

Psalms-CCLI Comparison via Natural Language Processing

Loc Nguyen, Nate Chu, Roshani Gladson, Joobong Bae, Chris Jaus, Jay Hamilton

Abstract and/or Background

Background: The Psalms is a body of worship given to us by God, so comparing the Psalms to the CCLI top 100 will reveal differences between Psalms and how we worship nowadays. We hypothesize that the Psalms, as a divinely inspired collection, show greater lexical diversity, emotional depth, and complexity in worship compared to the more streamlined expressions in CCLI music.

Methods: We focused on the texts' linguistic and theological characteristics. We leveraged Stanford's CoreNLP AI to process 150 Psalms and the CCLI Top 100 songs across six analytical dimensions: Lexical Analysis, Sentiment Analysis, Named Entity Recognition (NER), Parts-of-Speech (POS) Tagging, Repetitiveness (N-grams), and Coreference Resolution. All texts were lemmatized and cleansed of unwanted words. Metrics such as word count, type-token ratio (TTR), word frequency distributions, sentiment word clouds, dependency distance, and entity mentions were computed to enable a systematic comparison.

Results: Although the Psalms contain over three times the words of the songs, they show higher lexical diversity and more frequent references to God. The Psalms express a wider range of sentiments—including lament, fear, and anger—while the songs lean toward personal, positive themes. Additionally, the Psalms offer a richer array of historical and geographical references, more complex syntax, and a formal depiction of God, as evidenced by n-gram and coreference analyses.

Conclusions: The Psalms demonstrate a richer, more theologically varied mode of worship, while CCLI songs reflect a contemporary shift toward relational and simplified expressions. This research highlights significant linguistic and theological changes in worship across time.

Introduction and/or Research Question

Research Question: What significant linguistic and theological differences are there between the Psalms (ESV) and modern Top 100 CCLI songs?

The following research hypotheses were established as baseline to answer this question:

Lexical Analysis: Compare the diversity and richness of vocabulary used in the Psalms versus CCLI songs.

- **H1**: The Psalms will have greater lexical diversity than the CCLI Top 100 due to the wide range of topics (lament, praise, theology).
- H2: CCLI songs may have a smaller, more repetitive vocabulary, focusing on simpler language and a narrower set of themes (e.g., praise, love, joy).

Sentiment Analysis: Assess the emotional tone and diversity of sentiment in both

- **H3**: The Psalms will display a greater range of sentiments (including lament, fear, anger, joy, praise)
- **H4**: CCLI songs will likely be skewed towards positive sentiment, focusing more on joy, praise, and love, with fewer expressions of negative emotions (like despair or anger).

Theological Entities and Named Entity Recognition (NER): Compare how often and in what way both the Psalms and CCLI songs refer to theological entities.

- H5: The Psalms will have more diverse and frequent references to Old Testament figures, places, and events (e.g., Zion, David).
- **H6**: CCLI songs will focus more on New Testament themes, with Jesus as the central figure, and may have fewer geographical or historical references.

Parts-of-Speech (POS) Tagging and Grammatical Structure: Analyze the grammatical structure of lyrics to assess complexity and the roles of different parts of speech.

- H7: The Psalms will use a more varied and complex sentence structure, reflecting the
- poetic style and depth of the original texts. **H8**: CCLI songs will have simpler sentence structures, with more repetition of certain grammatical forms (e.g., verbs of praise, adjectives related to God's attributes like "good," "great," "holy").

Repetitiveness (N-grams Analysis): Identify repetitive phrases or concepts in both datasets.

- **H9**: CCLI songs will show more repetition, with phrases like "I worship," "praise," or "God is good" recurring frequently.
- H10: The Psalms will have less direct repetition, focusing instead on thematic and structural variation.

Coreference Resolution: Explore how consistently both the Psalms and CCLI songs refer to God or theological entities, such as whether CCLI songs focus predominantly on personal or relational language (e.g., "You" or "He") versus the formal references in the Psalms.

- H11: The Psalms will use more formal titles for God (e.g., "The Lord," "Almighty"), while CCLI songs will use more intimate and personal language (e.g., "You").
- H12: Coreference patterns in the Psalms will reflect a deeper exploration of God's various attributes (e.g., protector, judge, redeemer), whereas CCLI songs might focus more narrowly on relational or worship-based titles.

1. Lexical Analysis

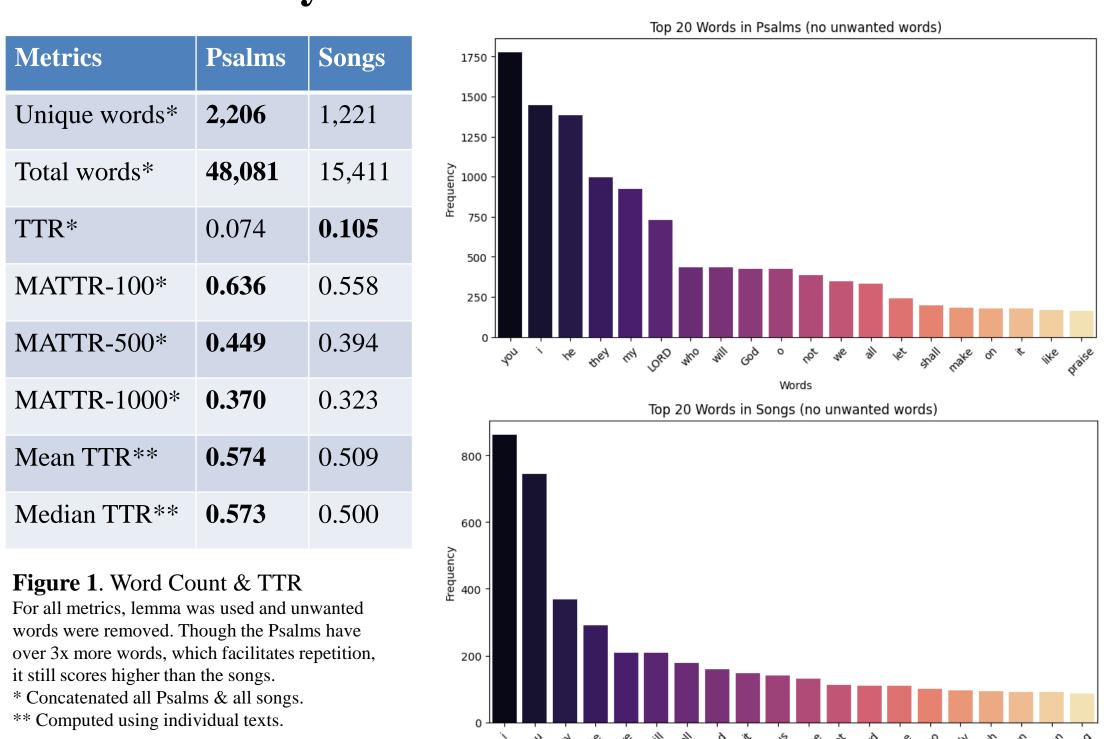


Figure 2. Word Frequency Analysis The Psalms refers to God more frequently than the first-person self, while the songs shows the opposite trend.

2. Sentiment Analysis

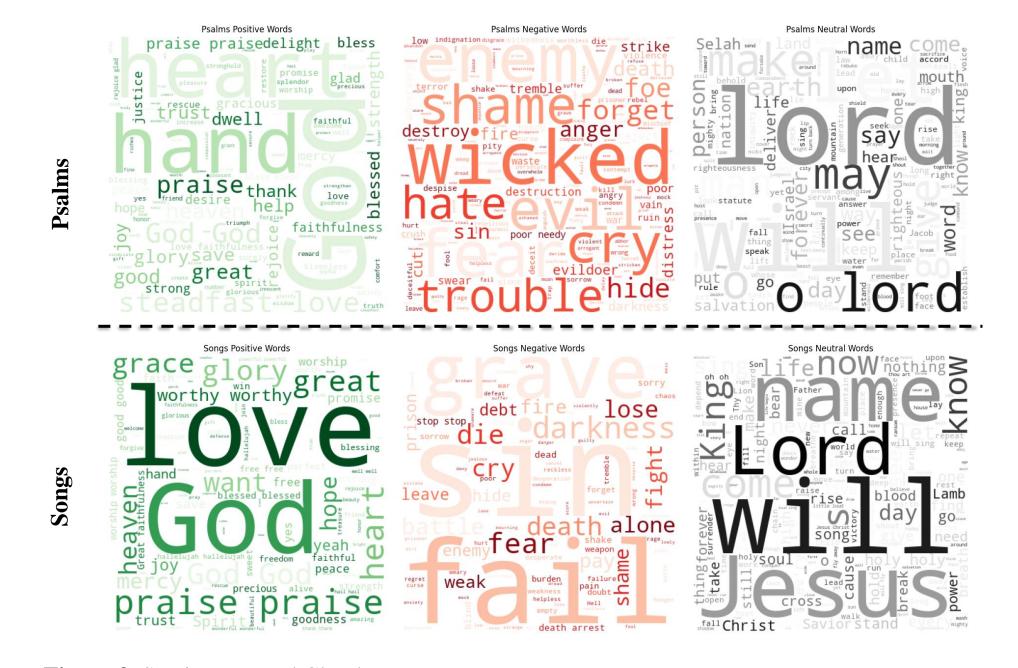
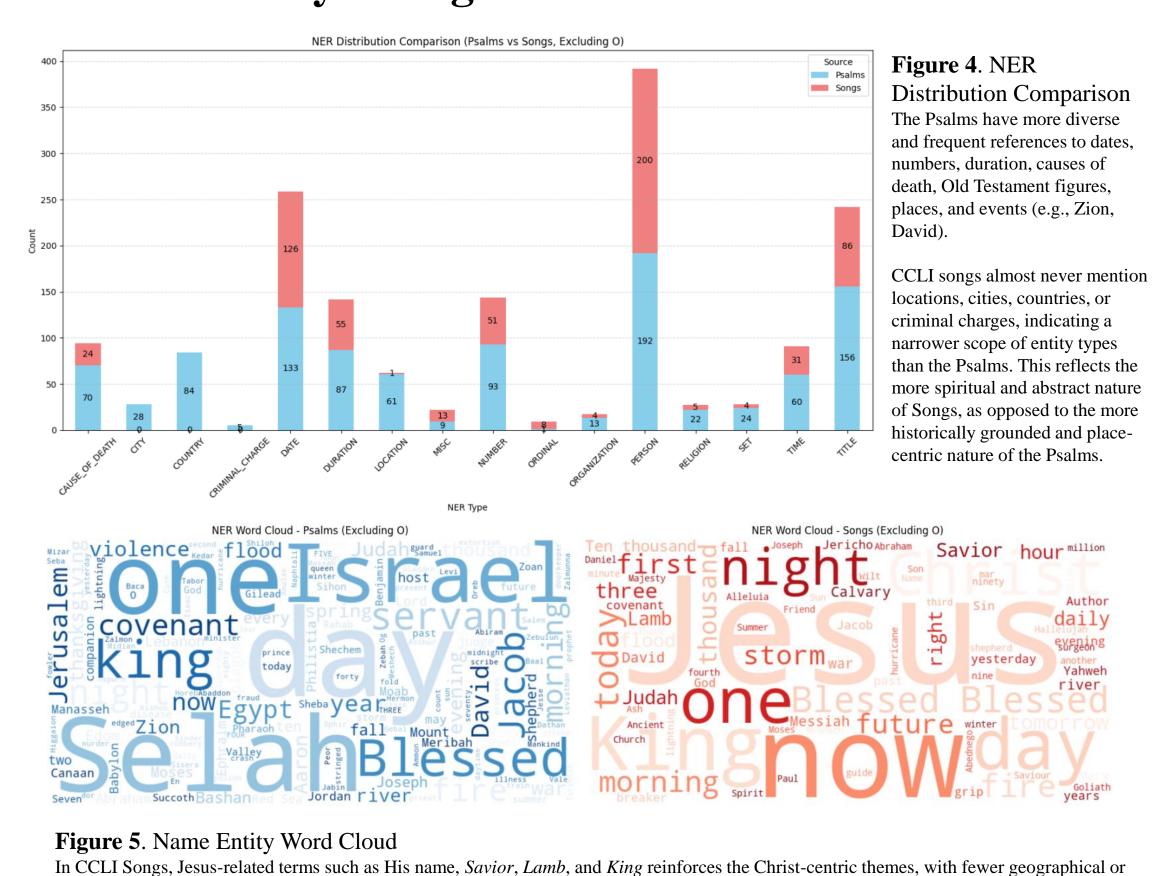


Figure 3. Sentiment Word Cloud In the songs, frequent words like *love*, *heaven*, *grave*, *sin*, and *fail* carry general themes of divine love, redemption, and human frailty. In the Psalms, major mentions of hand, heart, God, enemy, and wicked are specific and emphasizes themes of outward (hand) through prayer or service and inward (heart) worship through sincerity, devotion, intimate connection with God.

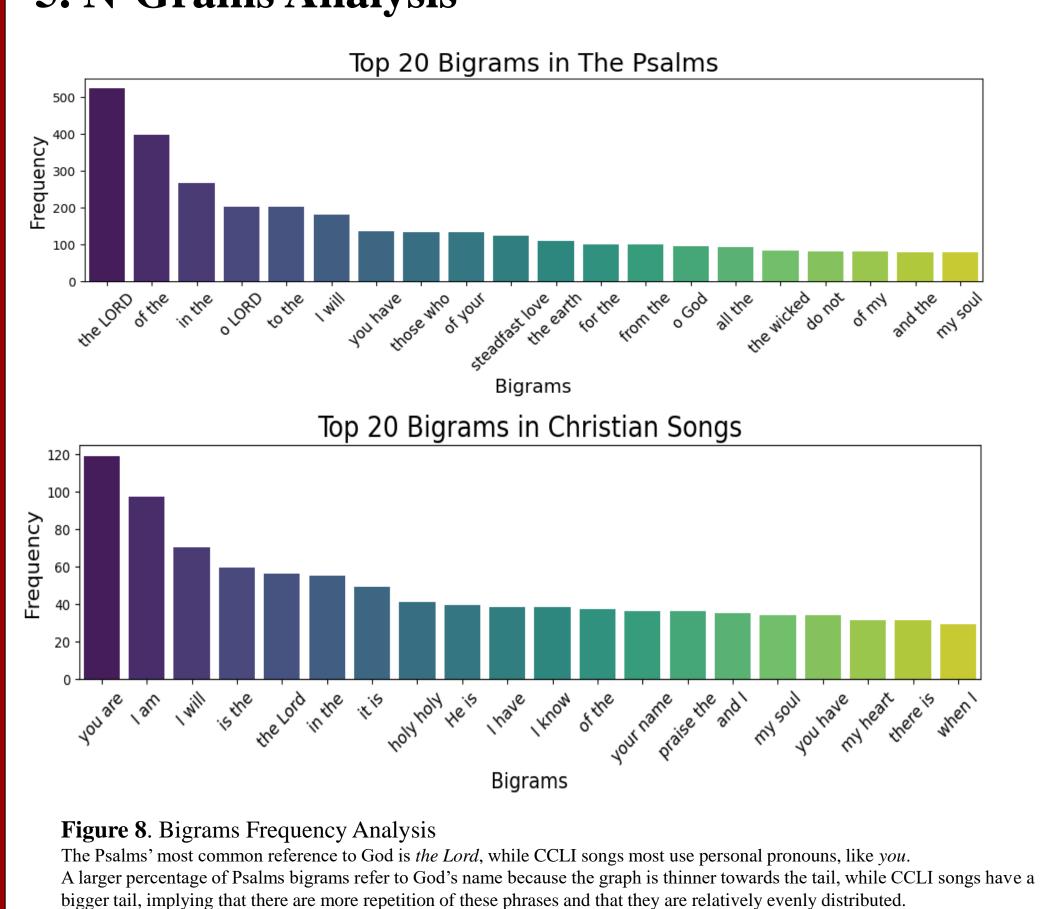
In the Psalms, the negative words cloud is denser than in the songs, meaning Psalmists explore a greater range of sentiments,

3. Name Entity Recognition



historical references than the Psalms. The major words like *Jesus*, *King*, and *now* reflect a focus on divine kingship of Jesus & present-day faith.

5. N-Grams Analysis



4. Part of Speech Tagging

Psalms		Songs			
POS Tag	Frequency	Normalized (per 1k)	POS Tag	Frequency	Normalized (per 1k)
NN	6015	125.10	NN	6015	125.10
IN	5839	121.44	IN	5839	121.44
PRP	4266	88.73	PRP	4266	88.73
DT	3702	76.99	DT	3702	76.99
PRP\$	2914	60.61	PRP\$	2914	60.61
,	2816	58.57	,	2816	58.57
VB	2595	53.97	VB	2595	53.97
•	2445	50.85	•	2445	50.85
NNS	2245	46.69	NNS	2245	46.69
VBP	1676	34.86	VBP	1676	34.86

Figure 6 POS Tagging Frequency Analysis The Psalms uses a similar array of high-frequency tags but with heavier emphasis on prepositions and punctuation, aligning with more formal, clause-rich statements. CCLI Songs focuses comparatively more on direct personal pronouns (PRP) and present-tense verbs (VB, VBP). The

This measures how far apart (in terms of token indices) a governor and its dependent are on average in a sentence. The large difference in values implies that, on average, related words in Psalms are

be tagged as NN ("Lord," "God").

presence of NNP in All Songs could reflect references to personal names or cultural elements, whereas Psalms references to the Divine might often **Figure 7**. Mean Dependency Distance (MDD)

> farther apart, creating longer or more complex syntactic structures. The Psalms likely employs more prepositional phrases, embedded clauses, or formal rhetorical constructs, causing words that depend on each other to span greater distances. In contrast, the simpler syntax of CCLI songs (e.g., short, direct lines like "I love you," "You are mine") keeps

6. Coreference Resolution

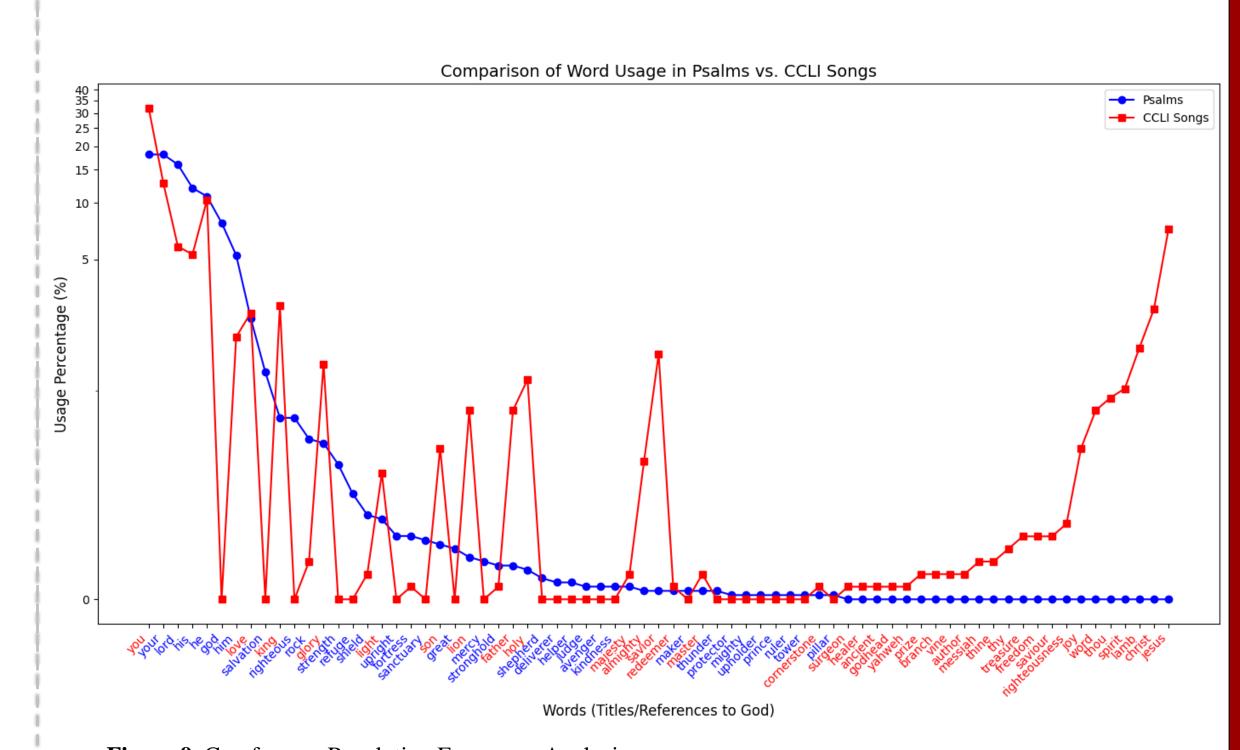


Figure 9. Coreference Resolution Frequency Analysis The Psalms explore more complex and vivid attributes of God, described using nature or battle-inspired imagery, like refuge, fortress, stronghold, ruler, and upholder. Meanwhile, the songs focuses more on characteristics of God as laid out in the New Testament, like Jesus, Christ, light, vine, lamb, and messiah.

Methods

We began by establishing clear, testable hypotheses drawn from a deep dive into existing literature on computational linguistics, digital humanities, and theology. Our goal was to compare the language and sentiment of the Psalms with modern CCLI songs, focusing on aspects like lexical diversity, emotional range, historical and geographical references, and syntactic complexity. Using Stanford's CoreNLP, we processed a corpus comprising 150 Psalms and the CCLI Top 100 songs. I developed a Python pipeline that automated text preprocessing lemmatizing words and filtering out unwanted terms—before running the data through CoreNLP's suite of tools. Specifically, we extracted metrics for word count, type-token ratio, word frequency, sentiment scores, named entities, parts-of-speech tags, dependency parse trees, and n-gram frequencies. Each measure was carefully chosen to test our hypotheses; for example, lexical diversity was evaluated using typetoken ratios, while sentiment was analyzed through CoreNLP's recursive neural network model. We also examined syntactic complexity using mean dependency distance and POS tagging, and identified repetitive patterns via bigram analysis and coreference resolution. Our approach combined rigorous statistical tests with visualizations—like word clouds and histograms—to provide both quantitative and qualitative insights. This comprehensive, mixed-methods strategy allowed us to robustly assess the nuanced differences between the ancient Psalms and modern CCLI worship songs.

Results and/or Conclusion

Results: Even though the Psalms have over three times the total words compared to the CCLI songs, they still show higher lexical diversity. This confirms our expectation that the Psalms use a richer vocabulary despite having more repetition potential. In our word frequency analysis, the Psalms consistently refer to God more often, while the CCLI songs focus on first-person expressions—suggesting a shift toward a more personal worship style today. The sentiment analysis backs this up: the Psalms cover a broader emotional range (including lament, fear, and anger), whereas the songs tend to be overwhelmingly positive. Our Named Entity Recognition results reveal that the Psalms include a wide array of historical and geographical references, like dates, places, and figures such as Zion and David, which contrasts sharply with the more abstract, Christ-centric language found in the songs. When looking at the grammatical side, the POS tagging and mean dependency distance show that the Psalms use more complex, clause-rich sentences, while the songs lean toward simpler structures. Lastly, the bigram analysis and coreference resolution indicate that the Psalms depict God in a more formal and multifaceted way, whereas the songs use more repetitive and relational language. Overall, these findings support our initial hypotheses and highlight clear linguistic and thematic differences between the ancient Psalms and modern CCLI worship songs.

Future Work

Future works can focus on the following areas that our current research did not yet

- For analysis, we used Psalm excerpts from the English Bible in ESV. Other languages, including the original language, was not considered. More detailed and insightful analysis could stem from researching a version of the text that is closer to the source, like Hebrew manuscripts.
- Future research can consider more facets of analysis, such as Theological Richness Goal, which measure the depth and variety of theological terms and concepts in both texts. The Psalms will likely show richer theological vocabulary, exploring complex topics like atonement, justice, divine wrath, and covenant, while CCLI songs will focus more on relational and affective terms (love, peace, joy), with less exploration of complex theological ideas.

References and/or Acknowledgments

English Standard Version Bible. (2001). ESV Online.

https://esv.literalword.com/

CCLI. (2023). Christian copyright licensing international: License agreement. https://www.ccli.com/license/