Supporting Biblical Analysis Using Natural Language Processing:

A Preliminary Study with the Book of Psalms

Timothy Lee

Teachers College, Columbia University

**Abstract**

The Book of Psalms, a part of the Old Testament, is of high religious significance in the Judeo-Christian tradition. Scholarship attracted to the Bible and Book of Psalms for this reason have mostly relied on humanistic approaches from literature and history. The emergence of computational criticism in the highly related field of literary analysis suggests the promise of computational approaches for biblical analysis. This study is a preliminary exploration of this idea. An unsupervised machine learning technique, Latent Dirchlet Allocation (LDA), is applied to the Book of Psalms with each psalm as a text document. Output from the LDA model is used for thematic discovery and exploring type distinctions in relation to traditional biblical research on these issues. The LDA model seemed to affirm themes identified in Form-Critical theory. However, it was difficult to draw conclusions regarding psalm types, with multiple potential reasons for the pattern of results uncovered. Despite these findings, the questions opened by the study and finding limitations offer further opportunities for research, affirming the promise that computational methods have in developing our understanding of biblical texts.

**Introduction**

**The Book of Psalms**

The Book of Psalms, or ‘Psalter’, is a collection of 150 ‘psalms’. These are poems, prayers or songs of worship from Biblical Israel (Bullock, 2001). It is part of the Old Testament Books of Wisdom.

The psalms have high significance in both Jewish and Christian religious traditions. Some psalms were written for public worship in the Temple and Synagogues, while others were used for private worship (Bullock, 2001). Their use remains in the Jewish daily prayers and synagogue services. In Christianity, the psalms are also prominent in both public and private worship, and have been very influential in the development of the history of individual Christian piety (Weiser, 1962).

**The Book of Psalms in Research**

The importance of the Book of Psalms to religious practice has attracted scholarship to develop a deeper understanding of the text and its use in religious practice. Two major interests are the types and themes of the Psalms, as well as their historical context (Weiser, 1962).

**Types and Themes of the Psalms.** As a whole, a major theme of the psalms is the connection between the individual and collective in worship of the divine and the dispensation of salvation (Weiser, 1962). At a finer level, the Form-Critical approach of studying the psalms do so according to main ‘types’, each with structures and thematic focuses.

There are competing Form-Critical approaches (Bullock, 2001). Westermann pioneered one approaches (Bullock, 2001), arguing for a distinction between Psalms of Lament and Psalms of Praise. Psalms of Lament tend to have themes of judgement, calamity, supplication, penitence, and the saving mercy of God, while psalms of praise, or hymns, tend to focus on praising and proclaiming knowledge of God in worship. Psalms of lament are further classified according to the speaker (individual or community), whereas Psalms of Praise are classified into declarative or descriptive (Bullock, 2001). Declarative praise is general praise of God for helping the people, whereas descriptive praise specifies God’s action (Bullock, 2001). Form-Critical theory thus provides a quick overview of themes in the psalms, with different ‘types’ having different thematic focuses.

**Historical Context of the Psalms.** Understanding of the historical context is a major strain of modern biblical scholarship (Richardson, 1963), including the study of the Book of Psalms. Some of this research is related to Form-Critical theory; classifications of psalm ‘types’ were based on scholars’ understanding of the historical purpose of the psalms (Bullock, 2001).

However, historical research into the writing, editing, and compilation of the Book of Psalms complicates a clean division of types suggested by a staunch adherence to the Form-Critical theory approach. Much of this research has questioned a simplistic view of the psalms as authored by those attributed to in the psalms’ superscriptions, and the Book of Psalms as a cohesive text (Bullock, 2001; Weiser, 1962). Scholars suggest that the psalms were likely anonymously written and only later attributed (Weiser, 1962). The book also likely went through phases of collection and editing, leading some to suggest that the Book of Psalms consists of five books (Bullock, 2001). During the processes of collection and editing, some of the psalms may also have been altered or stitched together (Bullock, 2001). Differences in authors and time periods of changes to the text make the notion of clear types or themes mapping to a set of psalms difficult. Thus, study of the historical context of the psalms often bolsters the notion that the form-critical approach is insufficient for a full understanding of the psalms (Weiser, 1962).

**Natural Language Processing and Biblical Analysis**

While the study of biblical texts has traditionally been a largely humanistic endeavour, computational techniques have potential to support or add to human insights. This potential is suggested by the practice of “computational criticism” in the closely related field of literary analysis (Long & So, 2016). In computational criticism, Natural Language Processing (NLP) techniques are applied as a form of literary criticism. Long and So (2016) argue that close reading, historicist, and computational approaches to textual analysis provide different understandings of the text tied to the ontology of the technique. Rather than viewing one approach as better than another, critical perception of the text can be enhanced by paying attention to the interface between the approaches (Long & So, 2016).

A similar application of computational techniques to biblical analysis may similarly enhance our understanding of the Book of Psalms. Relying on different ontological emphases from historical interpretation and Form-Critical approaches, NLP techniques may yield insights about the text that humans may not have seen. They may also lend support or question literary interpretations or classifications, such as those suggested by Form-Critical approaches. Finally, these insights from NLP may support the progress of historical criticism by suggesting a reinterpretation of historical inferences or textual classification. Taking a computational approach to biblical analysis is hence worth at least some exploration.

**Latent Dirichlet Allocation (LDA).** One NLP technique that could be valuable for analysis of the Book of Psalms is Latent Dirichlet Allocation (Blei, Ng, & Jordan, 2003). LDA is a Bayesian probabilistic model often used for topic modelling of text corpora. LDA takes a text corpus, a set of texts, as an input. The model then represents each text as a mixture of latent topics (topic dominance). Each topic is characterised by a weighted distribution of individual words or n-grams (Blei et al., 2003), encompassing the themes of the whole text corpus.

LDA is particularly suited as a computational approach of deriving topics of literary or biblical texts to be interpreted as themes alongside humanistic reading approaches. One problem with supervised machine learning approaches in topical extraction is that there are never unambiguous ‘topic labels’ that map on individual literary or biblical texts. The use of supervised machine learning approaches in textual analysis requires an expert categorisation phase that the model uses as training and validation to classify texts according to the labels (Long & So, 2016). This undermines the objective of computational approaches as a reading approach complementary to humanistic methods, rather than reliant on them. As an unsupervised machine learning approach, LDA relies on the distribution of words in the text alone, without the need for clear labels for training and validation. It thus does not rely on expert labels, but its output can serve as an alternative set of themes, mapped to individual texts in the corpus, to be interpreted alongside human perspectives.

However, the lack of clear labels also makes the use of LDA difficult. There is no way of saying how good the output of the LDA model is when there are no ‘right’ answers with which to validate its output. This is problematic as LDA requires the user to set the number of topics in the text, and output can vary depending on how this parameter is set. Thus, rather than just picking the ‘best’ model based on a performance metric, the user faces a paradox. He or she must know the number of topics in the text to run an LDA model which is supposed to provide a set of topics for the text that the learner is trying to discover without prior reading assumptions.

Topic Coherence (Röder, Both, & Hinneburg, 2015) can be used as a marker of model quality to work around this problem. Topic coherence measures provide a quantitative metric of the interpretability and consistency of topics outputted by LDA (Röder et al., 2015). Given that the objective of using LDA is to provide interpretable topics about the text to be used as themes, an LDA model with more interpretable topics (a higher topic coherence score) could be reasonably taken as a ‘better’ model. Thus, topic coherence can be used to choose between LDA models with differing parameters. The ‘best’ model may be selected to generate topics to be interpreted alongside humanistic analyses.

**The Present Study**

The present study is a preliminary effort in using LDA to provide greater insight in understanding the Book of Psalms. LDA is applied to the Book of Psalms as a text corpus, with each individual psalm as a text. The topics that are outputted by the model are then used to scrutinise assumptions implied by Form-Critical and historical understandings of the psalms. Specifically, we ask two questions about the Book of Psalms in relation to these prevailing approaches.

First, does LDA affirm the themes of psalms identified by Form-Critical approaches? Strong support for a strict Form-Critical approach would be suggested by topics with high coherence scores, with the topics clearly reflecting the different themes identified in the psalms, such as praise or judgement. Poorly coherent and similar topics would also suggest a prevalence of generic or interacting themes in the psalms over strict distinctions between themes and psalms that may be related to historical influences of mixed authorship, stitching, and editing.

Second, does LDA affirm types of psalms identified by Form-Critical approaches? Strong support for a strict Form-Critical approach may be indicated by individual psalms being dominated by one, or a few, highly coherent topics reflecting the themes identified by Form-Critical approaches. On the other hand, the historical influence of editing, stitching, and mixed authorships may suggest mixed topic dominance in at least some psalms if topics are highly coherent. If topics are not highly coherent, topic dominance would offer little insight into whether there are distinct types of psalms. High dominance of singular topics may occur as topics are generic rather than supporting distinct themes that differentiate between types of psalms, whereas mixed dominance of poorly coherent topics is at best weak evidence of a distinction between types.

**Methodology**

**Data**

The full text of the Book of Psalms World English Bible (WEB) Translation was taken from an online webpage. WEB is a free publicly available translation of the Bible published online since 1997 (Johnson, 2019). It is an updated version of the American Standard Version (Johnson, 2019).

An initial attempt was made to use the American Standard Version taken from another webpage. However, the use of archaic word forms such as ‘standeth’ were inappropriate for modern natural language processing tools that assume that modern English word forms are used, such as the Natural Language Tool Kit (NLTK; Bird, Klein, & Loper, 2009) WordNet Lemmatizer, used in this study. As the WEB translation explicitly sought to convert archaic word forms to modern equivalents, it is more appropriate for use with common natural language processing tools.

**Computational Tools**

The full text of the Book of Psalms was copied from the webpage using a web browser and saved in a local text file. This was then read into Python as a text object. All subsequent processing was done through Python 3 (Van Rossum & Drake, 2009) and associated packages, through Jupyter Notebook (Kluyver et al., 2016). Packages used included Pandas for data manipulation, matplotlib (Hunter, 2007) and pyLDAvis (Sievert & Shirley, 2014) for visualisation, NLTK (Bird et al., 2009) for text processing and genism (Řehůřek & Sojka, 2010) for LDA implementation.

**Procedure**

**Text Cleaning.** Before training an LDA model, the text needs to be processed. Text should be cleaned to remove redundant word features that could distract the model from meaningful insight. Text also needs to be wrangled into a format that can be used by the LDA Model implementation.

First, text was cleaned to remove titles and common phrases in the superscriptions of the psalms such as “The Book of Psalms”, “Chief Musician”, “A Psalm of David”. These phrases describe the numbering, usage, or attribution of the psalms and their inclusion in the text corpus are distractions from the intended discovery of text themes. The full corpus was then split into individual psalms, producing 150 individual documents, each of one psalm, within the corpora. Text was then converted to lower case letters to prevent the same words with different capitalisations from being treated as different words. Chapter and verse numbers were also removed. The psalms were then ‘tokenised’ into individual words associated with each psalm. Tokenisation is important for building a ‘dictionary’ of features that is used by the LDA model.

Further text cleaning was conducted to remove stop words, common words that serve mostly syntactic purposes that do not add much meaning to texts, from the corpus. An initial list of stop words was taken from the NLTK English stop words corpus. The presence of stop words in text corpora is conventional in NLP as they occur very frequently despite having little meaning, skewing most models to place greater emphasis on words of little interest to the user. Iterations of the procedure mentioned here led to additional stop words being removed, and the model being run again. These words showed high prevalence in the topics that resulted from the model but were deemed as stop words as they added little value to the interpretation of meaning. In this list were the syntactic words “like”, “shall”, “let”, and “also”, as well as the proper nouns “Yahweh” and “God”. While the latter two words had clear meaning and significance, their importance and prevalence amongst topics offers no insight to any interpretation of the psalms, which are already known to be focused on Yahweh or God.

The words were then lemmatised using the NLTK WordNet Lemmatiser. The WordNet Lemmatiser relies on a large, publicly available English language database that matches words to their base lemma forms. Lemmatisation references words in the corpus to this database, reducing inflected words to their base lemma forms. This prevents words with the same semantic meaning from being treated differently due to morphological and other inflections, which may affect textual analysis.

Finally, genism functions were used to associate the tokenised texts and corpus with word IDs and frequencies within the corpus and texts. These were stored as objects to be used to train the LDA model.

**LDA Model Training.** To select the best model, 30 LDA models, ranging from having one topic to 30 topics, were trained. Their overall topic coherence scores were compared to each other by plotting a graph. The ‘best’ model was chosen as the one that balanced a high coherence score with a reasonable number of topics suited for human interpretation. As the procedure was run iteratively multiple times to improve the topics by removing additional stop words, the number of topics of the ‘best’ model changed with each iteration. Despite the iterations and use of multiple models, all models had coherence scores between 0.2 to 0.4, with most around 0.25 to 0.33. The final ‘best’ model had 5 topics with an overall topic coherence of 0.291. This was only marginally poorer than the maximum topic coherence in the group, 0.305, in a model with 26 topics.

**LDA Model Interpretation.** Aside from overall topic coherence, The LDA model produces other outputs that can be interpreted for the research questions.

*Themes of the psalms*. The themes of the psalms identified by Form-Critical approaches can be compared to the weighted distribution of words for each of the five topics produced by the model. These can be investigated further using an LDA visualisation tool, pyLDAvis (Sievert & Shirley, 2014). pyLDAvis (see Figure 1 on the last page) provides an interactive means of analysing LDA topics. On the left, it represents topics in terms of their intertopic distance along principal components. Selecting topics on the graph or using the selector above it displays the most highly weighted words for the topic on the right as a bar chart. The bars are divided into blue and red components, with the blue components showing overall term frequency while the red portion shows term frequency given the selected topic. Moving the slider above the word weightings adjusts the relative weighting of the “Top-30 Most Relevant Terms” for the topic displayed. While this is the same as the topic’s word weightings when λ = 1, adjusting the slider to reduce the value of λ increases the emphasis of terms given the topic over overall term frequency. Thus, adjusting λ provides an additional tool for interpreting topics, particularly in comparing topics that may be highly similar.

*Types of Psalms*. The types of psalms identified by Form-Critical theory may be investigated by analysing topics not just in terms of their content, but also their dominance within individual psalms.

**Results**

**Does LDA affirm the themes of psalms identified by Form-Critical approaches?**

The themes of psalms identified by Form-Critical approaches can be compared against the topics generated by LDA. The top 8 weighted words for each topic generated by the model are presented in Table 1. These words were analysed qualitatively to draw cohesive ideas for each topic.

Topic 1, containing words like “hearts”, “souls”, “kindness”, “loving”, and “wicked” suggest a theme of hearts and souls singing praise to God who is loving kind, deals with the wicked. The topic is indicative of proclamation of God’s nature, which is in line with praise form, where proclaiming knowledge of God was a part of praise and worship.

Topic 2 contains many words that are similar to topic 1, suggesting a highly similar topic. This is also supported by the high overlap between topics 1 and 2 in the intertopic distance map. However, at low λ, topics 1 and 2 differ in topic 1 having more focus on the comfort and mercy of God in words like “refuge”, “place”, and “salvation”, while topic 2 has more generic words such as “mouth”, and “law” (the Jewish law). This difference, and the presence of words like “wicked” in topic 1 compared to “earth” and “hand” suggest that topic 1 covers praise of more specific actions of God, such as punishing the wicked and providing places of refuge. In contrast, topic 2 suggest more generic praise of God and his “hand” in praise by the “earth”. This corresponds to Westermann’s distinction between descriptive and declarative praise (Bullock, 2001), mapping roughly to topic 1 and 2 respectively.

Topic 3 is similar to topics 1 and 2, although the two most highly weighted words, “us” and “people”, suggest a greater emphasis on the humans as the initiators of praise rather than the divine object of praise.

Despite no overlap with topics 1 on the intertopic distance map, the most highly weighted words of topic 4 seem to have trivial differences from those of topic 1, save from a greater emphasis on divine attributes. At low λ, topic 4 suggests armed conflict such as “raze”, “army”, “burned”, and “foreign”, in contrast with topic 1’s emphasis on refuge and sanctuary. One possibility is that that topic 4 captures either the process of achieving salvation in God defeating enemies. Another possibility is that topic 4 captures the themes of the psalms of lamentation, which follows a format of judgement, calamity, supplication, penitence, and the saving mercy of God (Weiser, 1962). Where the earlier portions of judgement and calamity match the conflict related terms of topic 4 at low λ, the later portions of thanksgiving and God’s mercy which draw on similar ideas as the psalms of praise are reflected in topic 4 at high λ parts that are similar to topics 1, 2, and 3.

The fifth topic differs from the others in words such as “Nation” and “Israel”. The ninth most salient word, excluded from the table, is “David”, despite the removal of “David” as an attributed author in the superscriptions during text cleaning. At low λ, topic 5 shows words such as “tribe”. These terms suggest that topic 5 reflects a generic reference to the Jewish nation (David was the second King of the Jewish nation, formed of the 12 tribes of Israel).

Overall, poor coherence and the fact that topics were highly similar to each other in their most highly weighted words suggests a prevalence of generic and interacting themes. However, these do not seem to be clearly attributable to historical influences of authorship, editing, or stitching, but the nature of the Book of Psalms itself. The words seen in the topics indicate that the generic, interacting theme would likely be an overall emphasis on praise. This is unsurprising given that the psalms are a genre focused on praise and worship of God, and themes of praise cut across all type distinctions made by Form-Critical approaches (Weiser, 1962).

Despite this assessment, analysis of the topics from the LDA model continues to indicate at least loose support the themes and offered by the Form-Critical approach. Despite low overall topic coherence that would suggest difficulty in interpreting the individual topics, all topics seemed to be interpretable with some effort and prior knowledge of the themes identified by humanistic approaches. Individual topics seemed to capture different aspects associated with praise, from the actions of God (topic 1), nature of God (topic 2), people and nation praising God (topics 3 and 5), and the relation between praise and other aspects of the psalms of lamentation (topic 4). Although these individual topics support the more specific themes identified by Form-Critical theory, the overarching theme that unites these smaller themes is still one of praise.

**Does LDA affirm types of psalms identified by Form-Critical approaches?**

Although the overall coherence score of the topics was low, topics were still interpretable in a manner that seems to support Form-Critical classifications, as discussed above. Also, few psalms exhibited mixed topic dominance; only 27 of the 150 psalms had dominant topics weightings lower than 0.75. Combined, these results suggest promising potential that the LDA model was able to discover and make similar psalm-type matching as Form-Critical scholars, lending support to their approach. They also suggest that the influence of editing, stitching, and melding of forms may be limited.

However, comparing psalms’ dominant topics to a list of psalm-type classifications (Bratcher, 2018) erodes support for this view. Taking the psalms of lament as a n example, only 9 of 34 psalms with topic 4 as their dominant topic were classified as psalms of lament. Meanwhile, 8 out of 13 psalms with topic 3 rather than topic 4 as their dominant topic were classified as psalms of lament. Also, with about a third of psalms classified as psalms of lament (Bratcher, 2018; Weiser, 1962), the model viewed more than half the psalms of lament as having a different dominant topic. Thus, although the LDA model seemed to be able to extact the same themes as biblical scholars in the psalms, it did not seem to match sets of themes to specific psalms in the same way as scholars have.

Mixed topic dominance also appears to be a misleading marker of the influence of stitching and editorship. Psalm 42-43, a textbook example of a stitched psalm (Bullock, 2001), was given a single dominant topic of over 90% weigtages for both psalm 42 and 43, rather than mixed weightage. This might reflect the editor stitching the psalms to develop more cohesive themes.

Overall, the LDA model seems to have affirmed the types of psalms and their thematic focuses in a manner similar to biblical scholars. However, the model’s attribution of these types and themes to psalms appear to diverge from Form-Critical approaches. The influence of the historical evolution of the Book of Psalms was also not clear in the model output. Where it was initially hypothesised that stitching of psalms would result in mixed topic dominance, stitched psalms had high dominance of single topics. Yet, alternative historical accounts of editors’ desire for coherence may account for the results observed.

**Discussion**

This study was a preliminary exploration of using computational methods to draw insight from biblical texts, serving as a complement to humanistic methods of inquiry. An LDA model was trained on the Book of Psalms as a text corpus with each individual psalm as a text document.

Without making prior assumptions about the Book of Psalms, I asked what themes the model would be able to extract from the corpus, and how they compare to themes identified by Form-Critical approaches. The themes drawn from the model were taken from the weightings of words that are used to form the model’s outputted topics. The chosen model had 5 topics, with themes that were interpretable despite a low overall topic coherence score. The topics were highly similar in expressing a theme of praise, supporting a notion that praising God permeated all the psalms. Individual topics captured various aspects of praise, including the actions and nature of God, the people and nation praising, as well as judgement and calamity that precedes praise in the form of psalms of lamentation. These topics are quite similar to aspects of praise that are main themes within different psalm types as identified by the Form-Critical approach (Bullock, 2001; Weiser, 1962). Hence, the model seemed to have drawn the same broad themes from the Book of Psalms as biblical scholars.

I also asked whether LDA would identify distinct psalm types, comparable to the types of psalms identified by Form-Critical approaches. The dominance of different topics from the model for each individual psalm was used to identify psalm types. Given the interpretability of the topics in similar terms to themes based on Form-Critical approaches, we were confident that psalms that shared the same dominant topic could be seen as the same type, and these would be similar to the types as identified by Form-Critical approaches. Most of the psalms were overwhelmingly dominated by a single topic. Unfortunately, there was poor corroboration between model topic-psalm mappings and scholars’ theme-psalm mappings.

The conclusions that can be drawn from these findings are not entirely clear, and open new questions for researchers. One potential reason interpretation of LDA topics made in this paper may be overly liberal. That topics with poor coherent scores could be so nicely interpreted with results very similar to prevailing approaches relying on different ontological bases may be too good to be true. My interpretations may have been skewed by prior knowledge of previous classifications and the text itself. The interpretation of the topics from the LDA model, and their relation to previous thematic analyses of the book of psalms, deserves closer scrutiny.

Another reason for the results may be related to limitations of the computational approach. The prominence of the overarching theme of praise across the psalms may have confused attribution of dominant topics. LDA also relies heavily on word frequencies and co-occurrence. The technique does not account for narrative structure, which is important to the interpretation of the psalms given that they often follow forms with different thematic emphases spread across the text. This is particularly true of the psalms of lamentation, where the structure contrasts calamity with salvation at opposite ends of the psalm. Consideration of narrative structure may lead biblical scholars to diverge from LDA’s word frequency and relation-based analyses in classification of types. Nevertheless, while scholars may have picked up on aspects of the psalms that computational models miss, the LDA model may have also identified different frames of classification that may be worth further study using traditional humanistic methods of biblical analysis. The differences in approaches makes both pursuits and comparison across them worthwhile in understanding the psalms more deeply.

The influence of editing, mixed authorship, and stitching was also unclear from the pattern of topic dominance, with differing accounts of history able to support or challenge the findings. This highlights the need to scrutinise the output of computational criticism research in light of prevailing historical and literary work.

**Other Limitations and Future Directions**

At a simple level, further efforts at removing stop words, using other LDA implementations, or newer unsupervised machine learning methods may improve on the methods used in this study. This may develop findings and insights that are potentially richer than those drawn here.

Other choices made in this study may also suggest specific areas for future work. For instance, the lack of prevalence of words related to lament in the topics produced by the model may have been due to psalms of lamentations being underrepresented in the book of psalms relative to psalms of praise, and the overall theme of praise. Psalms of lamentation only take up about a third of the Book of Psalms (Bratcher, 2018; Weiser, 1962). Future work may seek to equalise the representation of psalms of lamentations with psalms of praise to see if more distinct and coherent topics result.

In this study, no a priori assumptions were made regarding the topics that the model could extract from the texts. Pre-seeding the LDA model with terms drawn from Form-Critical approaches may steer the model towards finding topics and psalm-topic relations more similar to themes and types that biblical scholars are familiar with. An even stronger emphasis on building on prior work in computational criticism of the Book of Psalms would be to use supervised machine learning techniques with expert labels drawn from Form-Critical type classifications to classify the psalms. The model that emerges could be investigated for insights into the psalms. Starting on closer common ground between traditional biblical scholarship and computational criticism may be of greater value to comparisons between the approaches and developing deeper understandings of biblical texts.

Another promising area for future work would be to expand the use of LDA and other computational criticism approaches to related texts from the Near East. This may develop further insights into the development and influence between canonical biblical texts, non-canonical texts from the biblical period, and texts from the region.

**Conclusion**

Biblical research has largely focused on methods from literature, history, and the social sciences. This study provides a preliminary proof-of-concept of the value of computational approaches to studying the bible. Latent Dirichlet Allocation is applied to the Book of Psalms with each psalm as a text document, with output interpreted in light of prevailing ideas from historical and Form-Critical approaches. Where the LDA model seemed to affirm themes identified in Form-Critical theory, LDA raised more questions than answers when investigating the notion of psalm types. I believe that these questions, other future directions proposed, and the countless other opportunities for the application of computational techniques in biblical analysis provide promising avenues of deepening our understanding of a book so significant to human culture and history.

**References**

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit*. O’Reilly Media Inc. Retrieved from https://www.nltk.org/book/

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Bratcher, D. (2018). Types of Psalms: Classifying the Psalms by Genre. Retrieved December 16, 2019, from http://www.crivoice.org/psalmtypes.html

Bullock, C. H. (2001). *Encountering the Book of Psalms: A Literary and Theological Introduction*. Grand Rapids, MI: Baker Academic.

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Johnson, M. P. (2019). The World English Bible (WEB) FAQ. Retrieved December 14, 2019, from http://www.magister.msk.ru/library/bible/english/web/

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., … Jupyter Development Team. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In *Proceedings of the 20th International Conference on Electronic Publishing* (pp. 87–90). Göttingen, Germany. https://doi.org/10.3233/978-1-61499-649-1-87

Long, H., & So, R. J. (2016). Literary Pattern Recognition: Modernism between Close Reading and Machine Learning. *Critical Inquiry*, *42*(2), 235–267. https://doi.org/10.1086/684353

Řehůřek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45–50). Valletta, Malta: ELRA.

Richardson, A. (1963). The Rise of Modern Biblical Scholarship and Recent Discussion of the Authority of the Bible. In *The Cambridge History of the Bible* (pp. 294–338). Cambridge University Press. https://doi.org/10.1017/CHOL9780521042543.009

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining - WSDM ’15* (pp. 399–408). New York, New York, USA: ACM Press. https://doi.org/10.1145/2684822.2685324

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (pp. 63–70). Baltimore, Maryland, USA: Association for Computational Linguistics. https://doi.org/10.3115/v1/W14-3110

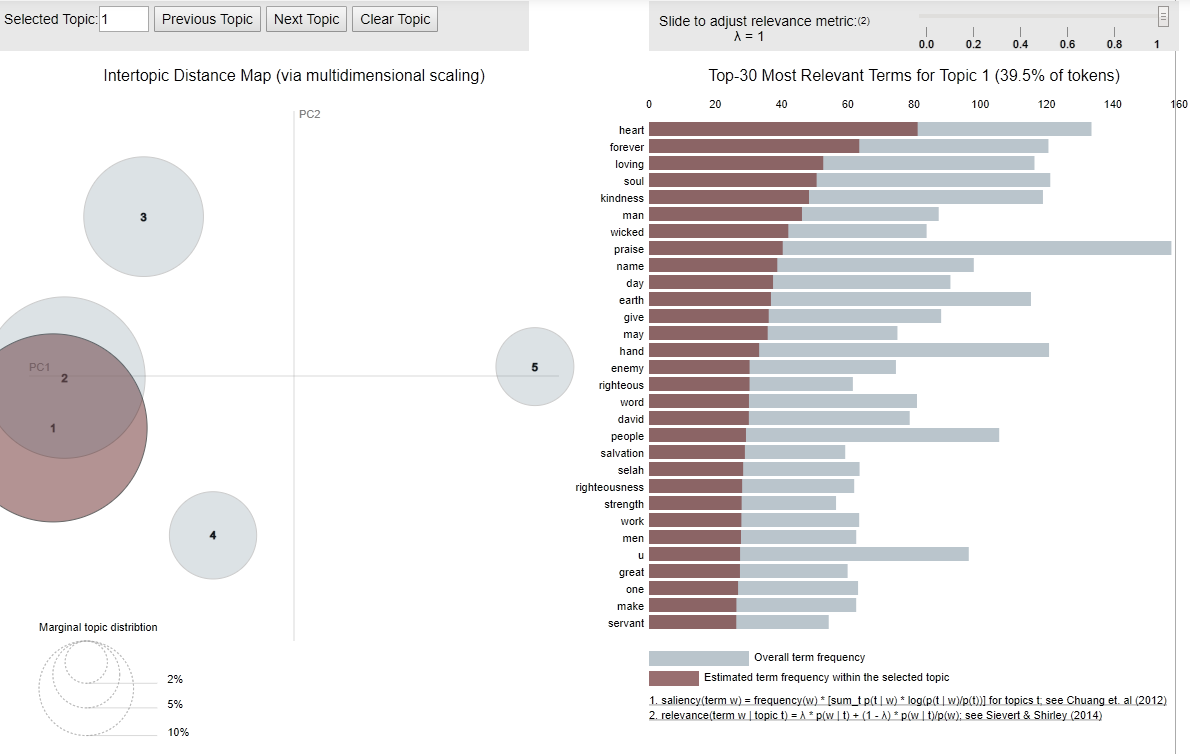
Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. Paramount, CA: CreateSpace.

Weiser, A. (1962). *The Psalms: A Commentary*. (G. E. Wright, J. Bright, J. Barr, & P. Ackroyd, Eds.) (Fifth Revi). Philadelphia: Westminster John Knox Press.

Table 1

*Top 8 words and their weightings for each topic from the LDA model*

|  |  |  |
| --- | --- | --- |
| Topic Number | Weighting | Word |
| 1 | 0.012 | Heart |
|  | 0.010 | Forever |
|  | 0.008 | Loving |
|  | 0.008 | Soul |
|  | 0.007 | Kindness |
|  | 0007 | Man |
|  | 0.006 | Wicked |
|  | 0.006 | Praise |
| 2 | 0.021 | Praise |
|  | 0.012 | Earth |
|  | 0.010 | Us |
|  | 0.009 | People |
|  | 0.009 | Hand |
|  | 0.007 | Name |
|  | 0.006 | Soul |
|  | 0.006 | Day |
| 3 | 0.011 | Us |
|  | 0.008 | People |
|  | 0.008 | Hand |
|  | 0.006 | Day |
|  | 0.006 | Word |
|  | 0.005 | Name |
|  | 0.005 | Enemy |
|  | 0.005 | Heart |
| 4 | 0.011 | Kindness |
|  | 0.010 | Praise |
|  | 0.009 | Soul |
|  | 0.009 | Hand |
|  | 0.009 | Loving |
|  | 0.007 | Forever |
|  | 0.007 | People |
|  | 0.007 | Heart |
| 5 | 0.007 | Hand |
|  | 0.006 | People |
|  | 0.006 | Name |
|  | 0.006 | Earth |
|  | 0.005 | Nation |
|  | 0.005 | Forever |
|  | 0.005 | Enemy |
|  | 0.005 | Israel |

****

*Figure 1*. Output from pyLDAvis