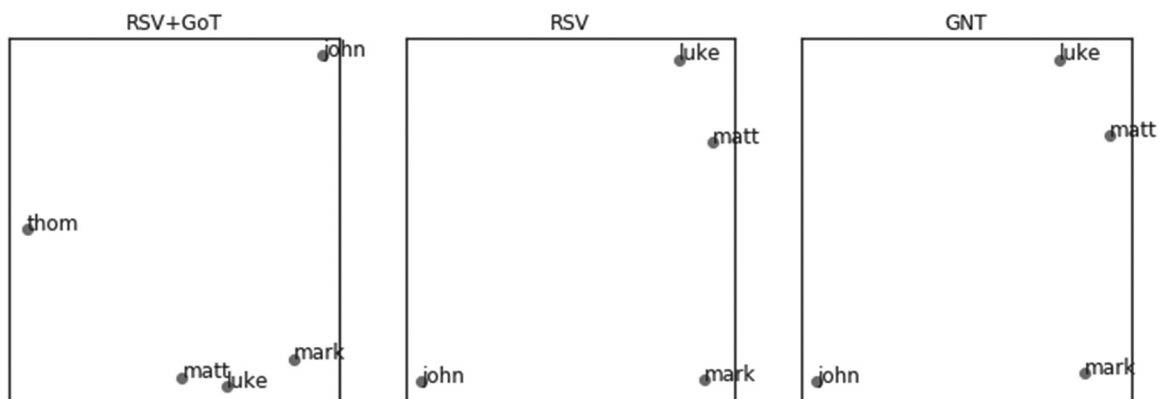
## 3.2 Analysis 1: document distances

### 3.2.1 Methods

In the first analysis we carried out, we used the dataset of Jesus' sayings to formally evaluate similarity between the five/four Gospels by drawing on the vocabulary they employ. This method is based on the idea that documents sharing a larger amount of words are more similar to each other than documents that have less words in common. This similarity is calculated by a comparison of multidimensional vectors based on document-term co-occurrence matrices, by which for each pair of documents within a corpus we obtain one value expressing their numerically expressed similarity. This value can then be used to produce a visualization of distances between documents in a two-dimensional space, where more similar documents are plotted as being closer to each other within the space. Here the similarity between two documents is inverted to express their distance on a scale between 0 and 1, with 1 meaning that the documents are completely unrelated. However, the projection into a two-dimensional space has to be interpreted with caution, since it represents only a rough approximation of the numerical values.

### 3.2.2 Results

We calculated the distances between three sets of documents containing the sayings of Jesus. The first set is based on the five Gospels in the English translation (*RSV + GoT*), the second is based on the four canonical Gospels in the English translation (*RSV*), and the third is based on the four canonical Gospels in Greek (*GNT*). The results can be seen in Figure 1.



**Figure 1:** Visualizing distances between canonical Gospels and the Gospel of Thomas.

On the left plot, visualizing distances between the five Gospels in the English translation, we clearly see that the Gospel of Thomas and the Gospel of John are quite far from each other. Expressed numerically, their distance is 0.40978. This is a higher distance than the one dividing them from the Synoptics. For instance, the distance between Thomas and Mark is 0.28595, while the distance between John and Mark is 0.35930. These are still much greater distances than the ones dividing the Synoptics from one another: e.g., the distance between Matthew and Luke is only 0.08327.

The middle and the right plots are here to demonstrate that we obtain the very same results when we compare the canonical Gospels in English and Greek. This has also been evaluated statistically by means of the Pearson correlation ( $\rho = 0.99564$ ,  $p < 0.001$ ). This is an important finding, as it demonstrates that the similarities and differences in Jesus' vocabulary across the Gospels we observe in the English text are actually not an artificial product of the translation but rather closely mirror the original.

The results of this method, which identify the large distance between John and Thomas in comparison to the rest of the Gospels, appear to support the second perspective within the scholarly debate, i.e., the

one that highlights differences between these two Gospels. With this in mind, it does seem a bit less likely that these two texts shared the same sources or that they could have been created by the same community wanting to proclaim the same theological concept. Hypothetically, such a large distance between the two texts, even farther than their distance from the Synoptic Gospels, might explain the thesis, which considers that the Gospel of John proclaims a theology that is not in accordance with the theology of the Gospel of Thomas. However, with some degree of certainty, it can only be said that the distance between these texts indicates their differences. The conclusion that the contradiction or direct controversy highlighted by some of these researchers may be the reason why these texts are so far removed cannot be directly deduced from this analysis.

It is also possible that this analysis could support Dunderberg's thesis that there is no literary relationship between these two Gospels. Dunderberg argues that it is necessary to consider whether the similarities between the two Gospels imply some particular connection or conceptual relationship. It should be considered whether these parallels are also commonly attested outside the two Gospels in early Christian literature. Based on the analysis of I-sayings in the Gospel of Thomas and their comparison with similar statements in John, Dunderberg concludes that no intimate relationship between the Gospel of Thomas and the Johannine writings can be assumed. In his view, even close contact between the communities behind them cannot be postulated – they shared no similar theology, nor were they in conflict. Dunderberg suggests that each of them employed similar Jewish and Christian traditions in quite different ways and did not know each other's positions. Moreover, the emphasis that is common to both Gospels can be found also in other early Christian texts.<sup>67</sup> The results of our first analysis might be interpreted also in agreement with these points, but since we are interested in whether distance reading can support the view of either the researchers who emphasize similarities or the researchers who reveal differences, our further analysis will focus on what terms appear most often in the gospels.

### 3.3 Analysis 2: frequency distributions

#### 3.3.1 Methods

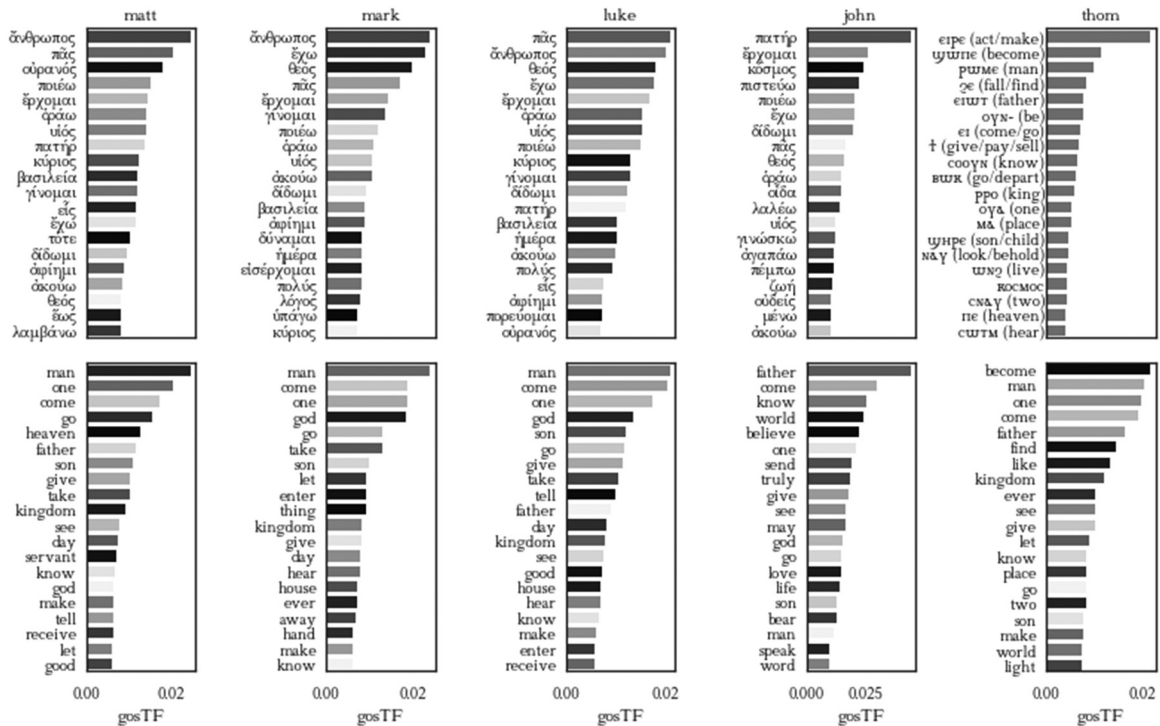
Our second analysis focused in more detail on the findings obtained by means of the first analysis. It consisted of measuring word frequencies in both the English and Greek versions of the canonical Gospels and in both English and Coptic versions of the Gospel of Thomas. In the case of texts available in Greek and English, we made two measurements for each unique word within each Gospel. First, we calculated its term frequency within the Gospel (*gosTF*) by counting all its occurrences and dividing this number by the total number of words within the Gospel. Second, to reveal words of special importance for each individual Gospel, we generated a second metrics, dividing the *gosTF* value by a term frequency of the same word across all Gospels together (i.e., five Gospels in the case of the English text and four canonical Gospels in the case of the Greek text). In instances where the value of these second metrics is significantly higher than 1, we can conclude that we are dealing with a term specific to the given Gospel. In the case of the Coptic version of the Gospel of Thomas, we conducted only the first measurement, as we do not possess other Gospels in a suitable format for comparison.

#### 3.3.2 Results

To report and interpret results of this analysis, we produced Figure 2 with 10 subplots: 4 for the canonical Gospels in Greek, 1 for the Coptic version of the Gospel of Thomas, and 5 for English versions of the canonical Gospels in tandem with the Gospel of Thomas. Each subplot depicts the 20 most frequent words

<sup>67</sup> Dunderberg, *The Beloved Disciple in Conflict?*, 113–5; Dunderberg, “Thomas’ I-Sayings and the Gospel of John,” 41–3, 63–64.

contained in Jesus' sayings within the Gospel. Although there is evidently an overlap between the Greek and English version of the canonical Gospels, we should not ignore the fact that there are also some differences, especially in verb usage. Therefore, in the subsequent analysis, we will pay special attention to nouns.



**Figure 2:** Twenty terms with the highest *gosTF* in the five Gospels. Where differentiated, the color scale depicts the term frequency within the Gospel in ratio to its frequency in all Gospels together; darker color signifies a higher ratio, and thus a higher specificity of the given term for the given Gospel.

Quite unsurprisingly, we see that the most frequent term in Jesus' sayings in the Gospel of John is the word “father” (πατήρ), while the word “man” (ἄνθρωπος), which is so predominant in the Synoptics, is almost missing. Overall, the figure appears to capture the same general pattern as in the case of distances: Looking at the 20 most frequent words, Jesus sayings' in the Gospel of John and the Gospel of Thomas do not seem to be very much related. Moreover, the table of normalized frequency of words seems to reflect the differences in theology, especially soteriology, between these two Gospels. It has been suggested that, for Thomas, the most important message is that everyone who seeks can find God, that every man has a divine capacity to know God and thus to become like Jesus. It appears that key terms for this message – “man” (ἄνθρωπος), “become” (γίνωσκω), “find” (εὑρίσκω), “know” (γινώσκω) – are also among the most common in the Gospel of Thomas. On the other hand, the main message of the Gospel of John is that God, the divine Father, because he loved us, came to this corrupt world and gave his Son to save us. In John, you cannot only seek and find; you do not have this kind of capacity within yourself, nor do you recognize the Father from the world. You must believe in Jesus, whom the Father sent into this world, know his word, and obey it. As in the previous case, the keywords of John's theology seem to be reflected in our analysis – we see that the frequency distribution of words (*gosTF*) has shown the importance of terms like “father” (πατήρ), “come” (ἔρχομαι), “world” (κόσμος), “believe” (πιστεύω), “give” (δίδωμι), “son” (υἱός), “know” (γινώσκω), “love” (ἀγαπῶ), and “send” (πέμπω). The same analysis was also made for Jesus' sayings in the Synoptic Gospels to test whether our results were accidental or not. Although these three Gospels do not translate their theological concepts into Jesus' words as strongly as John and Thomas, we can also note here that some of the characteristic features and the typical theological vocabulary of

the Gospels' authors are manifested here. An eloquent example is the importance of the term "heaven" (οὐρανός) in the Gospel of Matthew. At the same time, the similarities between the 20 most frequent terms of these three Gospels make it very clear that they are indeed Synoptic Gospels.

### 3.4 Analysis 3: word co-occurrences

#### 3.4.1 Methods

The two analyses introduced above managed to successfully capture similarities and differences in adopted vocabulary. However, these methods are completely blind to the semantic content of the words. Nevertheless, there is a whole family of QTA methods attempting to surpass this limitation. These methods form the core of distributional semantics models of meaning.<sup>68</sup>

Distributional semantics draws upon a general assumption that two words that tend to co-occur in the same contexts within a corpus also tend to have a similar meaning (the so-called "distributional hypothesis").<sup>69</sup> When a context is defined as a sentence, two words occurring repeatedly together in the same sentences are considered to be closer to each other in terms of semantic relatedness than two words that tend to occur in different sentences.<sup>70</sup>

Adopting this approach in our analysis, we first construct a word-by-word co-occurrence matrix with cell values corresponding to the number of instances that any two words co-occur within a sentence.<sup>71</sup> Each row or column within this square matrix might be considered a vector, informing us about a location of a given word within a multidimensional space (the number of dimensions equals the number of words). However, since many elements within such vectors are typically zeros (i.e., no sentence co-occurrence between two words within the text), two vectors cannot be directly compared (as more zeros imply higher similarity, which is misleading). Therefore, we have to apply standardized methods for reducing vector dimensionality, such as singular value decomposition.<sup>72</sup> As a result of this step, we obtain a matrix with rows corresponding to words and with columns corresponding to a vector with our preselected number of dimensions. This reduced word-vector matrix can finally be used to calculate similarities or distances between any word pair.

#### 3.4.2 Results

We have used this methodology to generate a word similarity matrix for four documents: Jesus' sayings within the Coptic version of the Gospel of Thomas, Jesus' sayings within the English version of the Gospel

<sup>68</sup> Lenci, "Distributional Models of Word Meaning," 151–71. For a more elaborated application of these methods in New Testament Studies, see Munson, *Biblical Semantics*.

<sup>69</sup> Harris. "Distributional structure," 146–62.

<sup>70</sup> The computational methods to calculate this kind of relatedness between two words are actually very similar to the ones we used in the case of calculating distances between two documents. The difference is that instead of analyzing and comparing documents as a whole, we are now calculating similarities between words within one document by comparing vectors based on the distribution of these words across sentences in this document.

<sup>71</sup> Instead of simple word counts, we employ Tf-Idf weighting here. Generally speaking, our method differs from Latent Semantic Analysis, which is actually constructed on the basis of a word-by-document co-occurrence matrix. Thus, while LSA attempts to capture similarities on the basis of word-by-sentence co-occurrence vectors, we are interested here in more basic patterns on the level of word-by-word co-occurrences vectors. Furthermore, LSA is designed to work with much larger corpora than the ones we are dealing with here. See Landauer, McNamara, Dennis, and Kintsch, *Handbook of Latent Semantic Analysis*; Altszyler, Sigman, Ribeiro, and Slezak, "Comparative Study of LSA vs Word2vec," <https://doi.org/10.1016/j.concog.2017.09.004>.

<sup>72</sup> For further details, see Martin and Berry, "Mathematical Foundations," 35–55.

of Thomas, Jesus' sayings within the Greek version of the Gospel of John, and Jesus' sayings within the English version of the Gospel of John.

In the historical section of this article, we introduced a set of concepts shared by the Gospel of John and the Gospel of Thomas, which have been discussed by scholars to demonstrate either similarities or differences in the theology of the two Gospels. Here we focus on the context of usage of terms which were directly identified in both Gospels using the same wordings and which could be more or less straightforwardly identified in either the Greek or Coptic version of the two Gospels.<sup>73</sup> Thus, we ended up with six terms:

“father” (πατήρ, εἰωτ), “begin/beginning” (ἀρχή, 2H/ ΑΡΧΗ/ ΑΡΧΕΙ), “find” (εὕρισκω, σῆε/ 2E), “light” (φῶς, οὐοειη), “seek” (ζητέω, ωῆε), “world” (κόσμος, κοσμος).

For each language variant of the document, we can now extract a chosen number of words with which these six target terms most often co-occur. These words might be considered to be their nearest neighbors. Since we are working with two documents in two languages, for each of the six target terms we now have two versions of their nearest neighbors (see Table 1).

**Table 1:** Target terms' nearest neighbors in Greek and Coptic

John (gr) – term and neighbors		Thom (copt) – term and neighbors	
πατήρ	κατηγορέω, ἐλπίζω, ἄπτω, τιμάω, δαιμόνιον, ἀτιμάζω, ἀδελφός, εὐχαριστέω, πωλέω	εἰωτ (father)	ΜΑΞΗ (sign), ΟΥΩΩΤ (worship), ΠΩΣΤ (bend/fall), ΕΤΜΜΑΥ (who/that), ΜΟΥΤΕ (speak/call), ΠΟΡΗ (prostitute), ΩΩΤ (trader, merchant), ΛΟΓΟΣ (word), ΜΑΛΥ (mother)
ἀρχή	δέρω, κακῶς, κακός, λαλέω, ἀπαγγέλλω, παροιμία, στήκω, ἐπιθυμία, ἀνθρωποκτόνος	2H/ ΑΡΧΗ/ ΑΡΧΕΙ (begin /beginning)	ΤΩΟΥΗ (raise), ΣΩΟΥ2 (gather, collect), ΜΗΤΕΡΟ (kingdom), †ΗΕ (loins), 2ΑΕ (last/end), ΤΣΙΟ (sate), 2ΚΟ (be hungry), ΛΙΑΚΟΗΙ (serve), 2Η (belly/womb)
εὕρισκω	παραλαμβάνω, ὅπου, τόπος, ἐρμηνεύω, σάββατον, ἐνθάδε, ἄνθρωπος, περιτομή, ἐτοιμάζω	σῆε/ 2E (find/fall)	ΛΟ (cease, stop), ΩΕ (wood), ΟΥΩΗ (be open), ΑΗΑΠΛΥΣΙC (repose), ΩΗΕ (seek/ask), ΤΚΑC (pain), ΟΒΕ+ (thirst), ΠΩ2 (tear/break)
φῶς	σκοτία, προσκόπτω, περιπατέω, φαῦλος, πράσσω, λύχνος, φαίνω, νύξ, μᾶλλον	οὐοειη (light)	ΚΑΚΕ (darkness), ΟΥΛΛ (alone, self), ΤΗΡ (all/every), ΟΥΟΗ (someone), ΩΗΑ (destroyed, be desert), 2ΙΚΩΗ (image), ΒΑΒΙΛΕ (grain), ΩΛΤΑΜ (mustard), ΣΡΩ2 (want, need)
ζητέω	ἀδικία, τεκνίον, δύναμαι, ὅπου, δόξα, κηπουρός, ὑπάγω, ποῦ, ἔτι	ωῆε (seek/ask)	ΛΟ (cease, stop), ΜΟΥΗ (remain), ΩΧΗ (cease), ΟΥΩΗ (be open), ΠΣΤΕ- (ninety), †ΨΙC (nine), 2ΗΕ- (will, desire), ΤΕΗΟΥ (now), ΧΗΟΥ (ask, question)
κόσμος	κάτω, μισέω, ὑπηρέτης, ἀγωνίζομαι, πονηρός, ἄνω, πῶτος, θαρσέω, νικάω	κοσμος (world)	ΜΗΩΛ (be worthy, worth), ΣΩΜΑ (body), ΑΡΗΑ (decline), ΤΜ- (not), ΣΑΜΒΑΤΟΗ (sabbath), ΣΑΡΞ (body), ΡΗΜΑΟ (rich man), ΟΥΛΛ (alone, self)

These results may be easily subjected to a qualitative inspection. We can identify some obvious similarities here. For instance, in both cases, the nearest neighbor of “light” (φῶς, οὐοειη) is “darkness” (σκοτία, κἀκε). At the same time, we clearly observe many differences. However, we cannot evaluate the extent of these similarities vs differences in a controlled and quantitative manner. To overcome this limitation, we have turned again to the English translations of these texts.

<sup>73</sup> Thus, for instance, we omitted words like “understanding/to understand” and “knowledge/to know,” since in the Greek original they sometimes correspond to γινώσκω, sometimes to οἶδα, etc.



When we focus only on words that appear at least twice within both documents, we find that there are 123 words which the two English-translation Gospels have in common. To make our models more comparable, we will focus especially on this shared vocabulary. For each of the target words in this shared vocabulary, we can now easily extract its relatedness to all these 123 words (including the target word itself with a value of 1, which means absolute similarity, or identity). In this manner, we obtain pairs of numerical lists of identical length, which can be inspected for correlation. Following the logic of the distributional hypothesis, the extent of the correlation between these two lists should mirror semantic similarities in usage of the target word across the two documents.

The results of this analysis are listed in Table 2. First, we see that there are a number of differences between the nearest neighbors here and in the original languages. This is not surprising considering the different structure of the languages and the very limited size of the corpus. However, we can also see that some overall patterns persist. In general, we might claim that the target terms are used quite differently within the two documents, as is evident from their nearest neighbors.

**Table 2:** Target terms' nearest neighbors and correlation

Term	John	Thom	Correlation	<i>p</i> value
Begin	Thing, witness, also, lie, truth, father, devil, desire, liar, accord	End, discover, wish, lend, interest, money, look, experience, death, place	0.16575	$p > 0.05$
Father	Know, work, one, come, love, believe, see, hear, give, send	Mother, kingdom, know, one, like, son, disciple, truly, ever, certain	0.23397	$p < 0.05$
Find	Side, right, cast, door, enter, save, Abraham, kill, place, seek	Seek, world, intoxicate, let, stone, look, ever, trouble, repose, bless	0.33914	$p < 0.001$
Light	Darkness, walk, world, deed, evil, come, may, stumble, one, become	Image, within, manifest, man, conceal, remain, put, become, come, say	0.21766	$p < 0.05$
Seek	Glory, judge, yet, go, kill, authority, tell, weep, man, Abraham	Find, let, near, worm, unable, destroy, treasure, alive, come, die	0.32701	$p < 0.001$
World	Even, come, may, hat, send, give, light, father, judgment, ruler	Find, ever, body, fast, cast, upon, superior, flesh, corpse, renounce	0.36347	$p < 0.001$

Furthermore, we also see that in the case of some words, such as “light” or “world,” the correlation is quite strong, while in the case of other words, like “begin” or “seek,” it is much weaker. A stronger correlation suggests that the usage of the term is more similar across the two documents. However, we should use discretion with these findings, as the reliability of these models is highly dependent on the size of the input language data. In our case, this size is very small.

With this in mind, we can still see that by using these technological tools we are able to discern that the same vocabulary and the same metaphors are used to convey semantically different messages. Despite the similarities and parallels shared by the two Gospels, they seem to be quite far from each other according to this analysis. By focusing especially on terms that scholars consider proof of the similarity of the two documents, we were able to capture quantitatively that these terms are used very differently here. This analysis therefore seems more to support the view shared by scholars who focus on the differences and assume that the same vocabulary is used to present different attitudes in these Gospels, as we can see that key common terms in both Gospels are associated with a very different set of expressions, which is shown in the two tables above.

## 4 Conclusion

In summary, all of the quantitative text analyses we employed in the previous section seem to support the second view in the scholarly debate, i.e., the view that emphasizes differences between the two Gospels and considers these differences to be more fundamental than the similarities visible at first sight. However,

it is not possible to decide based on these analyses whether these Gospels are in direct conflict with each other or whether they are different because – despite the use of similar terminology – they represent different theological concepts. With these results in mind, it appears that at least in our particular case it is meaningful and useful to complement traditional close reading, i.e., exegetical interpretation, with methods of “distant reading,” and that even very basic methods of this type can provide relevant insights. It was also very important to find that these tools have the ability to capture semantic differences when the same terminology is used to transmit different meanings.

The essential benefits therefore include the acquisition of a global view of the content and structure of the examined texts or corpora; furthermore, we can also compare textual corpora with one another. In a controllable and replicable form, it is possible to support and complement the researcher’s view obtained via a detailed close reading of individual texts. A very important benefit we see in these methods is the possibility to correct pre-understandings and subjective views, which is difficult to avoid in close reading and classical exegesis. Especially researchers who have been dealing with certain texts and topics for a long time and know them very well already have a wealth of knowledge and information that naturally shapes and influences their understanding of related problems and topics. Computer analysis can thus enable them to look at the problem without all of these burdens, to evaluate their hypothesis and interpretation in a replicable form, or to warn them in cases in which they have overlooked some essential aspects or overloaded an element at the expense of another.

Naturally, there are also indisputable limitations to this type of research, some of which have been revealed by this study. Linguistic questions have arisen, such as how texts preserved in different languages can be compared and whether the use of their translations into English is acceptable and meaningful or not. By comparing the analyses performed for the canonical Gospels in the English, Greek, and Coptic versions, we have attempted to show that the use of translations did not impair the comparability of these texts to such an extent that the results would not be relevant. Similarly, the question arises as to whether it is possible to compare texts differing in genre (narrative vs collections of sayings). In this case, we attempted to balance this difference in the genre of the texts by selecting only the statements of Jesus preserved in direct speech from all the Gospels. Since the author of the Gospel of John lets his theology speak to the reader through the words of Jesus, we believe that in this case the analysis was meaningful and justified. However, if we wanted to compare the theology of the Gospel of Thomas and the Synoptic Gospels solely on the basis of an analysis of Jesus’ sayings contained therein, it would not be possible and would not lead to relevant results, as Jesus’ statements in the Synoptics are not significant bearers of their theological concepts.

In conclusion, an important issue that we also pointed out concerns the informative value of comparing smaller text corpora, as, from a statistical point of view, the comparison of a larger volume of data naturally has much greater evidential value. On the other hand, the advantage of analyzing smaller text corpora lies in the controllability and comparability of the results of these quantitative analyses with the results obtained by close reading. This is the procedure we have chosen for this study, and it cannot be applied to cases in which the corpus analyzed is too large to be covered by close reading – in such a case, it is necessary to rely solely on the results of these quantitative analyses without comparing them with the results of qualitative methods. To summarize, we can conclude that despite the aforementioned limitations, we can see that these quantitative text analyses can provide a relevant and useful complement to traditional close reading and to correct or control it in a replicable fashion.

## References

- Altszyler, Edgar, Sigman, Mariano, Ribeiro, Sidarta, and Slezak Diego Fernández. “Comparative Study of LSA vs. Word2vec Embeddings in Small Corpora: A Case Study in Dreams Database.” *Consciousness and Cognition* 56 (2017), 178–87. doi: 10.1016/j.concog.2017.09.004
- Bauckham, Richard. “For Whom Were Gospels Written?” In *The Gospels for All Christians: Rethinking the Gospel Audiences*, edited by Richard Bauckham, 9–48. Grand Rapids: Eerdmans, 1998.

- Bird, Steven, Klein, Ewan, and Loper, Edward. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. Sebastopol: O'Reilly Media, Inc., 2009.
- Brown, Raymond. "The Gospel of Thomas and St. John's Gospel." *New Testament Studies* 9:2 (1962–1963), 155–177.
- Davies, Stevan L. "The Christology and Protology of the Gospel of Thomas." *Journal of Biblical Literature* 111:4 (1992), 663–82.
- Davies, Stevan L. *The Gospel of Thomas and Christian Wisdom*. New York: The Seabury Press, 1983.
- DeConick, April D. *Voices of the Mystics: Early Christian Discourse in the Gospel of John and Thomas and Other Ancient Christian Literature*. Sheffield: Sheffield Academic Press, 2001.
- Dunderberg, Ismo. "John and Thomas in Conflict?" In *The Nag Hammadi Library after Fifty Years: Proceedings of the 1995 Society of Biblical Literature Commemoration*, edited by John D. Turner & Anne McGuire, 361–80. Leiden: E. J. Brill, 1997.
- Dunderberg, Ismo. "Thomas' I-Sayings and the Gospel of John." In *Thomas at the Crossroads: Essays on the Gospel of Thomas*, edited by Risto Uro, 33–64. Edinburgh: T&T Clark, 1998.
- Dunderberg, Ismo. *The Beloved Disciple in Conflict? Revisiting the Gospels of John and Thomas*. New York: Oxford University Press, 2006.
- Fieger, Michael. *Das Thomasevangelium: Einleitung, Kommentar und Systematik*. Münster: Aschendorff, 1991.
- Harris, Zellig S. "Distributional Structure." *Word*, 10:2–3 (1954), 146–62.
- Jurafsky, Daniel & Martin, James H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Harlow: Pearson, 2017.
- Klink, Edward W. *The Sheep of the Fold: The Audience and Origin of the Gospel of John*. Cambridge: Cambridge University Press, 2007.
- Koester, Helmut. "Gnostic Writings as Witnesses for the Development of the Sayings Tradition." In *The Rediscovery of Gnosticism, vol 1*, edited by Bentley Layton, 238–61. Leiden: E. J. Brill, 1980.
- Koester, Helmut. *Ancient Christian Gospels*. Philadelphia: Trinity Press International, 1990.
- Koester, Helmut. *Introduction to the New Testament: History and Literature of Early Christianity*. New York: Walter de Gruyter, 2000.
- Landauer, Thomas K, McNamara, Danielle S., Dennis, Simon, and Kintsch, Walter (eds.). *Handbook of Latent Semantic Analysis*. New York–London: Routledge, 2007.
- Lenci, Alessandro. "Distributional Models of Word Meaning." *Annu. Rev. Linguist* 4 (2018), 151–71.
- Marcovich, Miroslav. "Textual Criticism on the Gospel of Thomas." *Journal of Theological Studies* 20 (1969), 53–74.
- Martin, Dale I. & Berry, Michael W. "Mathematical Foundations Behind Latent Semantic Analysis." In *Handbook of Latent Semantic Analysis*, edited by Thomas K. Landauer, Danielle S. McNamara, Simon Dennis, and Walter Kintsch, 35–55. New York–London: Routledge, 2007.
- Mirkovic, Alexander. *Johannine Sayings in the Gospel of Thomas: The Sayings Traditions in their Environment of First Century Syria*. [https://www.academia.edu/2775436/Johannine\\_Sayings\\_in\\_the\\_Gospel\\_of\\_Thomas\\_The\\_Sayings\\_Traditions\\_in\\_their\\_Environment\\_of\\_First\\_Century\\_Syria](https://www.academia.edu/2775436/Johannine_Sayings_in_the_Gospel_of_Thomas_The_Sayings_Traditions_in_their_Environment_of_First_Century_Syria)
- Moretti, Franco. *Distant Reading*. London–New York: Verso, 2013.
- Munson, Matthew. *Biblical Semantics: Applying Digital Methods for Semantic Information Extraction to Current Problems in New Testament Studies*. Aachen: Shaker Verlag, 2017.
- Pagels, Elaine. "Exegesis of Genesis 1 in the Gospels of Thomas and John." *Journal of Biblical Literature* 118:3 (1999), 477–96.
- Pagels, Elaine. *Beyond Belief: The Secret Gospel of Thomas*. New York: Random House, 2003.
- Patterson, Stephen J. *The Gospel of Thomas and Jesus*. Sonoma: Polebridge Press, 1993.
- Perrin, Nicholas & Skinner, Christopher W. "Recent Trends in Gospel of Thomas Research (1989–2011). Part II: Genre, Theology and Relationship to the Gospel of John." *Currents in Biblical Research* 11:1 (2012), 65–86.
- Popkes, Enno Edzard. "Ich bin das Licht – Erwägung zur Verhältnisbestimmung des Thomasevangeliums und der johanneischen Schriften anhand der Lichtmetaphorik." In *Kontext des Johannesevangeliums: Das vierte Evangelium in religions und traditionsgeschichtlicher Perspektive*, edited by Jörg Frey & Udo Schnelle, 641–74. Tübingen: Mohr Siebeck, 2004.
- Quispel, Gilles. "Qumran, John and Jewish Christianity." In *John and the Dead Sea Scrolls*, edited by James Charlesworth, 137–55. New York: Crossroad, 1991.
- Riley, Gregory J. *Resurrection Reconsidered: Thomas and John in Controversy*. Minneapolis: Fortress Press, 1995.
- Robinson, James. "The Johannine Trajectory." In *Trajectories through Early Christianity*, edited by Helmut Koester & James Robinson, 232–68. Philadelphia: Fortress Press, 1971.
- Robinson, James (ed.). *The Nag Hammadi Library in English*. Leiden: E. J. Brill, 1996.
- Sell, Jesse. "Johannine Traditions in Logion 61 of the Gospel of Thomas." *Perspectives in Religious Studies* 7 (1980), 24–37.
- Skinner, Christopher W. *John and Thomas – Gospels in Conflict? Johannine Characterization and the Thomas Question*. Eugene: Pickwick Publications, 2009.
- Slingerland, Edward, Nichols, Ryan, Neilbo, Kristoffer, and Logan, Carson. "The Distant Reading of Religious Texts: A "Big Data" Approach to Mind-Body Concepts in Early China (Preprint)." *Journal of the American Academy of Religion* 85:4 (2017), 985–1016.
- Underwood, Ted. "A Genealogy of Distant Reading." *Digital Humanities Quarterly* 11:2 (2017), 1–12.
- Zöckler, Thomas. *Jesu Lehren im Thomasevangelium*. Leiden: E. J. Brill, 1999.