religions

MDPI

Article

Old Wine in New Wineskins: Applying Computational Methods in New Testament Hermeneutics

Christian Houth Vrangbæk [1,*], Eva Elisabeth Houth Vrangbæk [2] and Jacob Mortensen [1]

1 School of Culture and Society, Research Unit for New Testament Studies, Aarhus University, 8000 Aarhus, Denmark; teojmo@cas.au.dk
2 School of Culture and Society Church History and Practical Theology, 8000 Aarhus, Denmark; eehv@cas.au.dk
* Correspondence: chv@cas.au.dk

Abstract: New Testament studies has over the past years seen an increase in the use of digital methods, but some of the more advanced methods still lack proper integration. This article explores some of the advantages and disadvantages in employing computational/algorithmic approaches, such as so-called semantic models of word embedding and topic modelling analysis. The article is structured into three main parts. The first part (1) introduces the reader to the field of computational studies in literary, historical, and religious research areas and outlines the computational methods, namely topic modelling and word embedding. The second part, (2) showcases two computational tools in analyzing New Testament narratives. The third part (3) discusses and compares how the methodology of applying computational techniques can maintain and advance a focus on the historical and literary context of New Testament texts. The specific problem the article addresses is how computational methods can be wielded and not sacrifice the contact to the text and the historical context. We argue that applying computational methods in New Testament hermeneutics necessarily involves methodological pros and cons. These computationally assisted analyses can be regarded as old wine in new wineskins—classic, hermeneutical questions can be posed with new methods.

Keywords: New Testament; topic modelling; word embedding; hermeneutics; literary–historical context; computational methods

check for updates

Academic Editors: Susanne Luther, Birke Siggelkow-Berner and Clarissa Breu

Received: 6 November 2024
Revised: 21 December 2024
Accepted: 23 December 2024
Published: 31 December 2024

Citation: Vrangbæk, Christian Houth, Eva Elisabeth Houth Vrangbæk, and Jacob Mortensen. 2025. Old Wine in New Wineskins: Applying Computational Methods in New Testament Hermeneutics. *Religions* 16: 28. https://doi.org/10.3390/rel16010028

1. Introduction

In this article, we wish—within the scope of the special issue on current and new approaches to New Testament studies—to outline, demonstrate, and discuss the value of how key New Testament words/concepts can be analyzed by employing the computational/algorithmic methods of topic modelling and word embedding. Our focus is on outlining, exemplifying, and applying the methods in a practical and accessible way. We want to demonstrate how a computational/algorithmic approach is a kind of "conceptual history" that can maintain a close contact with the source text and the literary–historical context.[1] We focus our analysis on New Testament narratives (Gospels, Acts, and Revelation) and compare the results to LXX historiographies.[2] We wish to showcase how computational methods in textual analysis do not minimize the importance of the text nor the historical and literary context. We argue that computationally assisted analyses can be regarded as old wine in new wineskins—classic, hermeneutical questions can be posed with new methods.

1.1. Computational Hermeneutics Past and Present

The use of digital and computational research methods has grown significantly across many disciplines, especially over recent years.[3] Computational methods were initially adopted by researchers to revolutionize the natural sciences, especially biology and physics. Subsequently, these techniques found their way into the social sciences, and, now, they have established their own niche in the humanities (Armand and Henriot 2023; Špiclová and Kaše 2020; Vrangbæk and Nielbo 2021; Vrangbæk 2024).

In New Testament studies, we have been using concordances extensively. Concordances, in likeness with computational methods, break down the textual structure and allows the reader to find information from the text in a decomposed form. Concordances are one-dimensional versions of what can be accomplished with basic computational methods. An often-heard criticism of computational methods is that the studies are conducted by non-experts or that the analyses are too distant from the text. But, if the hermeneutical approach is initiated by biblical scholars, as is the case with concordances, then computational methods will not be foreign to the investigation.

In the humanities, there has been vigorous debate regarding the adoption of computational analytical methods (Moretti 2000b; Underwood 2019; Pannapacker 2012; Nielbo et al. 2017; Brennan 2017; Conrad 2014; Da 2019a, 2019b; Vrangbæk and Nielbo 2021).[4] Proponents highlight the new opportunities offered, whereas critics insist on improved integration and demonstrate distrust or skepticism toward the results of computational methods. A significant aspect of the debate concerns the methodological basis of so-called "distant reading" (Moretti 2000a, p. 55; Jockers 2013, p. 31). Distant reading employs a largely quantitative, nomothetic approach, often rooted in a strong critique of traditional manual methods. In contrast, classic humanistic research is predominantly regarded as being idiographic, qualitative, and involving "close-readings". However, for us, this dichotomy appears false in practice because both modes are intricately intertwined. By integrating computational methods with traditional approaches, we demonstrate that, as theological, text-focused researchers, we engage in both quantitative and qualitative work, as well as idiographic and nomothetic analysis. We use quantitative and nomothetic techniques to uncover textual structures, while our interpretations are qualitative, idiographic, and close to the text, which has been described in similar studies (Jensen 2014, pp. 115–34; Porsdam 2013; Liu 2016; cf. Vrangbæk and Nielbo 2021, p. 151). We engage with the formal rigor of mathematics and computation and bring the results to a hermeneutic practice (Nielbo et al. 2019b). Our current study aims to embed computational analyses within the Biblical disciplines rather than treating them as external methodologies (Nielbo et al. 2017; Bizzoni and Lappin 2018).[5] The approach we show in this demonstration is meant as a supplementary approach to textual analysis and close readings. It is by no means meant to outmaneuver the traditional, qualitative, and hermeneutical approaches.[6] It is a supplement which can provide new interpretations, ideas, and perspectives.

1.2. Methods

The approach to this study is partly situated in computational studies and partly in conceptual history as has been conducted in recent studies (Wevers and Koolen 2020; Baunvig and Nielbo 2022). We understand conceptual history less as an analysis of the historical development of a word and more as the study of the use and dispersal of specific words in a given context (Ifversen 2007, pp. 81–82). The approach is "conceptual" in the sense that we focus on calculating the importance of first-order terms, which through computational methods will become second-order concepts. When we analyze texts with computational methods we can transform first-order words into second-order concepts and, thus, be able to map focal points in the New Testament "conceptual landscape" (Baunvig

2023; Vrangbæk 2024). A key component of conceptual history is the inclusion of historical and literary context. It is important to note that we work with context on two levels: In word embedding, the "meaning" of the words is calculated from the contextual position, i.e., the individual words gain values based on their position and co-occurrence with other terms in the text (cf. Section 1.2.2 below). In topic modelling, the "meaning" of words and topics are determined based on the contextual corpus from which the topics are calculated. Both word-embedding and topic modelling retrieves information and produces patterns only based on context, but context on different levels.

According to the father of conceptual history, Reinhardt Koselleck, a word has a wide range of different contextual meanings (Ifversen 2007, p. 89; Koselleck 1972). Whereas our approach does not survey a historical development, we map the contextual structure in a closed system. That is, the contextual approach of word embedding and topic modelling assists in analyzing the semantic structure of individual words in the target text, which for us is the New Testament texts. We find that the combination of topic modelling and word embedding can support the conceptual analysis of New Testament texts and can assist in analyzing the texts traditionally (Willkomm et al. 2018, p. 184; Baunvig and Nielbo 2022; Wevers and Koolen 2020).

1.2.1. General Methodological Steps in Computational Analysis

In employing computational analysis methods, certain fundamental methodological steps are involved, yet the process is far from being an automatic pipeline. Each step involves numerous qualitative and hermeneutical decisions with a range of interpretational pros and cons. Figure 1 presents a simplified overview of the basic steps.[7]

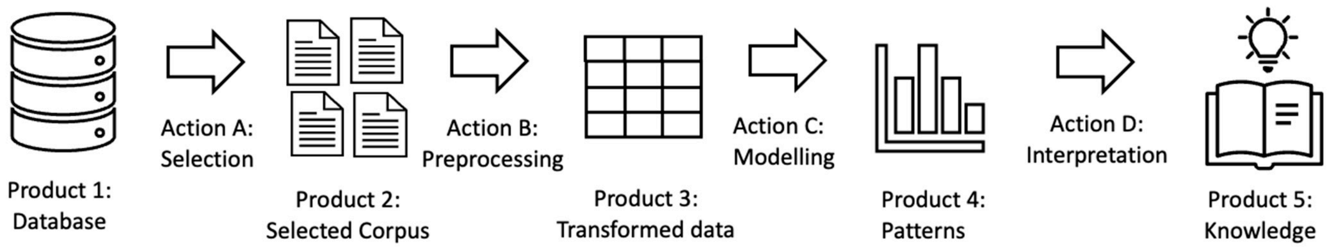

Figure 1. The process of working with computational methods.

The "arrows" in the figure symbolize decisions that must be made about how to handle the provisional "products" to achieve the final outcome of "knowledge" (Product 5). Each action involves multiple sub-actions and hermeneutical choices. As traditional close-reading scholars, we usually work with Action A and D, which pertain to data/source selection and interpretation, respectively. However, computational methods introduce two additional steps: Action B involves preprocessing and managing the Ancient Greek text in a digital form, and Action C involves computational modelling of the (Greek) digital text. Modelling can include all the various "computational methods". In our case, topic modelling and word embedding have been applied. Other models, such as clustering or decision trees, could also be used. Modelling produces patterns (Product 4), which we could call the "computational results". Preprocessing and modelling moves the interpretation of the text from first-order observations to second-order conceptualization, since the modelling presents new patterns and connections in the texts (transformed data), and it is due to this circumstance that we can speak of conceptual history, since we conceptualize first-order terms into second-order patterns produced by the chosen modelling-technique. The preprocessing and modelling are crucial parts of the "analysis", since hermeneutical preconditions are applied here, which are preconditions that need to be declared. The results and patterns of the modelling have to be interpreted (Action D) by a human interpreter in order to reach new knowledge (Product 5) (cf. Vrangbæk 2024).

1.2.2. Word Embedding

When the method of word embedding was introduced in 2013, it was claimed that previous techniques tended to "treat words as atomic units". Instead, the authors pleaded, we should consider context (Mikolov et al. 2013a; Boleda and Erk 2015). Each word does not derive its meaning from lexical meaning. Rather it gains contextual meaning based on the neighboring words and context windows. The "embedding" refers to the process of representing words as dense vectors in a semantic vector space (Jurafsky and Martin 2024; Li and Yang 2018; Mandikal and Mooney 2024; Goldberg 2017, 90ff).[8]

When a word embedding analysis is run, each word and its contextual embeddings are stored in a multi-dimensional vector matrix. Put simply, this means that all words are assigned numerical values and the words that have similar values are contextually similar to each other (Baunvig and Nielbo 2022, pp. 24–25). Word embedding is comparable to a collocation analysis, where collocated words are analyzed because they are next to (or close to) each other. However, word embedding investigates collocations within a set contextual window of, for instance, 100–300 words. This allows the scholar to discover semantic relations beyond the immediate collocated context and it will, thus, reveal more complex and "hidden" collocations.

The use of word embedding presupposes a theoretical view of language known as "distributional semantics". Distributional semantics is often explained by summarizing the phrase that a word is "characterised by the company it keeps" (Firth 1957, p. 11; Harris 1954, pp. 146–62; Boleda 2020). In this scheme, the meaning of words does not derive from their lexical meaning. Rather, their meaning pertains to which words they co-occur together with in the trained corpus. In distributional semantics, the meaning of words is represented across a multidimensional space. This involves understanding meanings through the "geometric relationships between words", which refers to their relative positions in a vector space. Essentially, meaning is situated within a network or system of semantic relationships (Boleda 2020, pp. 215–16). The vector space, or semantic vector space, enables the exploration of a concept's semantic landscape, including its associations and neighboring words. This vector space represents a decomposition of the original text, while preserving all its elements.[9]

The following Figure 2 is a hypothetical example of what a geometrical interpretation of distributional semantics may look like.

This plot illustrates the relationships among a random set of terms. Notably, fruits naturally cluster together, vehicles form another group, and terms related to humans are closer to each other compared to those referring to objects. Additionally, the relationship between "king" and "queen" mirrors that between "man" and "woman". These concept pairs, consisting of related terms and/or opposites, align with the binary opposites concept of Saussure and Levi-Strauss. Each pair complements rather than excludes the other, enhancing their meaning. We believe this is an important consideration when conducting semantic conceptual analysis using computational methods.[10]

The output of a word embedding analysis needs to be evaluated, and, further, it needs to be interpreted and integrated into the broader analysis. If we consider Figure 1 from Section 1.2.1, word embedding constitutes Action C, "modelling". The word embedding analysis produces outputs in the form of patterns or conceptual maps, which we as interpreters must relate to and interpret to create new knowledge (Product 5). At this point, the analysis is comparable to traditional methods, since we have to support interpretations with the sources and the literature (Nielbo et al. 2017; Vrangbæk 2024). There is no shortcut to more true or real interpretations of texts. Actual close reading interpretation (continuously in dialogue with secondary literature) still has to be performed. The word embedding model presents the decomposed target-text in a new way which opens up new possibilities for interpretation.

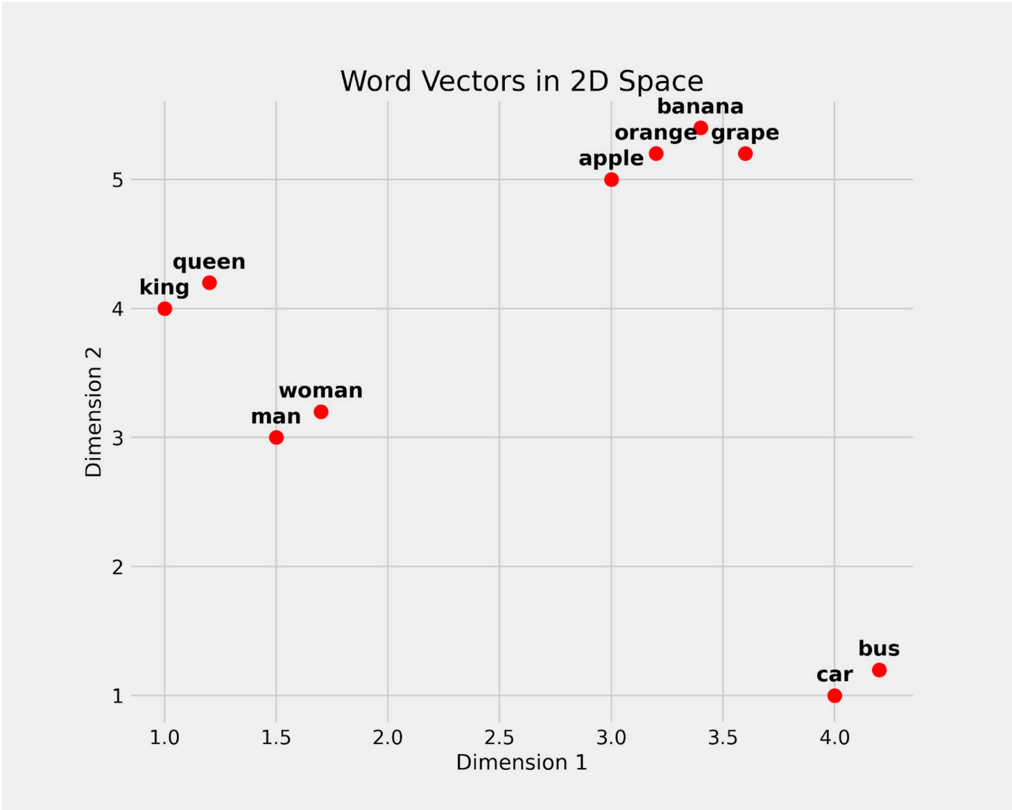

Figure 2. Hypothetical visualization of a word embedding analysis.

1.2.3. Topic Modelling

Topic modelling is generally used to classify documents within a large unstructured corpus of texts (Kherwa and Bansal 2019; Gillings and Hardie 2023; Uglanova and Gius 2020). For example, imagine a corpus of 1 million news articles that you want to organize the articles into the topics of "Politics", "Business", "Lifestyle", and "Sports". Topic modelling will assist in assigning such topical features to the documents, since the algorithm will uncover the topical structure of each document based on the actual words appearing in each article. Topic modelling can be applied to any corpus of texts. In a study on poetry, topic modelling was found to be able to uncover "rich deposits of hermeneutic possibility" (Rhody 2012, p. 33).

When we employ topic modelling, we operate under the assumption that our corpus of texts is made up of various topics, with these topics being composed of the words within the corpus. We operate with three textual levels: word–document–corpus. A "topic" is connected to all three layers: it is a group of words that appear with more or less likelihood in the documents (texts) across the corpus. The process can be described in a simplified form as the following: One can imagine the whole corpus as a bag of words. Then, the algorithm, in the first instance, randomly assigns twenty random words to each document. Since this first instance is based on a random assignment—e.g., a word group (topic) containing "ἰησοῦς" and "Ἰσαάκ" could randomly be assigned to the *Iliad* (!)—there is obviously a low likelihood of these words appearing in the *Iliad*, so the algorithm reassigns words until it finds the most probable match between word groups (topics) and documents, and iterates this process multiple times until there is a probable fit between topics and documents.

Each document (text) might end up being assigned to multiple topics. The composition of multiple topics that appear in a document is called the "topical distribution".[11] When a document is relatively dominated by one topic, it shows that this particular document with high probability consists of the words in the topic. If we take the Book of Jeremiah as an example, and run an experiment as we do below, it shows that this text will be 95% dominated by a theological topic (as defined below), whereas, if we take Philo's works, these are more ambiguous in their topical distribution, displaying tensions between Hellenistic philosophy and theology (topics defined below).

In this article, we employ a specific type of topic modelling known as non-negative matrix factorization (NMF) (Sherstinova et al. 2020; Lee and Seung 1999). NMF is a method that deconstructs a high-dimensional term–document matrix, where the matrices represent documents as columns and words as rows, with the cell entries reflecting the weighted frequency of word occurrences. These occurrence values are set as the term frequency-inverted document frequency (TF-IDF).[12] NMF identifies topical patterns by connecting words to topics and topics to documents based on "optimisation". Essentially, the NMF model uncovers latent topical patterns by clustering words that frequently co-occur, and which appear in the assigned texts within the corpus, so that, eventually, topics consisting of words, like θεὸς and υἱὸς, will end up being assigned to texts in which they appear with a higher probability than others (Kherwa and Bansal 2019).

It is important to note that topic modelling is a so-called "unsupervised" method. This means that the outcome is not controlled except by the input texts and the set parameters. The results to be interpreted are only based on algorithmic calculation. For the results of the topic modelling analysis, see the following Section 2.1.

2. Topical and Semantic Structures in New Testament Texts

2.1. Finding the Topical Structure of New Testament Texts

In our first experiment, we will employ the method of topic modelling, and execute an NMF analysis of our corpus of 2153 documents.[13] The texts have been preprocessed so all words are lemmatized, and so-called stopwords, i.e., most of conjunctions, prepositions, and particles have been removed (e.g., καὶ, μέν, δέ, ὁ, οὗτος, and ἀλλά).[14] The documents in the corpus are a collection of Classical Greek, Hellenistic Greek, New Testament, Septuagint, and Pseudepigraphal writings. When using topic modelling, it is important to choose a specific number of topics. If the requested number of topics is too large, one risks overfitting, so it is necessary through trial and error to find a fitting number of topics which the corpus is able to uphold. For example, a different corpus with millions of documents might be capable of holding 50 topics, whereas a corpus of our size is usually capable of 10–20 topics. In our case, we landed on 10 topics with 20 words each, which created the best interpretable topics (Vrangbæk et al. 2024).[15] The results of the topic modelling are shown in Table 1.

The resulting topics are in themselves a list of words, so it is up to the informed interpreter to assign a label to them. We thus interpret topic 0, which contains such words as πόλις, πὸλεμος, βασιλεύς, ἕλλην, and χώρα, to be called "Greek History", while topic 1, which contains θεός, ψυχή, λόγος, and χριστός, is called "Theology 1". Topic 2, which contains θεός, κυριός, λαός, ἄνθρωπος, is called "Theology 2". The difference between topic 1/theology 1 and topic 2/theology 2 could also be interpreted to be differences in the LXX texts and the New Testament texts, or, between "Judaism" and "Christianity" with the most important words of topic 1 being θεὸς and χριστός, which incline toward "Christianity/New Testament", while the most important word in topic 2 is κύριος, together with such terms as λαός, βασιλεύς, γῆ, υἱός, and πατήρ, which incline toward "Judaism/LXX". Topic 2 also has χριστός, but this word is much lower ranked than in topic 1. We assigned our labels based on the words in the topic as well as the texts that have a high percentage of the particular topic. Based on this principle and based on the observation that almost all New Testament texts have a high percentage of topic 2 (the topic inclining toward Judaism), we decided to call them, respectively, theology 1 and theology 2. This principle is also valid for how the other topics were labelled, based partly on the words and partly on which texts display a high percentage of the particular topic. In Figure 3, we show a full list of the words of topics 1 and 2 together with bar charts showing their importance for the topic (dark red) combined with their overall importance in the corpus (light red).

Table 1. Overview of resulting topics.

No.	Top Topic Words (Calculated and Chosen)	Translated Words	Label (Interpreted)	Authors with High Distribution of Topic (Observed)
0	πόλις, πὸλεμος, βασιλεύς, ἕλλην, χώρα	city, war, ruler, Hellene, land	Greek History	Thucydides
1	θεός, ψυχή, οὐρανός, λόγος, χριστός	god, soul, heaven, word, Christ	Theology 1	Galatians, Eusebius, Origen
2	θεός, κύριος, λαός, ἄνθρωπος	god, lord, people, human being	Theology 2	Jeremiah, Judges, Chronicles
3	λόγος, φύσις, ἀρετή, ἄνθρωπος	reason, nature, virtue, human being	Greek Philosophy	Plato's *Phaedrus* and *Politeia*
4	ζεύς, παῖς, ἔρως, κύπρις, ἀπόλλων	Zeus, child, love/Eros, Cypris, Apollo	Greek culture and religion	Homer, Hesiod
5	μόριον, τρόπος, οὐσία, ἁνάγκη, αἰτία	part, character, being, necessity, cause	Medicine and ontology	Galen
6	σῶμα, ὕλη, ἀήρ, γή, ὕδωρ, πῦρ.	body, matter, air, earth, water, fire	Element Philosophy	Aristotle's *Problemata*, Plato's *Timaeus*
7	λόγος, πόλις, νόμος, τύχη, χάρις	reason, city, law, fortune, favor	Rhetoric and Philosophy	Libanius, Philo, Josephus
8	κύκλος, γωνία, κέντρον, σημεῖον, ἐπιφάνεια	circle, angle, center, sign, appearance	Geometry	Scholia in Euclidem

It is important to note that all documents in the 2153-text corpus are assigned to multiple topics based on "optimisation" (Sherstinova et al. 2020). All documents in the corpus will end up with a topical distribution which reflects its content. The more unified the distribution, the clearer the expression of a particular document, and the more diverse topical distribution, the more ambiguous the assignment of one topic. Some texts in the corpus, e.g., the Book of Jeremiah, have a relatively unified distribution with almost 95% of topic 2. Other texts, like the Gospel of John, have two large portions of topics 1 and 2, together with a mix of topics 9, 3, 6, and 8 (law and justice, Greek philosophy, element philosophy, and geometry, respectively). Illustrative examples of topical distributions are, e.g., the Homeric Hymns, which are dominated by topic 1 and topic 4 (theology 1 and Greek religion), and IV Maccabees, which has topic 7 (rhetoric/philosophy) as its top topic, followed by a large portion of topic 1 (theology 1).

Figure 3. Word list comprising topic 1 and topic 2.

In the following paragraphs, we will go further and dive into the topical distribution of New Testament narratives (Gospels, Acts, and Revelation).[16] In our experiment, the algorithm has assigned a different topical distribution to all texts based on the resulting topics in Table 1. New Testament narratives are clearly dominated by topics 1 and 2, with topic 2 being the most dominant. The topical distribution is visualized in Figure 4.

Even without going into much detail, we can discern that Luke and Matthew are similar in terms of percentages, with both having the largest portion of topic 2, the topic which dominates all of these documents. It is noteworthy that the word "χριστός" only appears in topic 1, which figures as the second largest topic in all of the texts. Mark, together with John, has a larger portion of topic 1. In our interpretation, topic 2 appears to be a slightly more Jewish topic than topic 1, with λαός and κύριος as high ranking words. The topic of element philosophy (topic 6) is the third highest ranked in Mark, Matthew, and Revelation, while topic 9 is the third highest ranked in Luke, Acts, and John has topic 9, which with our label can be called the justice and law topic. Luke, Matthew, and Mark each contain eight topics, whereas Acts and John contain seven, and Revelation contains six topics. The presence of topic 4 (Greek culture and religion) in Mark and Revelation is also noteworthy in comparison to its minor prominence in Luke and its absence in Acts. This might be explained with the corresponding presence of topic 0, in Luke, Acts, and Revelation, namely the topic of Greek history with prominent words, like πόλις, πόλεμος, χώρα, and βασιλεύς.

Our first experiment with topic modelling has served the purpose of categorizing and navigating our texts in the large corpus at a distance. In Section 3, we will return to topic modelling and show how this method can also assist in engaging the literary and historical context. So far, we have made an important first step by extracting a rough pattern of the topical and, therefore, conceptual structure of NT narratives.[17] The following experiment in Section 2.2 will take the results of the analysis in experiment 1 and employ a new computational method to move further into the conceptual landscape of NT narratives and related texts.

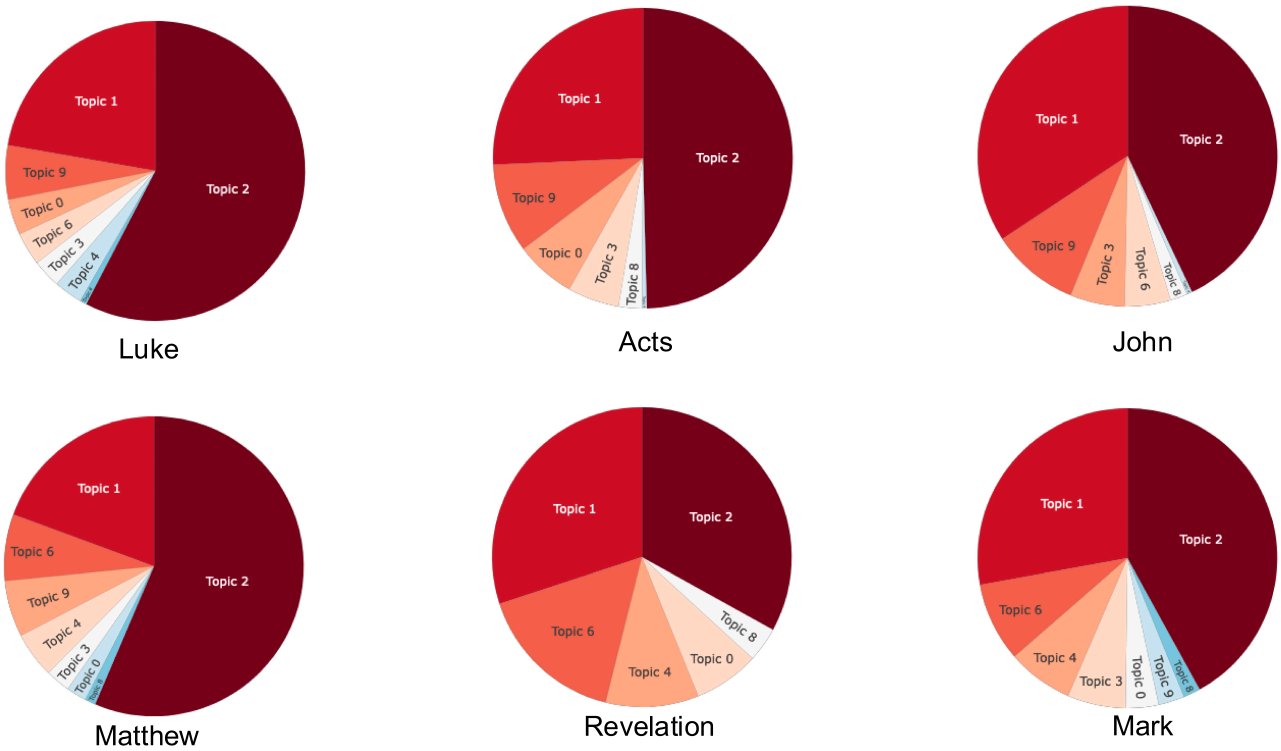

Figure 4. Visualisation of Topical Distribution of NT-narratives.

2.2. Conceptual Analysis with Word Embedding

Differently from topic modelling, word embedding enables the querying of single words and their significance in the corpus. This is a strong tool when investigating concepts and their dispersal. Similar to our topic modelling analysis, our word embedding analysis happens in two steps: the first is the computational analysis (cf. Figure 1, Action C), and the second is the interpretation of the produced results (cf. Figure 1, Action D).

When we create a word embedding analysis, we calculate contextual embeddings for every word in our corpus. Word embeddings have to have clearer demarcations than topic modelling analyses. One of the strengths of topic modelling is that this method can organize large unstructured corpora, whereas, if we were to run word embedding on the same large corpus, we could only gain information about the whole corpus and its contextual embeddings. In our previous topic modelling experiment, we could distinguish the topical distribution of each text in relation to the corpus. The textual setup in the word embedding experiment is different, since, here, we survey Gospels, Acts, and Revelation in one lump, labelled as "New Testament narratives" (NT narratives).[18] Our setup of the word embedding analysis is, thus, different from the topic modelling, since we are only analyzing NT narratives in isolation. This means that, on the one hand, we can dive deeper into the semantic landscape of the texts, but, on the other hand, we must be aware that we have grouped six texts together as one demarcated group. When we compare these with other texts, we will compare them with a new isolated group, like we do below with LXX historiographies.

In order to extract interpretable information on selected keywords, we have to query specific "seed words". Seed words are words we specifically search for; these are words which we know, as scholars and from the above experiment, are (or could be) important for our texts. In our second experiment, we will build further on the results from our first experiment, and query deeper into the top words of the topics that were found in the analysis in Section 2.1, i.e., the words which dominated NT narratives, namely θεὸς, ψυχή, λόγος, and χριστός (topic 1 words), and θεὸς, κύριος, λαός, and ἄνθρωπος (topic 2 words). The results of this query are shown in Figure 5.

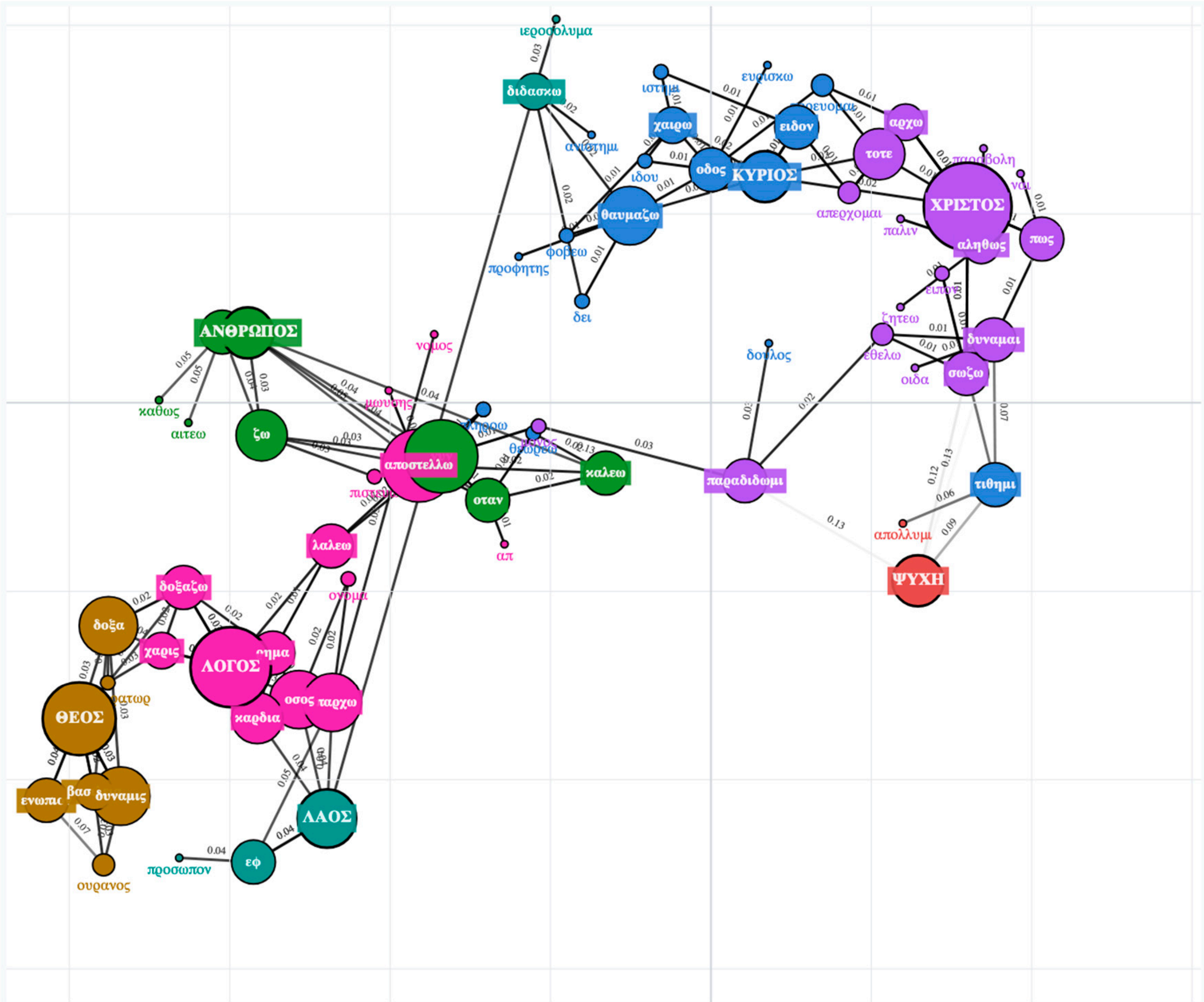

Figure 5. Semantic network with selected seed terms in the NT narratives.

What the results of the word embedding shows is a visualization of how the input terms are related to each other within the queried corpus (NT narratives).[19] This means that the contextual embeddings of all input terms across Mark, Matthew, Luke, John, Acts, and Revelation are visualized. The seed terms are written with capital letters. The other terms that spawn are, respectively, first-level associated words (words in lower case which are circled) and second-level associated words (words in lower case with no circles). When there is an overlap, for example, χάρις has a first-level association to λόγος, while it has a second-level association to θεὸς, χάρις is shown to be circled. First-level associative words are words that co-occur with the seed words, and second-level-associations are first-level associations of the first-level associations. For example, our seed word is θεὸς, and one first-level association is δύναμις, while the second-level association is οὐρανός. If δύναμις was our seed word, then οὐρανός would have been the first-level association of δύναμις. The associations are created based on the "training", i.e., calculations of vector similarity between terms in the training corpus. The words that appear close in our visualization are words that share similar contextual values between their vectors (Ethayarajh et al. 2019). When we use the word "co-occur", it is important to note that this means co-occurrence within the context windows that the training model has been set on. In our experiment, the context window is set on 15 words with a dimensionality of 50.[20]

The visualized semantic network shows two relatively tight groups, with the first being Χριστός and κύριος, and the other being λόγος, θεὸς, ἄνθρωπος, and λαός, whereas ψυχή appears to be relatively isolated. If we turn to the first group, we see that Κύριος is associated with ὁδός, θαυμάζω, χαίρω on the first level, and that Χριστός is associated with ἀληθῶς, σώζω, and δύναμαι. The grouping of Χριστός and Κύριος, and their relative distance to θεὸς, might inspire new literary and theological interpretations or hypotheses to be explored. For instance, the relative distance of "Christ" and "Lord" to "God" might suggest a lower Christology. Moreover, the terms associated with "Christ" and "Lord" are not divine terms, but rather terms meaning saving, being able, rejoicing, and wondering, which could signify a Christology leaning towards a wonderworking truth-teller.

In the θεὸς–λόγος–λαὸς-ἄνθρωπος cluster, we notice many connected terms; for example, λαός and λόγος share many associative terms, like ὅσος, καρδία, and ὑπάρχω, and, moreover, λόγος and θεὸς share χάρις and δοξάζω through δόξα.

The distance and similarity between the terms are represented in the visualization's proportionate distances between the queried terms (nodes) and in the similarity markers on the lines (edges) between the nodes. A low value (e.g., 0.01) signifies a high similarity and a frequent contextual co-occurrence of the terms. The Χριστός and κύριος cluster has distance values that indicate relative similarity between seed terms and associated terms, whereas the θεὸς, λόγος, and ἄνθρωπος cluster has relatively higher values (from 0.02–0.13) indicating a cluster with less tightly connected terms. The term ψυχή has distance values of 0.09–0.13 connected to the terms παραδίδωμι, τίθημι, σώζω, and δύναμαι. This also explains the coloring of the network. Each seed term has its own color. However, the associative terms of ψυχή are different colors, since the "closest" associative terms are in fact closer to other terms, thus making σώζω the color of the χριστός group, although σώζω in a network where ψυχή was the only seed term would have the same color as ψυχ ́ η. The associative words of ψυχή are simply further distanced than the associative words of the Χριστός–κύριος cluster. This network must be interpreted as showing distances and similarities relatively. In Figure 6, we have created a control network in order to position our input words in a larger network of value-laden terms.

This control network is made in order to compare our results with other input terms, like Σατανᾶς, λυπή, and σκότος, which helps to discover that all the input terms from topics 1 and 2 seem to be positioned relatively closer together than many of our control terms. The input words of topics 1 and 2 are positioned in the cluster to the right, whereas such terms as Σατανᾶς, λύπη, and σκότος are positioned to the left.

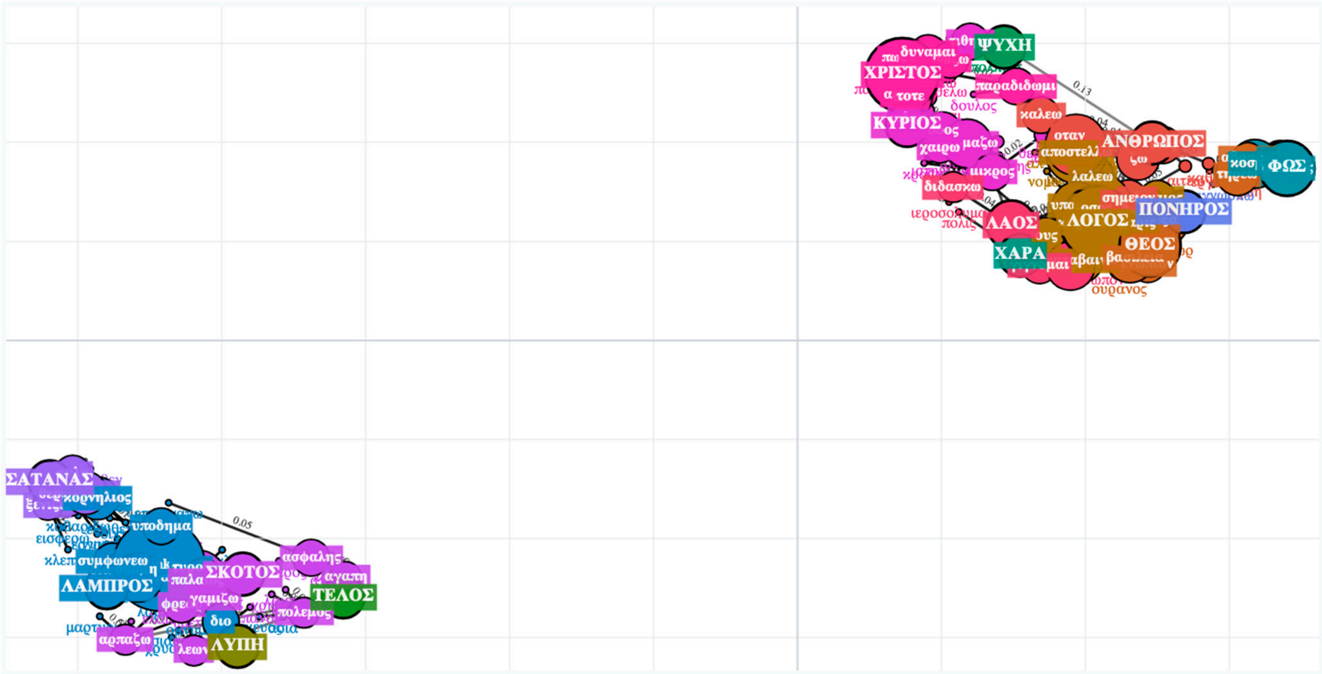

Figure 6. Control network with Topic 1 and 2 terms including "control" terms: λύπη, σκότος, Σατανᾶς, τέλος, πονηρός, χαρά, φῶς, and λαμπρός.

The methodological point to make here is that when we interpret these semantic networks, it should be regarded as a way of "reading" the texts we know so well again. This means that we regard the visualization, which is the result of a contextual embedding calculation, to be a valid representation of our source texts. There is nothing "new" or "different" in the visualization which is not in the NT texts. All the words are derived from the texts, and their distance from each other is based on how often and close they appear to each other in the sentences and on the pages of the NA28. The visualization is merely a new and different way of presenting the NT narratives as texts that can be read. Furthermore, this new visualization can help us as scholars explore new relationships between words and themes, which we might have missed by "merely" reading the sentences and pages of the NA28 from left to right and from start to end.

In the above analysis, we have surveyed New Testament narratives. We could dive deeper into the semantic structures by comparing semantic networks between individual terms. However, in the following section, we will conduct a third experiment, namely a word embedding analysis, but based on a different text corpus, namely "historiographic" texts from LXX (LXX historiographies).[21] This comparison will allow us to emphasize similarities and differences based on co-occurrence structures of the two corpora.

The results of the third experiment are shown in Figure 7. Here, we see a semantic network based on the input words, as in experiment 2, but the texts queried are LXX historiographies. The first observation to note is that the seed term χριστός is missing, and where we had two significant pairs in the previous analysis (χριστὸς and κύριος, and λόγος and θεὸς) we can now observe how κύριος is closer to λόγος. Moreover, a significant term of the New Testament (χάρις), which was situated closely between λόγος and θεὸς does not appear in the network of LXX historiography.[22]

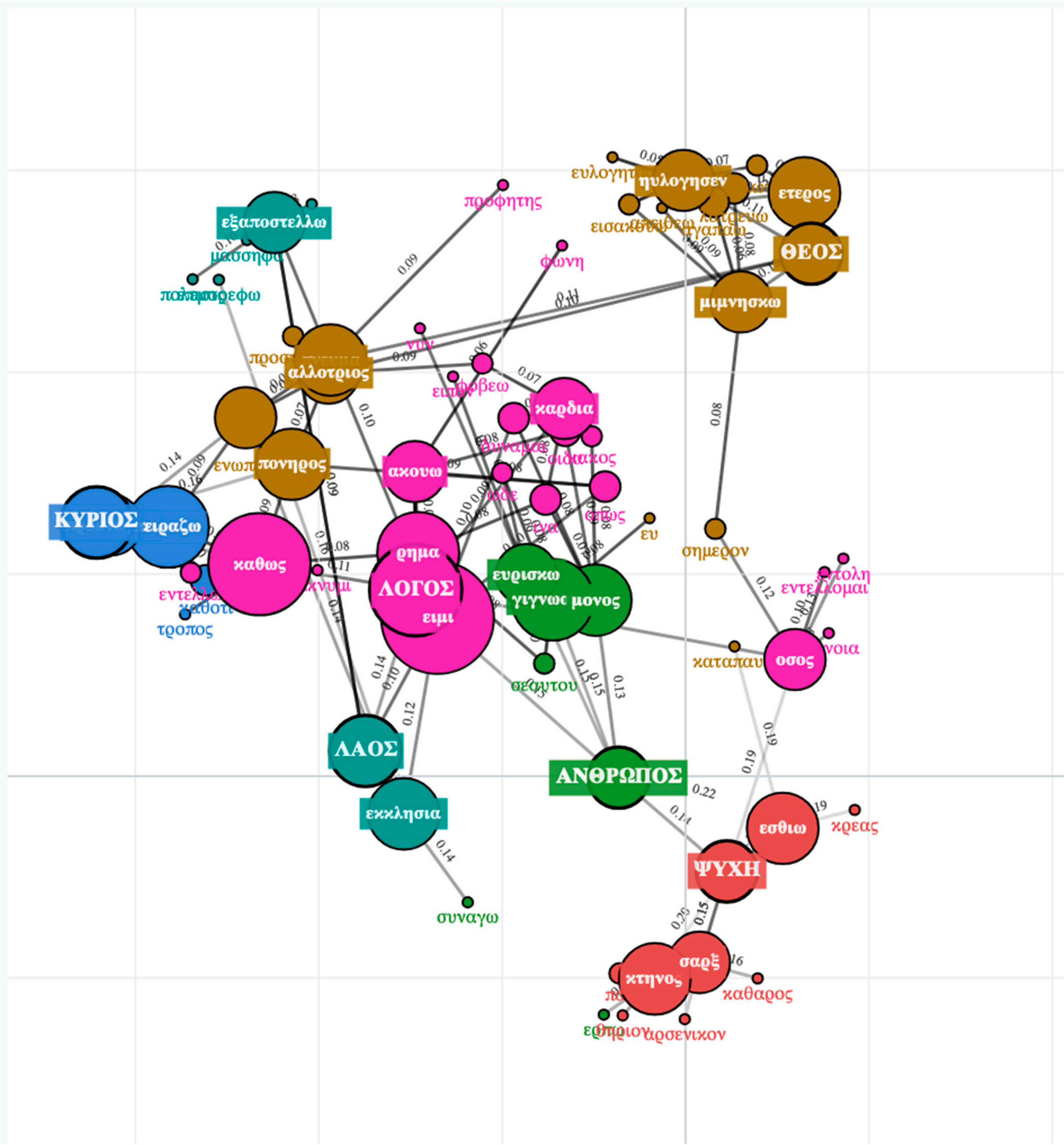

Figure 7. Semantic network of queried terms in LXX-historiographies.

The second main observation is that the cluster appears to be denser than the previous. However, if the distance measures are higher, this indicates less similarity between the terms. In the θεὸς group of the visualized semantic network in the LXX historiographies, there are many distinctive second-level associations, whereas in the New Testament narratives we could see that the θεὸς cluster shared many first-level associations with λόγος, e.g., δοξάζω. Moreover, χάρις had a first-level association to λόγος and a second-level association to δόξα. The circumstance that many terms in the NT narratives share first- and second-level associations together with the lower distance measures indicates that the terms in the NT narratives are contextually closer to each other than in the LXX historiographies.

If we concentrate on the individual co-occurrences of the seed words, it is noticeable that λόγος in the LXX historiographies as well as in the NT narratives is associated with the terms καρδία and ῥῆμα. Moreover, ἄνθρωπος and ψυχή are more closely related in the LXX historiographies than in the NT narratives. The terms related to ψύχη in the LXX historiographies are σάρξ, κτῆνος, ἐσθίω, and, as mentioned, ἄνθρωπος as first-level associations, with κρέας, καθαρός, θηρίον, and ἀρσενικός as second-level associations. In the NT narratives, ψυχή, with its close position to χριστός and its related terms σώζω δύναμαι, τίθημι, and ἀπόλλυμι, seems to be a concept that has ethical connotations, whereas in the LXX historiographies, ψυχή has different ethical connotations regarding meat, animals, food, cleanness, etc. This might not be completely new insights to scholars working with the soul in NT research and LXX research, respectively, but the visualizations have brought these two areas into very close contact and produced data ripe for interpretation.

Our results of the word embedding analyses contribute to ongoing New Testament research by highlighting variations in Christological perspectives and conceptual associations between New Testament narratives and LXX historiographies. In the New Testament context, the semantic network exhibited a distinct clustering around Χριστός and Κύριος, suggesting a focus on a wonderworking truth-teller aspect of Jesus, resonating with debates over low versus high Christology (Bauckham 2024, pp. 141–58; Culpepper 2024, pp. 113–26; Bühner 2021; Porter 2020). Furthermore, the close proximity between θεὸς and λόγος, along with shared associations in the New Testament, emphasized the centrality of divinity and grace, aligning with theological discussions on logos theology (Forger 2020; Anderson 2016). Comparatively, the LXX historiographies' semantic network provides insight into the evolving ethical connotations of various terms, like ψυχή, illuminating shifts in theological and ethical discourse between the Old and New Testaments (Runesson 2023). These findings invite further exploration into the contextual and theological developments within early Christian thought. Solid core competencies by NT and LXX scholars cannot be left out or eliminated. The computational analysis does not provide a shortcut to scholarly knowledge. Only competent scholars who know the texts and the secondary literature will be able to analyze, interpret, contextualize, and put into perspective the computational findings in a proper scholarly manner. In this sense, nothing has changed with the advent of computational methods. Although the scholarly "grasp" of the texts (and the "presentation" or "appearance" of the texts themselves) have changed, this should merely be welcomed as an opportunity to build new knowledge. It is in this sense that we perceive traditional close-reading approaches and computational approaches to supplement one other.

Whereas topic modelling in experiment 1 revealed the topical distributions on the one hand, on the other hand, our word embedding analysis in experiment 2 captured more nuanced differences between the chosen seed terms. Moreover, word embedding confirmed that these precise terms are closely knit together in New Testament narratives since the clusters overlap much more than in th LXX narratives, where the terms are relatively more distant and do not share as many terms.

3. Comparison and Discussion of Integration of Literary and Historical Context

In our conceptual analysis with word embedding, we could collectively analyze New Testament narratives, and we compared them with a contextually relevant group of LXX historiographies. These two analyses are identical in the computational setup of the parameters. However, the inputs are different. The results of these two experiments revealed interesting differences concerning the core concepts of these groups of texts. This means that, in addition to uncovering the topical and conceptual landscape of NT narratives, we have also engaged with the historical and literary context in our experiments. Now, we want to expand further on this.

In experiment 1, in the analysis of topical distribution, the analyzed corpus consisted of 2153 texts spanning from early to late antiquity. In the following experiment 4, we will show how topic modelling as a whole is unlike word embedding where one has to train a new model on the object of comparison, but that the very idea of topic modelling is, in actuality, a contextual approach.

In this fourth and last experiment, we will contextualize and place our six New Testament narratives within the large corpus based on the topical distribution. In Figure 8, we show a visualization of the results of this experiment. Figure 8 displays a scatter-plot where each dot represents 1 text in our 2153-text corpus. The colors of the dots mark which topic that particular text is dominated by. The position is based on the topical distribution. For example, if a text is light green, it means that it is dominated by the theology 1 topic, and when it is at the same time positioned close to dark green dots, this indicates that it also contains elements of the dark green topic (theology 2). This experiment can, thus, provide a detailed contextual insight, namely which texts are grouped close to each other based on the semantic distribution of their words ordered into topics. The results of experiment 4 are shown in Figure 8, where six scatter plot visualizations with each of the chosen texts (marked with a small black arrow) are shown.[23]

The visualization shows that all of the NT narratives are placed in the dark green group, but are very close to the light green group. This shows, as we know from experiment 1, that many of our texts in the corpus have high percentages of both topics 1 and 2. We notice that the light green group (theology 1) is divided, where one part is close to texts dominated by the dark green topic, and the other part of the group of light green texts are closer to the light blue group, which is topic 7, where a mix of rhetorical (Libanius) and philosophical texts were situated (Philo). This visualization shows that the topical distribution is not only descriptive of the individual texts, but also can relate the topical distribution with relevant texts that have consciously or unconsciously influenced our target texts representing a relevant portion of Jewish and Hellenistic textual culture. The very assignment of topical distribution is performed in relation to the literary and historical context. What topic modelling is best at revealing is the relation between texts based on their proportional topical distribution. Almost every Judeo-Christian text displays a high percentage of topics 1 and 2, but they also show topics from the philosophy and law and justice topics. Topic modelling can, thus, serve to position documents in a corpus, i.e., a relational function, as well as analyze their topical distribution individually, i.e., individual analysis. What can be discussed is how the interpretation of this placement should be understood. When we investigate texts through the lens of topical distribution, we begin to "read" texts as being multifaceted composites of topics. Concerning NT narratives, it is noticeable that our target texts are not positioned in the outskirts of the clusters, which would signify that they had a uniform distribution of topics. Instead, they display ambiguity and diversity, which, on the one hand, needs to be taken as a descriptive conclusion in itself, and, on the other hand, allows for further interpretation and discussion of the results.

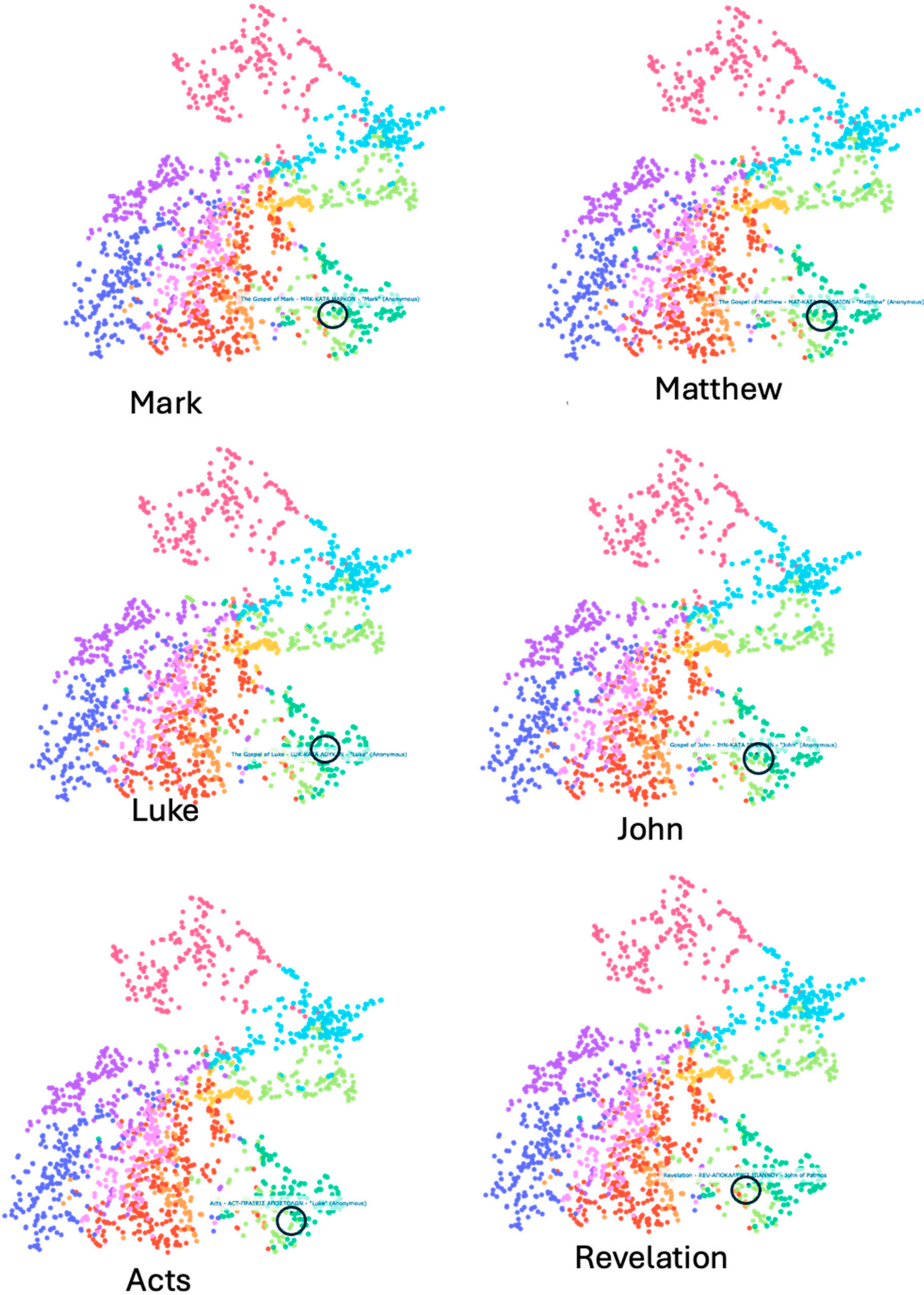

Figure 8. Scatterplot of target texts' position in the corpus.

4. Concluding Remarks

In this article, we have outlined, demonstrated, and discussed the use of computational methods, partly in engagement with a large corpus and partly as a group of texts consisting of New Testament narratives. We have, through four experiments, employed two methods, namely topic modelling and word embedding, with different purposes and with different input texts. The initial topic modelling investigation served to organize a large corpus of texts and uncover topics. Moreover, the topic modelling analysis could assist in analyzing the topical distribution of individual texts. In the final experiment, topic modelling also enabled us to position our target texts in a corpus and aided a contextual analysis. We chose to let the results of the topic model analysis guide our further analysis of the texts. In the following experiment (the word embedding analysis) we pre-trained a model on NT narratives in order to recognize contextual patterns. This allowed us to engage more closely with the texts and their conceptual landscape. Within the word embedding analysis, we compared the conceptual landscape of New Testament narratives to LXX historiographies and found, among other things, that these input words are more closely tied to each other in the NT narratives than in the LXX historiographies. Moreover, it was possible to analyze distinct terms, such as, for example, "soul", and how it is more connected to being saved and possibility in the NT narratives, whereas in the LXX historiographies, "soul" is contextually more related to eating and to the cleanness of animals. Word embedding allowed to query specific words and analyze their dispersion and distance values in relation to the other input terms in the queried corpus on the one hand, and, on the other hand, word embedding enabled, in addition to analyzing specific terms, aa comparison between them. Our aim has been to show how this approach is a kind of conceptual history that can maintain close contact with both the source text and the literary–historical context. We hold that computationally assisted analyses can assist traditional methods and be regarded as old wine in new wineskins—classic, hermeneutical questions can be posed with new methods.

Author Contributions: Conceptualization, C.H.V. and E.E.H.V.; Methodology, C.H.V.; Formal analysis, C.H.V.; Investigation, C.H.V.; Data curation, J.M.; Writing–original draft, C.H.V. and E.E.H.V.; Writing–review & editing, C.H.V., E.E.H.V. and J.M.; Project administration, J.M.; Funding acquisition, J.M. All authors have read and agreed to the published version of the manuscript.

Funding: The research in this article is funded by the Carlsberg Foundation in the Semper Ardens: Accelerate-project "Computing Antiquity: Computational Research in Ancient Text Corpora".

Data Availability Statement: The data presented in this study are openly available in the repositories https://github.com/centre-for-humanities-computing/computing-antiquity; https://github.com/centre-for-humanities-computing/classic-embeddings; https://github.com/centre-for-humanities-computing/classic-topic [Github] [https://github.com/centre-for-humanities-computing/computing-antiquity; https://github.com/centre-for-humanities-computing/classic-embeddings; https://github.com/centre-for-humanities-computing/classic-topic].

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1 On conceptual history cf. Sections 1.2 and 1.2.1 "Concepts" are in this context understood as second-order terms based on the calculated results through the computational experiments/analyses. Individually, words and terms are not concepts, but when we (through a computational/algorithmic model) have calculated their importance, we can treat them as second-order concepts, i.e., words that are calculated, and through this calculation, are considered to be "conceptual" and, thus, more than only first-order occurrences of important words. Terms, such as ψυχὴ and θεὸς, will have different nuances in New Testament and LXX texts (as we will show through our experiments), and computational methods, like topic modelling and word embedding, can assist in uncovering and analysing such differences based on the co-occurrence, i.e., the contextual structure of terms cf. Section 1.2. However, we are aware that no words are concepts "in themselves" in continuation of Wittgenstein's language philosophy and James Barr's critique of biblical "concepts" in *The Semantic of Biblical Language* (1961) and *Comparative Philology and the Text of the Old Testament* (1968).

2 The choice of these text groups is discussed in Section 2.2. We are aware that there are no "historiographies" in the strict (Greek) sense in the LXX, but we have chosen to designate the "history-writing" parts of the LXX (despite the ideological–religious biases) as historiographies. For scholars comparing the NT narratives with (Greek) historiographies, cf. Becker (2017) and Becker (2006).
3 For a definition of "computational humanities" as opposed to "digital humanities", see Johnson et al. (2024).
4 Proponents have been, e.g., (Moretti 2000b; Underwood 2019; Pannapacker 2012; Jockers 2013; Nielbo et al. 2017); examples of critical voices are (Brennan 2017; Conrad 2014) and Da (2019a, 2019b). Cf. also Vrangbæk and Nielbo (2021, p. 151).
5 Already Jockers has spoken of a "blended method", Jockers (2013, p. 26); for integrative approaches see also (Hu et al. 2021; Manjavacas Arevalo and Fonteyn 2022; Nielbo et al. 2019a, 2019b; Nielbo et al. 2017; Baunvig 2019, 2022, 2023; Agersnap et al. 2020; Agersnap et al. 2022).
6 An approach which we find supported in, e.g., (Hu et al. 2021; Nielbo et al. 2019a; Alkan et al. 2023; Gao and Xu 2021).
7 The figure and a similar description also appears in Vrangbæk (2024) (unpublished Ph.D. dissertation). The figure is inspired by Nielbo et al. (2017, p. 97) and Fayyad et al. (1996, p. 41).
8 Concerning choosing the right kind of vectorization, (see Mikolov et al. 2013b; Sineglazov and Savenko 2023; Turian et al. 2010, pp. 384–94).
9 Baunvig and Nielbo (2022, p. 24) comment that, "As is well known, 'the map is not the territory', but (functional) maps do, nevertheless, retain certain relevant properties of the territory it seeks to represent." (cf. also Goldberg 2017, pp. 90–92; Albrecht et al. 2021, pp. 270–71). The paragraph builds on passages from Vrangbæk and Nielbo (2023) and Vrangbæk (2024).
10 This hypothetical example also appears in Vrangbæk (2024).
11 As explored in Vrangbæk et al. (2024).
12 Concerning "term frequency-inverted document frequency" (e.g., Zhang et al. 2011; Albrecht et al. 2021, pp. 132–33).
13 In our case, the corpus consists of a conglomerate of the Perseus Corpus, First1KGreek, the Pseudepigrapha.org corpus, LXX, and NA28, in sum, 2153 documents. This database is hosted by associate professor Jacob P.B. Mortensen, Aarhus University, in the project "Computing Antiquity".
14 In our case, we carried out textual preprocessing with the OdyCy-pipeline (Kostkan et al. (2023)).
15 This was also the procedure of topic modelling in Vrangbæk et al. (2024).
16 In this analysis, we treat each NT text as a one, finished text following the main text of NA28. The topics that are assigned to the texts are calculated results that allow us to treat the words of the topics as descriptive concepts.
17 Cf. our conceptual approach that first-order terms based on modelling become second-order concepts.
18 The setting of NT narratives as one lump has clear methodological consequences, and for a future analysis of the individual New Testament narratives, a segmentation of each text is necessary. However, in this article, the textual setup serves the purpose of demonstrating and showcasing our approach.
19 Query parameters need to be set, and, in this case, we ask for six first-level associations and four second-level associations.
20 See parameters in the code: https://github.com/centre-for-humanities-computing/computing-antiquity/blob/main/src/embeddings/train_glove.py (accessed on 1 October 2024).
21 LXX historiographies is a label we have assigned to a group of texts that have historiographic elements, although not all texts are historiography proper. The text group comprises Genesis–Deuteronomium, Chronicles, Kings, and Samuel, including "apocryphal" writings, except III Maccabees.
22 We tested the results of the relation between χάρις and θεὸς with a different digital tool that is more common among scholars, namely the Thesaurus Linguae Graece. We set up an N-gram search of all Septuagint texts of χάρις and θεὸς with a 10-word distance. In this case, there are 4 occurrences in the texts that we have categorised as LXX historiographies, whereas we find 16 occurrences in the other Septuagint texts, among which many occur in the apocryphal texts which lie historically closer to the New Testament: Gen. 6,2 Νωε δέ εὗρεν χάριν ἐναντίον κυρίου τοῦ θεοῦ; 30;27 Εἰ εὗρον χάριν ἐναντίον σου, οἰωνισάμην ἄν· εὐλόγησεν γάρ με ὁ θεὸς τῇ σῇ εἰσόδῳ; Gen. 43,17 ὁ δέ θεός μου δῴη ὑμῖν χάριν ἐναντίον τοῦ ἀνθρώπου, καὶ ἀποστείλαι τὸν ἀδελφὸν ὑμῶν τὸν ἕνα καὶ τὸν Βενιαμιν; Kings 15,25 καὶ εἶπεν ὁ βασιλεὺς τῷ Σαδωκ Ἀπόστρεψον τὴν κιβωτὸν τοῦ θεοῦ εἰς τὴν πόλιν· ἐὰν εὕρω χάριν ἐν ὀφθαλμοῖς κυρίου. Esdras I 8,4; Judith 8, 18; 8, 23; 10,8; Tobias 1,12; 12,17; 12,18; II Macc 7,32; III Macc. 1,9; 6,36; Psalm 44,3; 83, 12; Prov. 18,22; Sir. 45,1; Baruch 1,13). Searches are based on Thesaurus Linguae Graecae© Digital Library. Ed. Maria C. Pantelia. University of California, Irvine. http://stephanus.tlg.uci.edu.ez.statsbiblioteket.dk (accessed on 12 October 2024).
23 The scatterplot is a frozen image of a 3D visualization, which is represented in 2D here.

References

Agersnap, Anne, Kirstine Helboe Johansen, and Katrine Frøkjær Baunvig. 2022. The Legacy of the Danish "Church Fathers" NFS Grundtvig and SA Kierkegaard in 11,955 Contemporary Sermons. Paper presented at the 6th Digital Humanities in the Nordic and Baltic Countries Conference (DHNB 2022), Uppsala, Sweden, March 15–18; pp. 159–67.

Agersnap, Anne, Ross Deans Kristensen-McLachlan, Kirstine Helboe Johansen, Uffe Schjødt, and Kristoffer Laigaard Nielbo. 2020. Sermons as Data: Introducing a Corpus of 11,955 Danish Sermons. *Journal of Cultural Analytics* 5: 1–27. [CrossRef]

Albrecht, Jens, Sidharth Ramachandran, and Christian Winkler. 2021. *Blueprints for Text Analytics Using Python: Machine Learning-Based Solutions for Common Real World (NLP) Applications*, 1st ed. Sebastopol: O'Reilly Media.

Alkan, Bilal Barış, Leyla Karakuş, and Bekir Direkci. 2023. Knowledge Discovery from the Texts of Nobel Prize Winners in Literature: Sentiment Analysis and Latent Dirichlet Allocation. *Scientometrics* 128: 5311–34. [CrossRef]

Anderson, Paul N. 2016. The Johannine Logos-Hymn: A Cross-Cultural Celebration of God's Creative-Redemptive Work. In *Creation Stories in Dialogue: The Bible, Science, and Folk Traditions*. Shinjuku: Brill, pp. 217–42. [CrossRef]

Armand, Cécile, and Christian Henriot. 2023. Beyond Digital Humanities Thinking Computationally: A Position Paper. Available online: https://shs.hal.science/halshs-04194570 (accessed on 1 October 2024).

Bauckham, Richard. 2024. The Waves, the Wind, and the Wilderness: Some High Christological Narratives in Mark and John. In *Early High Christology: John Among the New Testament Writers*. Edited by Christopher M. Blumhofer, Diane G. Chen and Joel B. Green. Augsburg: Fortress Publishers.

Baunvig, Katrine Frøkjær. 2019. Mapping Worldviews. Word-Embedding 'Heaven', 'Earth' and 'Hell' in the Writings of Danish Nation Builder and Church Father NFS Grundtvig (1783–1872). Paper presented at the EASR: Religion—Continuations and Disruptions, Tartu, Estonia, June 25–29.

Baunvig, Katrine Frøkjær. 2022. Ånd hos Grundtvig: En digitalhumanistisk skitse af begrebsbrugen i Grundtvigs værker. In *Åndsfrihed i det moderne samfund*. Edited by Ove Korsgaard and Michael Agerbo Mørch. Århus: Aarhus Universitetsforlag, pp. 65–102.

Baunvig, Katrine Frøkjær. 2023. A Computational Future? Distant Reading in the Historical Study of Religion. In *Stepping Back and Looking Ahead: Twelve Years of Studying Religious Contact at the Käte Hamburger Kolleg Bochum*. Edited by Maren Freudenberg, Frederik Elwert, Tim Karis, Martin Radermacher and Jens Schlamelcher. Leiden: Brill, pp. 325–52.

Baunvig, Katrine Frøkjær, and Kristoffer L. Nielbo. 2022. Mermaids Are Birds Embedding NFS Grundtvig's Bestiary. In *The 6th Digital Humanities in the Nordic and Baltic Countries Conference 2022, Uppsala, Sweden, March 15–18*. Edited by Karl Berglund, Matti La Mela and Inge Zwart. Ceur Workshop Proceedings. Uppsala: DHNB, pp. 23–32.

Becker, Eve-Marie. 2006. *Das Markus-Evangelium im Rahmen antiker Historiographie*. Wissenschaftliche Untersuchungen zum Neuen Testament. Tübingen: Mohr Siebeck.

Becker, Eve-Marie. 2017. *The Birth of Christian History: Memory and Time from Mark to Luke-Acts*, 1st ed. The Anchor Yale Bible Reference Library. New Haven: Yale University Press.

Bizzoni, Yuri, and Shalom Lappin. 2018. Predicting Human Metaphor Paraphrase Judgments with Deep Neural Networks. In *Proceedings of the Workshop on Figurative Language Processing*. Edited by Beata Beigman Klebanov, Ekaterina Shutova, Patricia Lichtenstein, Smaranda Muresan and Chee Wee. New Orleans: Association for Computational Linguistics, pp. 45–55.

Boleda, Gemma. 2020. Distributional Semantics and Linguistic Theory. *Annual Review of Linguistics* 6: 213–34. [CrossRef]

Boleda, Gemma, and Katrin Erk. 2015. Distributional Semantic Features as Semantic Primitives—Or Not. Paper presented at the 2015 AAAI Spring Symposium, Palo Alto, CA, USA, March 23–25; pp. 1–4.

Brennan, Timothy. 2017. The Digital-Humanities Bust: After a Decade of Investment and Hype, What Has the Field Accomplished? Not Much. *The Chronicle of Higher Education*, October 15.

Bühner, Ruben A. 2021. *Messianic High Christology: New Testament Variants of Second Temple Judaism*. Waco: Baylor University Press.

Conrad, Kathryn. 2014. What the Digital Humanities Can't Do. *The Chronicle of Higher Education*, September 8.

Culpepper, R. Alan. 2024. Davidic Messianism in Matthew and John. In *Early High Christology: John Among the New Testament Writers*. Edited by Christopher M. Blumhofer, Diane G. Chen and Joel B. Green. Augsburg: Fortress Publishers.

Da, Nan Z. 2019a. The Computational Case against Computational Literary Studies. *Critical Inquiry* 45: 601–39. [CrossRef]

Da, Nan Z. 2019b. The Digital Humanities Debacle. Computational Methods Repeatedly Come Up Short. *Chronicle of Higher Education*, March 27.

Ethayarajh, Kawin, David Duvenaud, and Graeme Hirst. 2019. Understanding Undesirable Word Embedding Associations. *arXiv* arXiv:1908.06361. [CrossRef]

Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. 1996. From Data Mining to Knowledge Discovery in Databases. *AI Magazine* 17: 37.

Firth, John Rupert. 1957. A synopsis of linguistic theory 1930–1955. *Studies in Linguistic Analysis*, 1–32.

Forger, Deborah. 2020. Jesus as God's Word(s): Aurality, Epistemology and Embodiment in the Gospel of John. *Journal for the Study of the New Testament* 42: 274–302. [CrossRef]

Gao, Jianbo, and Bo Xu. 2021. Complex Systems, Emergence, and Multiscale Analysis: A Tutorial and Brief Survey. *Applied Sciences* 11: 5736. [CrossRef]

Gillings, Mathew, and Andrew Hardie. 2023. The Interpretation of Topic Models for Scholarly Analysis: An Evaluation and Critique of Current Practice. *Digital Scholarship in the Humanities* 38: 530–43. [CrossRef]

Goldberg, Yoav. 2017. *Neural Network Methods in Natural Language Processing*. Edited by Graeme Hirst. San Rafael: Morgan & Claypool Publishers.

Harris, Zellig S. 1954. Distributional Structure. *WORD* 10: 146–62. [CrossRef]

Hu, Qiyue, Bin Liu, Mads Rosendahl Thomsen, Jianbo Gao, and Kristoffer L Nielbo. 2021. Dynamic Evolution of Sentiments in Never Let Me Go: Insights from Multifractal Theory and Its Implications for Literary Analysis. *Digital Scholarship in the Humanities* 36: 322–32. [CrossRef]

Ifversen, Jan. 2007. Begrebshistorien efter Reinhart Koselleck. *Slagmark-Tidsskrift for Idéhistorie* 48: 81–104.

Jensen, Kim Ebensgaard. 2014. Linguistics in the Digital Humanities:(Computational) Corpus Linguistics. *MedieKultur: Journal of Media and Communication Research* 30: 20. [CrossRef]

Jockers, Matthew. 2013. *Macroanalysis: Digital Methods and Literary History*. Philadelphia: University of Pennsylvania Press.

Johnson, Jessica Marie, Mimno, David, and Lauren Tilton, eds. 2024. *Debates in the Digital Humanities: Computational Humanities*. Minneapolis: University of Minnesota Press.

Jurafsky, Daniel, and James Martin. 2024. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Third Edition Draft. Available online: http://www.web.stanford.edu/~jurafsky/slp3/ed3book.pdf (accessed on 1 October 2024).

Kherwa, Pooja, and Poonam Bansal. 2019. Topic Modeling: A Comprehensive Review. *EAI Endorsed Transactions on Scalable Information Systems* 20: e2. [CrossRef]

Koselleck, Reinhart. 1972. Einleitung. In *Geschichtliche Grundbegriffe: Historisches Lexikon zur politisch-sozialen Sprache in Deutschland*. Edited by Otto Brunner, Werner Conze and Reinhart Koselleck. Band 1. Stuttgart: Ernst Klett Verlag, pp. XIII–XXVII.

Kostkan, Jan, Márton Kardos, Jacob Palle Bliddal Mortensen, and Kristoffer Laigaard Nielbo. 2023. OdyCy – A General-Purpose NLP Pipeline for Ancient Greek. In *Proceedings of the 7th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Dubrovnik: Association for Computational Linguistics, pp. 128–34. Available online: https://aclanthology.org/2023.latechclfl-1.14 (accessed on 1 October 2024).

Lee, Daniel D., and H. Sebastian Seung. 1999. Learning the Parts of Objects by Non-Negative Matrix Factorization. *Nature* 401: 788–91. [CrossRef]

Li, Yang, and Tao Yang. 2018. Word Embedding for Understanding Natural Language: A Survey. In *Guide to Big Data Applications*. Edited by S. Srinivasan. Cham: Springer International Publishing, vol. 26, pp. 83–104.

Liu, Alan. 2016. Is Digital Humanities a Field?—An Answer from the Point of View of Language. *Journal of Siberian Federal University. Humanities & Social Sciences* 7: 1546–52.

Mandikal, Priyanka, and Raymond Mooney. 2024. Sparse Meets Dense: A Hybrid Approach to Enhance Scientific Document Retrieval. Paper presented at the 4th Workshop on Scientific Document Understanding, Vancouver, BC, Canada, February 26; pp. 1–8.

Manjavacas Arevalo, Enrique, and Lauren Fonteyn. 2022. Non-Parametric Word Sense Disambiguation for Historical Languages. In *Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities*. Edited by Mika Hämäläinen, Khalid Alnajjar, Niko Partanen and Jack Rueter. Taipei: Association for Computational Linguistics, pp. 123–34.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Distributed Representations of Words and Phrases and their Compositionality. *arXiv* arXiv:1310.4546.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013b. Efficient Estimation of Word Representations in Vector Space. *arXiv* arXiv:1301.3781.

Moretti, Franco. 2000a. Conjectures on world literature. *New Left Review*, 54–68.

Moretti, Franco. 2000b. The slaughterhouse of literature. *MLQ: Modern Language Quarterly* 61: 207–27. [CrossRef]

Nielbo, Kristoffer L., Katrine F. Baunvig, Bin Liu, and Jianbo Gao. 2019a. A Curious Case of Entropic Decay: Persistent Complexity in Textual Cultural Heritage. *Digital Scholarship in the Humanities* 34: 542–57. [CrossRef]

Nielbo, Kristoffer L., Mads L. Perner, Christian Larsen, Jonas Nielsen, and Ditte Laursen. 2019b. Automated Compositional Change Detection in Saxo Grammaticus' Gesta Danorum. *Digital Humanities in the Nordic and Baltic Countries Publications* 2: 321–32. [CrossRef]

Nielbo, Kristoffer, Ryan Nichols, and Edward Slingerland. 2017. Mining the Past—Data-Intensive Knowledge Discovery in the Study of Historical Textual Traditions. *Journal of Cognitive Historiography* 3: 93–118. [CrossRef]

Pannapacker, William A. 2012. Digital humanities triumphant? In *Debates in the Digital Humanities*. Edited by Matthew K. Gold. Minneapolis: University of Minnesota Press, pp. 233–34.

Porsdam, Helle. 2013. Digital Humanities: On Finding the Proper Balance between Qualitative and Quantitative Ways of Doing Research in the Humanities. *Digital Humanities Quarterly*, 7.

Porter, Stanley E. 2020. Johannine and Pauline Christology. In *Johannine Christology*. Leiden: Brill, pp. 11–30.

Rhody, Lisa. 2012. Topic Modeling and Figurative Language. Publications and Research. January 1. Available online: https://academicworks.cuny.edu/gc_pubs/452 (accessed on 1 October 2024).

Runesson, Anders. 2023. *Judaism for Gentiles: Reading Paul beyond the Parting of the Ways Paradigm*. Tübingen: Mohr Siebeck.

Sherstinova, Tatiana, Olga Mitrofanova, Tatiana Skrebtsova, Ekaterina Zamiraylova, and Margarita Kirina. 2020. Topic Modelling with NMF vs. Expert Topic Annotation: The Case Study of Russian Fiction. In *Advances in Computational Intelligence*. Edited by L. Martínez-Villaseñor, O. Herrera-Alcántara, H. Ponce and F. A. Castro-Espinoza. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 134–51. [CrossRef]

Sineglazov, Victor M., and Illia M. Savenko. 2023. Comparative Analysis of Text Vectorization Methods. *Electronics and Control Systems* 2: 21–27. [CrossRef]

Špiclová, Zdeňka, and Vojtěch Kaše. 2020. 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. *Open Theology* 6: 423–39. [CrossRef]

Turian, Joseph, Lev-Arie Ratinov, and Yoshua Bengio. 2010. Word Representations: A Simple and General Method for Semi-Supervised Learning. Paper presented at the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden, July 11–16; pp. 384–94.

Uglanova, Inna, and Evelyn Gius. 2020. The Order of Things. A Study on Topic Modelling of Literary Texts.Paper presented at the Worshop on Computational Humanities Research (CHR 2020), Amsterdam, The Netherlands, November 18–20; pp. 57–76.

Underwood, Ted. 2019. Dear Humanists: Fear Not the Digital Revolution: Advances in Computing Will Benefit Traditional Scholarship—Not Compete with It. *The Chronicle of Higher Education*, March 27.

Vrangbæk, Christian Houth, Eva Elisabeth Houth Vrangbæk, Kristoffer Nielbo, Marton Kardos, and Jacob Mortensen. 2024. Non-Canonical Acts and their Topical Distribution. Paper presented at the First Semantic Digital Humanities Workshop at ESWC, Hernissos, Crece, May 27.

Vrangbæk, Eva Elisabeth Houth. 2024. What is Love Other than Will. An Investigation of Augustine's Concept of Love and Will Assisted by Computational Methods. Ph.D. dissertation, Aarhus University, Aarhus, Denmark.

Vrangbæk, Eva Elisabeth Houth, and Kristoffer L. Nielbo. 2021. Composition and Change in De ciuitate Dei: A Case Study of Computationally Assisted Methods. In *Paper presented at the Eighteenth International Conference on Patristic Studies Held in Oxford 2019. Volume 14, Augustine of Hippo's De ciuitate Dei: Content, Transmission, and Interpretations*. Edited by Markus Vinzent, Anthony Dupont and Gert Partoens. Leuven: Peeters.

Vrangbæk, Eva Elisabeth Houth, and Kristoffer Nielbo. 2023. Love Between Desire and Will—An Investigation of Augustine's Concept of Love Assisted by Computational Methods. In *Augustine & Ethics, Augustine in Conversation*. Lanham: Rowman & Littlefield.

Wevers, Melvin, and Marijn Koolen. 2020. Digital Begriffsgeschichte: Tracing semantic change using word embeddings. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 53: 226–43. [CrossRef]

Willkomm, Jens, Christoph Schmidt-Petri, Martin Schäler, Michael Schefczyk, and Klemens Böhm. 2018. A Query Algebra for Temporal Text Corpora. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*. Fort Worth: ACM, pp. 183–92.

Zhang, Wen, Taketoshi Yoshida, and Xijin Tang. 2011. A Comparative Study of TF* IDF, LSI and Multi-Words for Text Clas sification. *Expert Systems with Applications* 38: 2758–65. [CrossRef]