# Detecting Narrative Patterns in Biblical Hebrew and Greek

**Hope McGovern**[1] **Hale Sirin**[2]
**Tom Lippincott**[2] **Andrew Caines**[1]
[1] Department of Computer Science & Technology, University of Cambridge, U.K.
[2] Center for Digital Humanities, Johns Hopkins University, Baltimore, U.S.A.
[1] firstname.secondname@cl.cam.ac.uk [2] {hsirin1, tom.lippincott}@jhu.edu

## Abstract

We present a novel approach to extracting recurring narrative patterns, or *type-scenes*, in Biblical Hebrew and Biblical Greek with an information retrieval network. We use cross-references to train an encoder model to create similar representations for verses linked by a cross-reference. We then query our trained model with phrases informed by humanities scholarship and designed to elicit particular kinds of narrative scenes. Our models can surface relevant instances in the top-10 ranked candidates in many cases. Through manual error analysis and discussion, we address the limitations and challenges inherent in our approach. Our findings contribute to the field of Biblical scholarship by offering a new perspective on narrative analysis within ancient texts, and to computational modeling of narrative with a genre-agnostic approach for pattern-finding in long, literary texts.

## 1 Introduction

In this study, we present a new approach to extracting recurring narrative patterns, or *type-scenes*, in Biblical Hebrew and Biblical Greek.

The term type-scene [*typische Szene*] was coined in the Homeric scholarship to refer to the formulaic and recurring narrative patterns or scenes, such as the arming or sacrifice scenes in the Iliad (Arend, 1933). This term was later adapted to refer to the phenomenon in Biblical narratives of repeated use of similar patterns to describe a narrative event (Alter, 1978).[1]

We pose type-scene extraction as an information retrieval task, where one verse (a query) and its surrounding context can be used to retrieve thematically similar verses (candidates). Cross-references provide structured information about linked passages, whether through direct text reuse, thematic parallelism, or common characters. Learning representations that maximize the similarity of cross-referenced passages may enable us to surface other forms of narrative connection, i.e. type-scene. In the context of the Bible, an example of such a recurrent pattern would be the story of a rivalry between "a barren but favored wife" and "a fertile concubine" (e.g. Sarah-Hagar, Rachel-Bilhah and Leah-Zilpah) (Alter, 1978).

Type-scenes connect to a larger body of literary scholarship concerned with narrative patterns in the form of recurrent structures that could be instantiated by a variety of different characters and settings, such as a hero is tempted. In literary theory, these include *dramatis personae* and their functions in folktales (Propp, 1968) or the concept of *figura* that refers to similarities in the structure of events or in the circumstances that accompany them in Biblical stories (Auerbach, 1959).

In this study, we focus on the Bible since it offers a rich example to study recurrent narrative structures because of its heavily self-referential composition; however, the model we propose can be applied to any long-text genre where repeating narrative patterns may be of interest, such as Shakespeare's plays, Greek tragedy, even movie scripts. We show that a network fine-tuned on a set of Bible cross-references improves recall@$k$ over its respective pre-trained models. Although recall values remain low in real terms, our models retrieve relevant candidates in the top-10 results for each type-scene query, showing that this approach holds great promise as a tool for Biblical scholars to uncover recurrent narrative patterns in the Bible.

## 2 Related Works

Recently, studies have emphasized the need to combine theoretical work on narrative within the humanities and computational methods for reason-

[1] For example, Alter writes, "[t]hree times a patriarch is driven by famine to a southern region where he pretends that his wife is his sister, narrowly avoids a violation of the conjugal bond by the local ruler, and is sent away with gifts. Twice Hagar flees into the wilderness from Sarah's hostility and discovers a miraculous well."



*Proceedings of the 1st Workshop on Machine Learning for Ancient Languages (ML4AL 2024)*, pages 269–279
August 15, 2024 

| Type-Scene Description | Example | Scholarly Support |
|---|---|---|
| 1. A man meets a woman by a well | Genesis 24:11 | Alter (1978) |
| 2. An angel announces a barren woman will bear a child | Judges 13:3 | Alter (1983) |
| 3. The younger son is preferred to the older son | Malachi 1:2-3 | Davies (1995) |
| 4. God sends a prophet to speak to a rebellious people | Jonah 1:1-3 | Long (1973) |
| 5. A man has an epiphany in the field | Ezekiel 1:1 | Alter (1978) |
| 6. A prophet veils his face | Exodus 34:33 | Britt (2002) |
| 7. A well is found in the desert | Genesis 16:7 | Alter (1978) |
| 8. The words of a dying hero | Genesis 50:25 | Alter (1978) |
| 9. A hero is tempted | Judges 16:6 | Alter (1978) |

Table 1: **Type Scene Queries**. We craft a set of queries designed to elicit type-scenes as postulated by scholars in the humanities. Here we show an English version of the query (N.B. we must translate it into our model's language before encoding it), an example in the text of where it occurs, and scholarly support for the query being a repetitious narrative element in the text.

ing about narrative, such as information retrieval, summarization, commonsense inference, and event detection, in order to advance our understanding of narrative and open up new practical applications (Piper et al., 2021). Specifically, Finlayson (2009) adapts Bayesian model merging to derive narrative morphologies by merging analogical stories in a corpus of Shakespearean plays. Reasoning through analogies between seemingly different but related constructions, ("the virus invades cells" v.s. "the burglar breaks into the house") remains a challenging problem (Jiayang et al., 2023). Type-scene detection formulated in this paper can contribute to analogical narrative reasoning in natural language processing.

**RELiC** (Thai et al., 2022): We draw inspiration from RELiC's methodology, which employs two separate encoder networks to embed primary source quotations (e.g. taken from "Pride and Prejudice") and quoting passages in the scholarly literature (e.g. commentaries on Jane Austen's work). Their approach involves minimizing a contrastive loss function with in-batch negative sampling. Notably, RELiC's work demonstrates the effectiveness of crafting out-of-domain queries, such as 'Elizabeth is frustrated with her mother,' and finding sometimes useful results. Building on this, we explore intra-Biblical allusions through the lens of type-scenes, adapting their approach to the unique context of Biblical narratives.

**Document Similarity for Information Retrieval**: Another relevant area of study is document similarity in information retrieval (IR) systems. One approach from Ostendorff et al. (2020) focuses on similarity within structured documents, creating an IR system for academic papers using citations in specific sections as signals. In this framework, the section title in which a citation occurs serves as a label for the pair of citing and cited papers, framing the task as a pairwise document classification problem. Although our focus is on narrative analysis within the Bible, the structured nature of the Bible with its imposed chapter and verse divisions renders our task conceptually similar to aspect-based document similarity in IR systems.

| | Ancient Hebrew | Ancient Greek |
|---|---|---|
| verses | 23,275 | 31,227 |
| chapters | 929 | 1132 |
| books | 39 | 64 |
| ref pairs | 17,899 | 45,297 |

Table 2: **Dataset Statistics.** We use two ancient manuscripts, one in Ancient Hebrew covering the Hebrew Bible (without the Apocrypha), and one in Ancient Greek including the Hebrew Bible and Christian New Testament. We report statistics about each manuscript as well as the number of cross-reference verse pairs in each using a vote threshold value of 5.

## 3 Data

**Biblical Manuscripts:** Our primary data source is a digitized version of the Leningrad Codex[2], the oldest complete extant manuscript of the Hebrew Bible (HB) in Ancient Hebrew. We obtain a copy of this in XML format, which we process to extract the text of every verse[3].

We additionally use an open-source Ancient Greek edition of the Bible, containing both the Hebrew Bible and the Christian New Testament

[2]Source: https://www.tanach.us/Pages/Technical.html

[3]Note that verse delineations were not added to copies of the text until the Middle Ages. See Appendix A for more information on versification.

(NT). This edition uses the Septuagint for the Hebrew Bible in Greek and the Society of Biblical Literature's Greek New Testament (SBLGNT)[4].

**Cross References:** We use a file of cross-references found on `openbible.info` originally created via crowd-sourcing. Each cross-reference contains a source and target verse and the number of crowd-sourced votes affirming this connection. We assume that the target verse is always earlier in the text according to the modern standard English Bible ordering, even though this may not be valid for historical reasons[5], and separate one-to-many verses into individual examples. Figure 5 in the Appendix shows a histogram of the cross-reference votes; their distribution is highly left-skewed, with most pairs amassing fewer than 3 votes. We keep all references with at least 5 votes, which filters out potentially spurious data points without overly pruning the dataset.

The original file contains nearly 350,000 cross-references; after filtering for our vote threshold and removing pairs for which either the source or the target does not appear in a given manuscript, we obtain 45,297 Ancient Greek pairs and 17,899 Ancient Hebrew pairs. See Table 2 for relevant dataset statistics.

For ablation experiments in section 5, we increase this vote threshold to 50, limiting the data to around 700 pairs across both the HB and NT, and around 400 within just the HB.

**Type Scene Queries:** To test whether our model can retrieve type-scenes, we craft a small set of queries designed to elicit typological narratives, e.g. "A man meets a woman by a well". A language expert helped us to craft the queries in Biblical Hebrew, and we used ChatGPT[6] to obtain translations in Biblical (Koine) Greek. These were then manually verified. Each type-scene for which we write a query is supported by scholarly work on textual criticism of the Bible[7]. The full list of the queries we use is seen in Table 1, and the translations into Ancient languages may be seen in Table 5 and Table 6 in the Appendix.

### 3.1 Preprocessing

The digitized version of the Leningrad Codex already contains some morphological segmentation, separating definite articles and function words which, in Hebrew, become preprended to the subsequent word (e.g. 'in', 'to', 'and'). We remove this separation and also remove cantillation marks, but we leave vowel markings (niqqud) intact. We also remove ellipses denoting missing text in the physical codex. The only text cleaning we perform for the Greek Bible is lowercasing and removing diacritics, to match the preprocessing methods of the models we use.

Once cleaned, we extract a window of $n$ verses on either side of the referencing verse. We analyze the impact of the context window size in Figure 4.

Finally, we separate our data into training and validation sets with a 90/10 split. Our test set is a manually created set of type-scene queries, which have been preprocessed in the same way as the training examples.

## 4 Method

**Problem Formulation**: Our study addresses the task of identifying the most relevant candidate passage within a set of candidates (the text of the Bible) given a query passage. Specifically, we are provided with a set of query vectors and candidate vectors. Query vectors represent passages consisting of $n$ verses surrounding a verse referencing an earlier passage, while candidate vectors represent passages of $m$ verses. The goal of our model is to discern the true candidate passage from this set of candidates.

**Models**: We create a Siamese network instantiated with a pre-trained BERT-like transformer model[8]. The Siamese network includes separate encoders for the queries and the candidates, which are then jointly trained with a contrastive loss function to push learned representations for queries and their ground-truth candidates closer together in latent space (and incorrect candidates further away). Intuitively, after training in this manner, a simple dot product of a given query with all possible candidates will surface the true candidates. This approach is adapted from that of Thai et al.

[4]Source: `https://github.com/LukeSmithxyz/grb`

[5]The assembly process of the books of the Bible is a complex and often contested area, but we can be certain that books are not always arranged in order of date of writing. So, a book early in the order of the Bible (written late) may in fact reference a later book (written earlier).

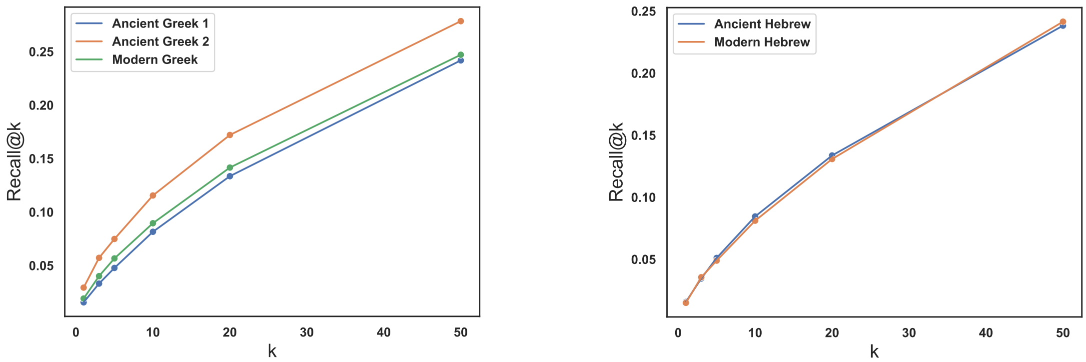
[6]`https://openai.com/chat/`

[7]We take liberties with the wording of the type scene. For example, Alter (1978) identifies, "the initiatory trial" as a type-scene, which we expect will not return good search results as it is highly vague. We rephrase it as "a hero is tempted". We find that even this is too vague to return many relevant passages

[8]All models use the Huggingface `Transformers` library (Wolf et al., 2020).

| Base Model | Language | Fine-tuning | R@1 | R@3 | R@5 | R@10 | R@20 | R@50 |
|---|---|---|---|---|---|---|---|---|
| Ancient Hebrew (Fono et al., 2024) | hbo | None | 0.01 | 0.02 | 0.03 | 0.05 | 0.08 | 0.14 |
| | | Leningrad | 0.02 | 0.03 | 0.05 | 0.08 | 0.13 | 0.24 |
| Modern Hebrew (Shmidman et al., 2023) | heb | None | 0.03 | 0.04 | 0.06 | 0.08 | 0.12 | 0.20 |
| | | Leningrad | 0.01 | 0.04 | 0.05 | 0.08 | 0.13 | 0.24 |
| Ancient Greek 1 (Yamshchikov et al., 2022) | grc | None | 0.02 | 0.04 | 0.05 | 0.06 | 0.09 | 0.15 |
| | | SEPT+SBLGNT | 0.02 | 0.03 | 0.05 | 0.08 | 0.13 | 0.24 |
| Ancient Greek 2 (Singh et al., 2021) | grc | None | 0.03 | 0.04 | 0.05 | 0.07 | 0.10 | 0.16 |
| | | SEPT+SBLGNT | 0.03 | 0.06 | 0.07 | 0.12 | 0.17 | 0.28 |
| Modern Greek (Koutsikakis et al., 2020) | gre | None | 0.01 | 0.01 | 0.01 | 0.02 | 0.04 | 0.07 |
| | | SEPT+SBLGNT | 0.02 | 0.04 | 0.06 | 0.09 | 0.14 | 0.25 |

Table 3: **Recall@$k$ for correct chapter of true candidate.** For base encoder models for our Siamese network, we include models trained on Ancient Hebrew (hbo), Modern Hebrew (heb), Ancient Greek (grc), and Modern Greek (gre). The 'Manuscript' column indicates whether or not the model was finetuned on a Biblical manuscript before being used to rank all possible candidates (i.e. every verse) in the Bible. We present R@$k$ for $k = 1, 3, 5, 10, 20, 50$ for the model returning the chapter of the true candidate in the top-$k$ ranked candidates. While our contrastive fine-tuning improved R@$k$ in every scenario, the numerical scores are still quite low, indicating that this is a difficult problem (retrieving one verse from $\tilde{3}$0k options) and that our current fine-tuning setup is not fully optimized for the needle-in-a-haystack nature of the problem.




(a) Ancient Greek vs. Modern Greek

(b) Ancient Hebrew vs. Modern Hebrew

Figure 1: **Ancient vs. Modern Language Models**. We investigate whether encoders trained on modern variants of ancient languages will outperform those trained on ancient languages as a result of the considerably larger training data available for modern languages. Figure 1a shows the R@$k$ for two Ancient Greek models compared to their Modern Greek base. Figure 1b shows the R@$k$ for a model trained exclusively on Biblical Hebrew compared to R@$k$ for a model trained on Modern Hebrew. For Hebrew, there is no distinction between using the modern and ancient variants, whereas there is more discrepancy in the Greek results. Although Ancient Greek 1 and 2 are instantiated from the same base Modern Greek model, Ancient Greek 1 outperforms both the Modern Greek and the other Ancient model.

(2022), RELiC, and we modify it as needed for our purposes.

For each language, we experiment with language models trained on either the modern or ancient variant of the language. Intuitively, a language model trained on Ancient Greek should outperform one trained on modern Greek, but there is a sizeable difference in the amount of available training data for ancient languages, which may lead to subpar model performance.

**Training Procedure**: The training of our model is guided by a contrastive loss function, which encourages the model to minimize the distance between embeddings of positive pairs (i.e., matching query-candidate pairs) while maximizing the distance between embeddings of negative pairs.

To generate negative examples for the contrastive loss, we simply choose a random verse, ensuring that it is not the true candidate.

Specifically, we use a triplet margin loss (Balntas et al., 2016) with a margin value of 1.0, Adam optimizer (Kingma and Ba, 2015) with 1$e$-5 initial learning rate, and a mini-batch size of 16. Our models were trained on an A100 NVIDIA GPU.

**Evaluation**: We compute the dot product of a normalized context vector to the candidate em-

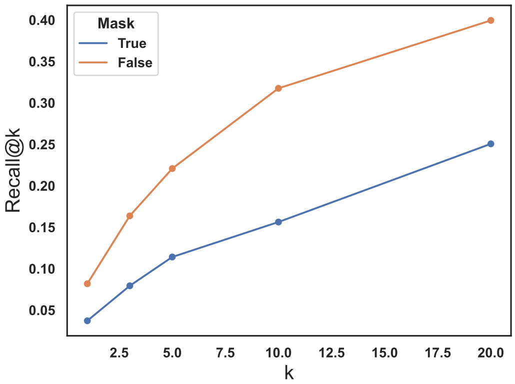



Figure 2: **Masking**. In Thai et al. (2022)'s implementation, the actual query vector is masked out while keeping the context on either side. We experiment with keeping the query vector unmasked and find that this provides a substantial increase in performance over masking.

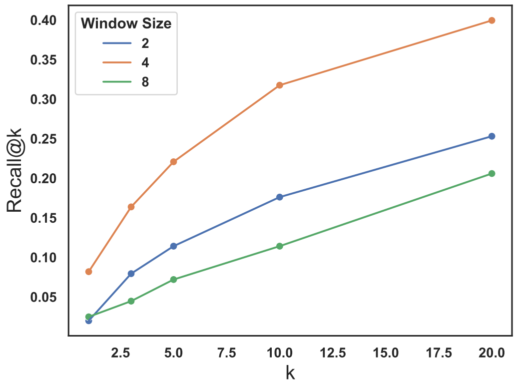



Figure 4: **Context Window Size.** We vary the number of verses on either side of the query verse. N.B. that a window size of 2 indicates 1 verse prior and 1 verse following, 4 indicates 2 verses prior and 2 verses following, etc. We find an optimal window size to be 4.

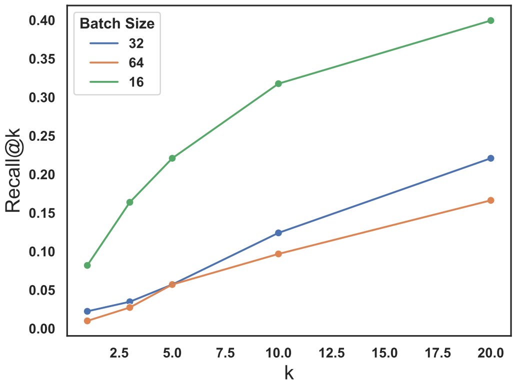



Figure 3: **Training Batch Size.** The effect of training batch size on R@$k$. We find that a smaller batch size of 16 outperforms larger values of 32 and 64.

beddings and sort to obtain candidate rankings. We then evaluate with standard recall@k metrics, where $k$ ranges from 1 to 50, to measure the proportion of relevant candidates identified within the top $k$ retrieved passages.

## 5 Results

We present the recall metrics for both untrained and trained models across Ancient and Modern Greek/Hebrew in Table 3.

We further investigate the performance disparity between Ancient and Modern language models, as illustrated in Figure 1a and Figure 1b. 'Ancient Greek 1' (Yamshchikov et al., 2022) and 'Ancient Greek 2' show a marked difference in their performance despite sharing the same backbone BERT model. 'Ancient Greek 1' is fine-tuned on a corpus of Plutarch texts and data from the Perseus Digital Library[9] and First Thousand Years of Greek[10] and was created to examine authorship identification in Plutarch's works. 'Ancient Greek 2' (Singh et al., 2021) is fine-tuned on a varied corpus of Modern, Ancient, and Post-classical Greek and was intended for downstream use for POS-tagging.

An ablation study on batch size, a critical parameter for contrastive loss functions, is conducted and depicted in Figure 3 We find a batch size of 16 to be optimal while increasing the batch size harms recall metrics. Additionally, we explore the impact of context window size through an ablation study, as shown in Figure 4 Two verses on either side of the query verse (a window size of 4) outperform over both smaller and larger context windows.

We also assess the effect of masking out the context verse, a strategy employed in the original RELiC implementation, on our problem, as presented in Figure 2. Masking has a substantial impact on recall metrics. Not masking the query verse nearly doubles performance over masking.

Finally, we offer selected examples of verses returned by our model for each type-scene query in Table 4. Manual analysis of these results is provided in section 6.

## 6 Analysis

The task of recalling thematically linked narrative passages presents significant challenges. Even when expanding the scope from individual verses to entire chapters, the recall scores remain low. Our

[9]https://github.com/PerseusDL/canonical-greekLit

[10]https://github.com/ThomasK81/TEItoCEX

| Query | Returned Verse | Rank |
|---|---|---|
| A man meets a woman by a well | (2SA.11.2) It happened, late one afternoon, when David arose from his couch and was walking on the roof of the kings house, that he saw from the roof a woman bathing; and the woman was very beautiful. | 3 |
| | (1KI.17.10) So he arose and went to Zarephath. And when he came to the gate of the city, behold, a widow was there gathering sticks. And he called to her and said, "Bring me a little water in a vessel, that I may drink." | 5 |
| | (2SA.17.19) And the woman took and spread a covering over the wells mouth and scattered grain on it, and nothing was known of it. | 7 |
| An angel announces a barren woman will bear a child | (HOS.1.8) When she had weaned No Mercy, she conceived and bore a son. | 1 |
| | (JUD.13.3) And the angel of the LORD appeared to the woman and said to her, Behold, you are barren and have not borne children, but you shall conceive and bear a son. | 2 |
| | (GEN38.27) When the time of her labor came, there were twins in her womb. | 5 |
| The younger son is preferred to the older son | (PRO.17.25) A foolish son is a grief to his father and bitterness to her who bore him. | 1 |
| | (GEN.37.32) And they sent the robe of many colors and brought it to their father and said, "This we have found; please identify whether it is your sons robe or not." | 10 |
| God sends a prophet to speak to a rebellious people | (JOB.38.31) "Can you bind the chains of the Pleiades or loose the cords of Orion? | 1 |
| A man has an epiphany in the field | (ISA.13.6) Wail, for the day of the LORD is near; as destruction from the Almighty it will come! | 2 |
| | (ZEC.2.3) And behold, the angel who talked with me came forward, and another angel came forward to meet him | 8 |
| A prophet covers his face | (JOB.41.6) Will traders bargain over him? Will they divide him up among the merchants? | 1 |
| | (PSA.10.11) He says in his heart, "God has forgotten, he has hidden his face, he will never see it." | 3 |
| A well is found in the desert | (GEN.26.19 ) But when Isaacs servants dug in the valley and found there a well of spring water, | 1 |
| | (GEN.16.7) The angel of the LORD found her by a spring of water in the wilderness, the spring on the way to Shur. | 4 |
| The words of a dying hero | (JOB.9.25) "My days are swifter than a runner; they flee away; they see no good. | 1 |
| A hero is tempted | (JOB.13.8) Will you show partiality toward him? Will you plead the case for God? | 10 |

Table 4: **Selected examples returned for type-scene queries.** Highlighted in green are words and phrases indicating that it is a relevant search result. Highlighted in red are returned verses that seem to have no relation to the query. In general, there tend to be a few relevant results returned in the top 10 for each query, but some queries work better than others, and we find that queries that mention concrete items, e.g. man, woman, desert, etc. fare better on average. N.B. we use English translations from the English Standard Version 2011.

contrastive loss fine-tuning approach has shown promise, yet it requires further optimization to develop a type-scene retrieval network with high recall.

**Insights from ablations.** Contrary to expectations, the performance differences between modern and ancient language models were minimal. Notably, one Ancient Greek model outperformed both the Modern Greek and another Ancient Greek model, potentially due to its training data containing more Septuagint-like Greek examples.

Ablation studies highlighted the importance of factors such as batch size, context window size, and masking. Our primary results were obtained using the optimal batch size (16) and window size (4 verses – 2 on either side), with masking applied.

The key insight from these studies is that retraining the models without masking the query verse in the context could significantly improve performance. This discrepancy likely arises because, unlike the setup in Thai et al. (2022), where a scholarly quote served as context and primary source quotes as candidates, our system uses passages from the same book for both context and candidates.

## 6.1 Type-Scene Retrieval Analysis

**Positive Examples**

*1. The Woman at the Well.* Retrieved verses for this query often involve scenarios featuring men, women, and water, even if the source of water is not necessarily a well. The system highly ranked verses such as 1 Kings 17:10 and 2 Samuel 17:19, both of which depict interactions involving a man and a woman in the vicinity of water or a well.

Interestingly, the system returned the scene of King David observing Bathsheba bathing on a rooftop among the top 3 candidates. This scene is not traditionally connected with other, "woman at the well" type-scenes, yet bears a clear resemblance, underscoring the potential utility of the model in returning unconventional examples for further study. However, the system did not retrieve some of the more obvious type-scenes in this category, such as those involving Jacob and Rachel, Moses and Zipporah, or Isaac and Rebekah. Capturing these more intricate narratives would likely require the system to process larger narrative windows than single verses.

*2. The Miraculous Birth.* Encouragingly, one of the top returned verses for query 2 is a prototypical example of this type-scene: Judges 13:3, where an angel announces the birth of Samson. Additionally, other retrieved verses frequently discuss themes of conception and the birth of a son, often within the context of overcoming barrenness.

*3. The Younger Son is Chosen.* The retrieval system returned numerous verses from Proverbs, which is consistent with the nature of Proverbs as ancient wisdom literature containing didactic proscriptions for healthy family relations.

Notably, an exciting retrieval was the instance of Joseph receiving the many-colored coat from his father, a clear symbol of a younger son being preferred over his older siblings. Additionally, the system retrieved Genesis 27:23 at rank 2, which describes Isaac mistakenly blessing Jacob instead of Esau due to Jacob's disguise. This verse is another significant example of the younger brother gaining prominence over the older brother.

*7. Water in the Desert.* The retrieval system successfully identified two clear examples of finding water in the wilderness. Almost every returned verse mentioned the wilderness, with most also referencing water or wells.

**Negative Examples**

*4. God Sends a Prophet.* The retrieval system failed to return any relevant examples for this type-scene. Interestingly, 8 of the top 10 returned verses were from the Book of Job. While Job is not traditionally considered a prophetic book, its strong negative sentiment may have been misinterpreted by the system as aligning with the concept of "rebellious."

*5. The Epiphany in the Field.* This type-scene poses a significant linguistic challenge. As shown in Table 5, the query translates more closely to "the Lord came upon him in the field." Consequently, it is not surprising that the returned verses predominantly relate to God's judgment. There is one example involving a meeting with angels, which aligns somewhat with the query's intended meaning, but does not include a reference to a field.

*6. The Veiled Prophet.* The system returned verses primarily about covering one's face, mostly from wisdom literature, which include poetic descriptions of feeling rejected by God (e.g., God hiding or turning His face away). These verses lack the appropriate narrative context of the life of a prophet. One expected response might be Moses veiling his face when he descends from Mount Sinai after the giving of the Law.

*9 & 10: The Dying Hero and The Hero's Trial* These type-scenes were particularly challenging due to the lack of a precise word for "hero" (as modern readers understand it) in Ancient Hebrew. The concept is typically represented either by *gibbor* ( גִּבּוֹר), denoting military prowess or a mighty warrior, or *nabi* (נָבִיא), denoting a prophet with divinely appointed power. Given this linguistic limitation, it is unsurprising that the system did not retrieve relevant results.

# 7 Discussion

Our findings indicate that concrete queries containing specific terms like "man," "woman," "well," and "son" tend to yield better results than more abstract or vague queries. This suggests that the precision of language used in queries significantly impacts the effectiveness of narrative type-scene retrieval.

Linguistic discrepancies further complicate this process. For instance, the absence of a direct equivalent for the modern concept of a "hero" in biblical texts highlights the challenges of aligning contemporary understandings with ancient languages.

This linguistic gap can lead to difficulties in accurately retrieving and interpreting type-scenes that involve heroic figures.

Circularity in data interpretation poses another challenge. Crowd-sourced votes for identifying type-scenes often come from English readers of English translations of the Bible. These translations may have been influenced by the translators' biases, potentially overemphasizing or underemphasizing certain narrative connections. This feedback loop can reinforce certain interpretations, making them seem more canonical than they might be in the original texts.

Additionally, confirmation bias can lead to the perception of type-scenes where none may exist. For example, the narrative of David and Bathsheba was identified as part of the "woman at the well" type-scene. However, this might represent an *inversion* of the type-scene. Instead of Bathsheba drawing David water, David sees and takes her, echoing the language of Eve "seeing and taking" the forbidden fruit, which connotes sinfulness. Although their union results in marriage, it is marked by tragedy rather than the celebration of a family lineage. The death of their child as a consequence of David's sin underscores this inversion.

This example illustrates how patterns might be perceived where there are none, with the model potentially focusing on superficial cues like "man," "woman," and "water." While such tools can inspire creative scholarly thinking, they cannot replace rigorous scholarly analysis. Researchers must remain vigilant against over-interpreting patterns and maintain a critical approach to ensure accurate and meaningful interpretations of biblical narratives.

## 8 Conclusions

Our study demonstrates that training a model to optimize embedding similarity between cross-referenced verses enhances an information-retrieval model's ability to identify repeating narrative patterns, known as type-scenes, in the Ancient Hebrew and Greek versions of the Bible. By fine-tuning a Siamese network with a contrastive loss function and using queries written in ancient languages, we were able to elicit type-scenes effectively.

Through manual error analysis and discussion of the returned verses, we conclude that while there is a detectable signal indicating the presence of type-scenes, the current setup has significant room for improvement. The model's potential utility is notable for Biblical scholars and personal devotees who are interested in exploring narrative parallelism within the Bible. However, further refinement and optimization of the system are necessary to enhance its accuracy and reliability. Future work will include a more sophisticated method for dividing books into narrative sections rather than individual verses and will remove masking from the training method. Finally, although we use the Bible as a case study due to its particularly fascinating self-referential narrative patterns, our approach may be adapted for long-form literary text of any genre to surface underlying narrative patterns.

## Limitations

Our research faces several limitations. Firstly, the negative sampling strategy plays a crucial role in achieving good results. However, our approach did not experiment with different strategies; we simply selected a random verse from the Bible, ensuring it was not related to the query verse. This approach may not have been optimal.

Secondly, for more accurate retrieval of type-scenes, we should have varied the candidate window. A single verse is often insufficient to represent an entire narrative scene. A more effective approach would have involved using a model capable of identifying narrative breaks or section headings to delineate narrative scenes.

Thirdly, the Greek type-scene queries were generated by a large language model (LLM) rather than a language expert. This may have adversely affected the results, as the LLM-generated queries might lack the nuanced understanding that a language expert would provide.

## Ethics Statement

To ensure the reproducibility of our experiments, we utilized the build program SCons[11]. Additionally, we are committed to transparency and openness in research, and thus, we have made our code publicly available at www.github.com/anonymous/repo.

## Acknowledgements

We thank Juan Moreno Gonzalez, Ph.D. candidate in the Faculty of Asian Middle Eastern Studies at Cambridge, for his enormous help in translating

[11]Source: https://scons.org/doc/production/HTML/scons-man.html

and validating the type-scene queries into Biblical Hebrew. This work is supported by the Woolf Institute for Interfaith Relations and the Cambridge Trust.

## References

Robert Alter. 1978. Biblical type-scenes and the uses of convention. *Critical Inquiry*, 5(2):355368.

Robert Alter. 1983. How convention helps us read: The case of the bible's annunciation type-scene. *Prooftexts*, 3(2):115–130.

Walter Arend. 1933. *Die typischen Scenen bei Homer*. Weidmann.

Erich Auerbach. 1959. *Scenes from the Drama of European Literature: six essays*. Meridian Books.

Vassileios Balntas, Edgar Riba, Daniel Ponsa, and Krystian Mikolajczyk. 2016. Learning local feature descriptors with triplets and shallow convolutional neural networks. In *British Machine Vision Conference*.

Brian Britt. 2002. Prophetic concealment in a biblical type scene. *The Catholic Biblical Quarterly*, 64(1):37–58.

Eryl W. Davies. 1995. *The Journal of Theological Studies*, 46(2):572–575.

Mark Finlayson. 2009. Deriving narrative morphologies via analogical story merging.

Niv Fono, Harel Moshayof, Eldar Karol, Itai Assraf, and Mark Last. 2024. Embible: Reconstruction of Ancient Hebrew and Aramaic texts using transformers. In *Findings of the Association for Computational Linguistics: EACL 2024*, pages 846–852, St. Julian's, Malta. Association for Computational Linguistics.

Cheng Jiayang, Lin Qiu, Tsz Chan, Tianqing Fang, Weiqi Wang, Chunkit Chan, Dongyu Ru, Qipeng Guo, Hongming Zhang, Yangqiu Song, Yue Zhang, and Zheng Zhang. 2023. StoryAnalogy: Deriving story-level analogies from large language models to unlock analogical understanding. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 11518–11537, Singapore. Association for Computational Linguistics.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference for Learning Representations (ICLR)*.

John Koutsikakis, Ilias Chalkidis, Prodromos Malakasiotis, and Ion Androutsopoulos. 2020. Greek-bert: The greeks visiting sesame street. In *11th Hellenic Conference on Artificial Intelligence*, SETN 2020, page 110117, New York, NY, USA. Association for Computing Machinery.

Burke O. Long. 1973. 2 kings iii and genres of prophetic narrative. *Vetus Testamentum*, 23(3):337–348.

Malte Ostendorff, Terry Ruas, Till Blume, Bela Gipp, and Georg Rehm. 2020. Aspect-based document similarity for research papers.

Andrew Piper, Richard Jean So, and David Bamman. 2021. Narrative theory for computational narrative understanding. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 298–311, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Vladimir Propp. 1968. *Morphology of the folktale*. Austin: University of Texas Press.

Shaltiel Shmidman, Avi Shmidman, and Moshe Koppel. 2023. Dictabert: A state-of-the-art bert suite for modern hebrew.

Pranaydeep Singh, Gorik Rutten, and Els Lefever. 2021. A pilot study for bert language modelling and morphological analysis for ancient and medieval greek. In *The 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2021)*.

Katherine Thai, Yapei Chang, Kalpesh Krishna, and Mohit Iyyer. 2022. Relic: Retrieving evidence for literary claims.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Ivan P. Yamshchikov, Alexey Tikhonov, Yorgos Pantis, Charlotte Schubert, and Jürgen Jost. 2022. Bert in plutarchs shadows. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*.

| English Query | Translation (hbo) | Literal Translation |
|---|---|---|
| A man meets woman by a well | וַיָּבֹא אִישׁ וַתֵּצֵא אִשָּׁה לִקְרָאתוֹ עַל בְּאֵר | Arrived a man and came out a woman to meet-him by a well |
| An angel announces a barren woman will bear a child | וַיֹּאמֶר מַלְאַךְ לְאִשָּׁה עֲקָרָה וְיָלַדְתְּ בֵּן | Said the angel to the barren woman "you will bear a son" |
| The younger son is preferred to the older son | והוא אָהַב הַצָּעִיר מהבְּכוֹר | And he loved the young (more than)-the firstborn |
| God sends a prophet to speak to a rebellious people | וַיִּשְׁלַח יְהוָה נָבִיא לְדַבֵּר אֶל־עַם סוֹרֵר | And sent the Lord a prophet to speak into a rebellious people |
| A man has an epiphany in the field | וַתְּהִי עָלָיו יְהוָה בַּשָּׂדֶה | And came upon him the lord in the field |
| A prophet veils his face | וַיְכַס הַנָּבִיא אֶת־פָּנָיו | And veiled the prophet his face |
| A well is found in the desert | וַיִּמְצָא בְּאֵר בַּמִּדְבָּר | And he found a well in the desert |
| A dying hero gives a speech | דְּבַר גִּבּוֹר מֵת | The words of a dead mighty man (warrior) |
| A hero is tempted | גִּבּוֹר נִסָּה | A mighty man (warrior) is tempted |

Table 5: **Ancient Hebrew Type Scene Queries**

.

| English Query | Translation (grc) |
|---|---|
| A man meets woman by a well | Ἄνθρωπος ἦλθεν καὶ ἐξῆλθεν γυνή εἰς ὑπάντησιν αὐτῷ πρὸς φρέαρ. |
| An angel announces a barren woman will bear a child | Ἄγγελος ἀναγγέλλει γυναῖκα στεῖραν ὅτι τεκεῖν ἕξει παῖδα. |
| The younger son is preferred to the older son | Καὶ ἠγάπα τὸν νεώτερον πλείονα τοῦ πρωτότοκου. |
| God sends a prophet to speak to a rebellious people | Καὶ ὁ κύριος ἀπέστειλε προφήτην λαλῆσαι εἰς λαὸν ἀπειθοῦντα. |
| A man has an epiphany in the field | καὶ ἐπέρχεται ἐπ' αὐτὸν ὁ κύριος ἐν τῷ ἀγρῷ. |
| A prophet veils his face | Καὶ ὁ προφήτης ἐκάλυψε τὸ πρόσωπον αὐτοῦ. |
| A well is found in the desert | Καὶ εὗρε φρέαρ ἐν τῇ ἐρήμῳ. |
| The words of a dying hero | Οἱ λόγοι τοῦ ἀποθνήσκοντος ἥρωος. |
| A hero is tempted | Ἥρωι πειραζεται. |

Table 6: **Ancient Greek Type Scene Queries**

.

# A Appendix

## Versification Discrepancies in Biblical Manuscripts

The standard chapter and verse numbering found in modern translations and editions of the Biblical texts does not appear in the earliest extant manuscripts. These numbering systems are later additions from the scribes of the Middle Ages, with modern versions largely adopting the verse system codified in the 1611 King James Version. For instance, a verse appearing as Exodus 21:37 in our manuscripts is numbered as Exodus 22:1 in the King James Version.

In most cases, these discrepancies are minor, differing by only one or two verses. However, the two manuscripts we use represent distinct textual traditions: the Leningrad Codex follows the Masoretic Text, while the Septuagint follows a different tradition. As a result, some verses are unique to one tradition and do not have a parallel in the other. For example, certain verses in the Leningrad Codex do not appear in the Septuagint, and vice versa.

To manage these discrepancies, we exclude any source and target pair where either the source or the target is missing. Consequently, there are slight differences between the examples from the Septuagint and those from the Leningrad Codex.

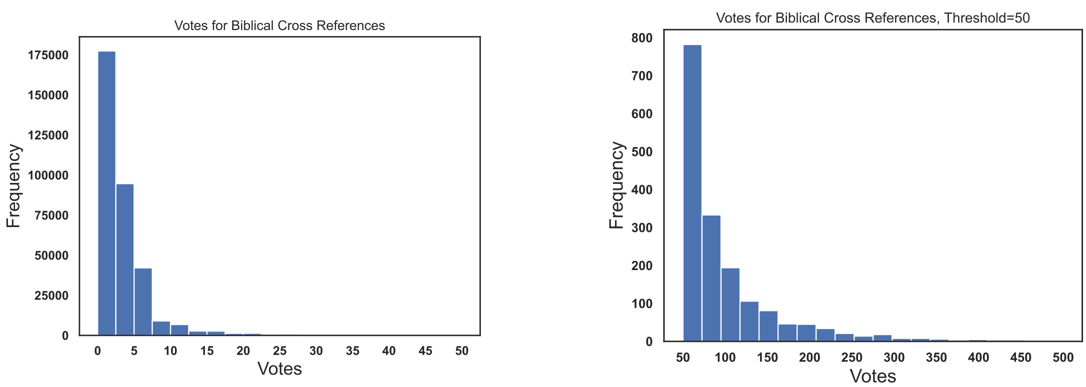



Figure 5: Cross-Reference Votes.

Additionally, the Greek version used was slightly messy, with a few instances of misnumbered chapters (e.g., two consecutive chapters numbered 17). These issues were manually corrected to ensure consistency in our analysis.