



Abstract

This poster explores the differences between the Bible and Psalter translations of the Psalms, both derived from the Greek Septuagint but serving distinct purposes. Using machine learning techniques like Term Frequency-Inverse Document Frequency (TF-IDF) and Principal Component Analysis (PCA), we analyze the thematic and linguistic variations between the two. Bible Psalms focus on doctrinal themes like salvation, while Psalter Psalms are more poetic and suited for liturgical use. This study highlights how translation purpose influences engagement with the Psalms, offering insights into their role in Orthodox Christian tradition.

Data

This study is only focusing on Psalms from the Christian Orthodox Church. One of them comes from The Book of Psalms within the Orthodox Study Bible, and the other set of Psalms comes from *The Psalter According to the Seventy*.

Psalms vs. Psalter

The Orthodox Church is built on tradition established by Jesus Christ, the Apostles and the Church Fathers. The Orthodox Strives maintain this tradition and any modifications or adjustments can lead us further away from this reality. Because the Orthodox Church follows the liturgical and scriptural tradition of the early Church, the Psalter remains central to the prayers and services. The Psalms form the foundation of The Orthodox Study Bible also has commentary from the early Church Fathers and a perspective rooted in the Orthodox tradition, aiming for a deeper understanding and living of faith.

Text Analysis

Term Frequency-Inverse Document Frequency (TF-IDF) represents two different calculations that, when combined, result in a score for each word in a given document within a set of documents. The entire calculation is composed of the following:

$$TF-IDF = Term\ Frequency(tf) \times Inverse\ Document\ Frequency$$

Term Frequency measures the occurrence of a word within a document. The equation to calculate this is:

$$TF(t, d) = \frac{\text{Count of term } t \text{ in document } d}{\text{Total number of words in } d}$$

Inverse Document Frequency measures how important a term is across a collection of documents. It is calculated as:

$$IDF(t)=\log\left(\frac{N}{DF(t)}\right)$$

Where:

N is the number of documents in the collection

$DF(t)$ is the number of documents that therm t is in.

Dimensionality Reduction

Principal Component Analysis (PCA) simplifies large datasets for easier analysis. PCA retains essential data while reducing redundancy.

The First Principal Component (PC1) is the direction along which the data has the highest variance. It represents the most significant relationship between the features in the Psalms.

The Second Principal Component (PC2) is the direction orthogonal (perpendicular) to PC1 and represents the second-highest variance. It captures the second-most important relationship between the Psalms.

Understanding the Graphs

The graphs below represent the TF_IDF Matrix after it has been transformed by Principal Component Analysis. Each point on the graph represents a single Psalms from either the Bible (blue) or the Psalter (red). The highlights points indicate the top 5 results for a specific query. Each document is plotted according to its PC1 and PC2 scores. *Figure 1* illustrates the search results for a highly liturgical query—an exact quote from the Divine Liturgy—which yields results exclusively from the Psalter. In contrast, *Figure 2* depicts the results of a less liturgical query, resembling language used in a Bible study or everyday conversation outside of a worship service- offering a majority of results from, the bible.

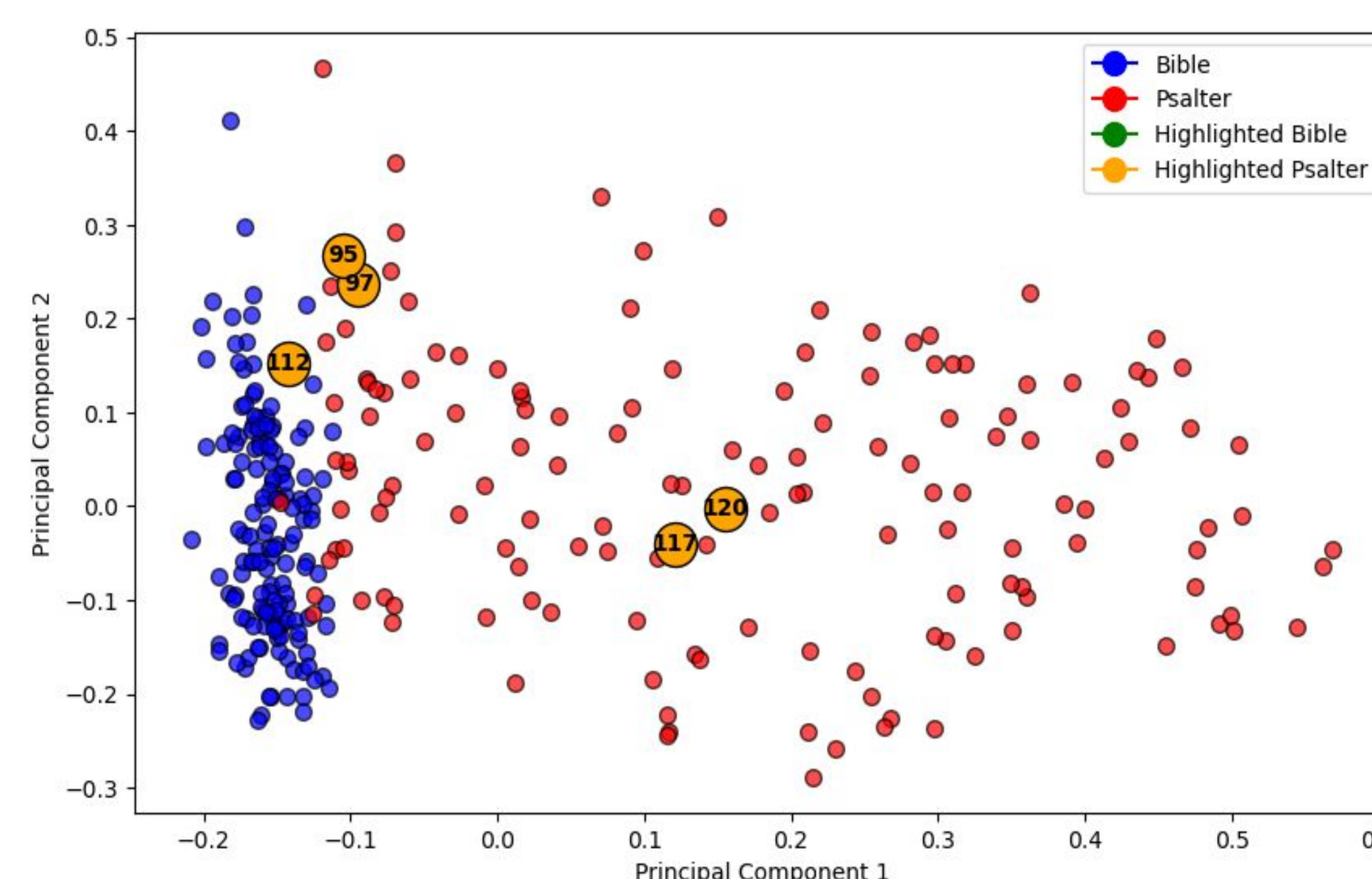


Figure 1. **Searching For:** Blessed is he that cometh in the Name of the Lord

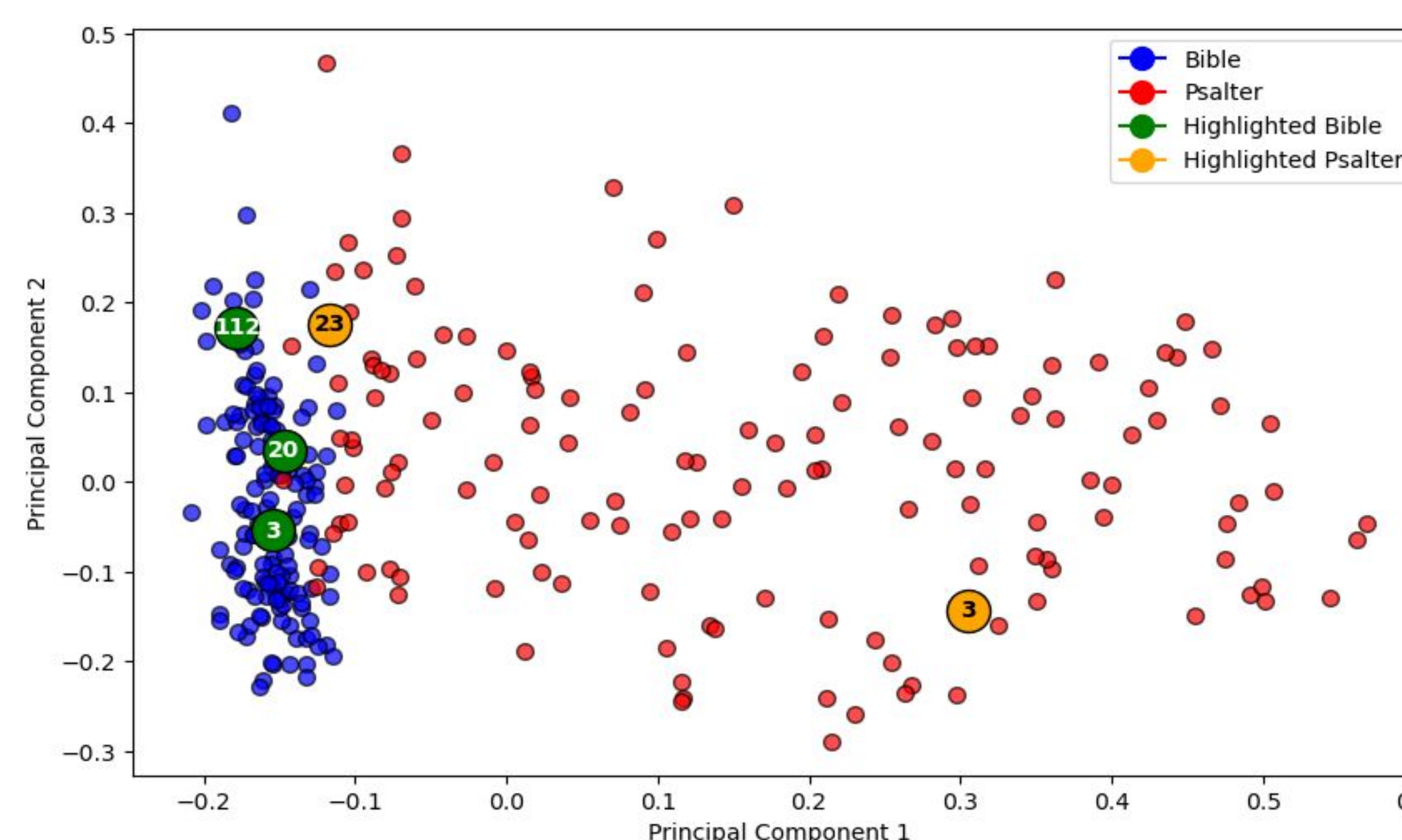


Figure 2. Searching For: Those who come in the Name of the Lord are granted blessings

Word Frequency



Results

The graphs visually highlight key differences in word choice and semantics, illustrating the distinction between the Book of Psalms and the Psalter. Psalms from the Psalter appear more widely dispersed, indicating greater variation compared to those from the Bible. Since both texts originate from the Greek Septuagint, this suggests that translation choices played a significant role in shaping their differences.

The primary cause of this variation lies in the Psalter liturgical function. Its translation prioritizes preserving the exact wording of the Greek Septuagint to maintain the tradition of the Faith. This is evident in Figure 1, where a query taken directly from the Divine Liturgy yields top results exclusively from the Psalter. This consistency reflects the need to adhere closely to the received tradition, minimizing deviation from the Faith passed down through Christ.

Conversely, the Book of Psalms was translated with a different purpose—study and reflection outside of liturgical services, whether for personal devotion or instruction. Here, the priority is accessibility rather than strict adherence to the Septuagint. This contrast is clear in Figure 2: when the query is reworded for study rather than liturgy, the majority of results come from the Bible rather than the Psalter.

References

- Géron, Aurélien. *Hands-on Machine Learning With Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2nd ed. Vol. Dimensionality Reduction: PCA. O'Reilly, 2019.
- Jurafsky, Dan, and James H. Martin. "Question Answering and Summarization." In *Speech and Language Processing*, Vol. Chapter 23: Question Answering and Summarization. Prentice Hall, 2010.
<https://github.com/rain1024/slp2-pdf>.
- Scott, William. "TF-IDF From Scratch in Python on a Real-world Dataset." *Medium*, December 7, 2021.
<https://medium.com/towards-data-science/tf-idf-for-document-ranking-from-scratch-in-python-on-real-world-dataset-796d339a4089>.
- The Orthodox Study Bible*. Thomas Nelson Publishers, 2008.
- The Psalter: According to the Seventy*. Holy Transfiguration Monastery, 1974).