Title	Identifying Binary Oppositions in the Lord of the Rings : A primary quantitative approach to binary oppositions
Author(s)	Dimitra, Vogatza
Citation	言語文化共同研究プロジェクト. 2025, 2024, p. 43-67
Version Type	VoR
URL	https://doi.org/10.18910/102221
rights	
Note	

The Lord of the Rings における二項対立の識別
二項対立への主要な定量的アプローチ

Vogatza Dimitra

大阪大学大学院人文学研究科

〒 560-0043 豊中市待兼山町 1-8
E-mail:　dvogatza@yahoo.com

あらまし　本論文は、*the Lord of the Rings* における二項対立を特定するための主要な定量的アプローチを提示する。Tolkien の作品のみに基づいて構築されたカスタム Word2Vec モデルを使用し、BERTopic を統合してトピックモデリングを行い、意味論的な関係が Tolkien の物語構造とテーマの対立をどのように反映するかを探求する。分析の結果、癒しと苦悩、自然と戦争といった明確なテーマの二項対立が明らかになり、これらは文脈レベルおよび語彙レベルの両面から検討される。トピック間の高コサイン類似度および低コサイン類似度のスコアは、意味論上における距離の指標として使用され、整合したカテゴリ内および概念的に異なるカテゴリ間の対立を明らかにする。結論として、エンベディングモデルが文学分析にもたらす意味について議論し、Tolkien の伝説体系全体にわたって二項対立の広範な影響を理解するための更なる探求を提案する。

キーワード　Tolkien、　二項対立、　エンベディング モデル、　コサイン類似度、BERTopic

Identifying Binary Oppositions in *the Lord of the Rings*:
A primary quantitative approach to binary oppositions

Dimitra Vogatza

Graduate School of Language and Culture, University of Osaka

1-8 Machikaneyama-cho, Toyonaka, Osaka, 560-0043 Japan

Abstract　This paper presents a primary quantitative approach to identifying binary oppositions in *the Lord of the Rings*. Using a custom-built Word2Vec model trained exclusively on Tolkien's works, and integrating it with BERTopic for topic-modelling, the study explores how semantic relationships reflect Tolkien's narrative structure and thematic oppositions. These oppositions were revealed through analysing the results. The analysis showed binaries such as healing and trauma, and nature and war, which are examined both at a contextual and lexical level. High and low cosine similarity scores between topics are used as indicators of semantic distance, identifying oppositions both within the same thematic categories and across conceptually distinct ones. The study concludes with a discussion on the implications of an embeddings model for literary analysis, proposing further exploration across Tolkien's legendarium to understand the broader impact of binary oppositions in his work.

1.   Introduction

1.1.   Binary Oppositions and *The Lord of the Rings*

J.R.R. Tolkien's (1892-1973) *the Lord of the Rings*' primary and overall concept that might come to the mind of any reader of the work is the triumph of good over evil—a very apparent binary opposition found in many an epic story. Yet, one of the structural observations that are made on Tolkien's work is in fact a form of a story-long established recurring opposition: havens in perilous situations (Holmes, 2014: 147). He creates these moments of safety in increasingly more dangerous situations as the narrative progresses. This is not all however, there is also deliberate linguistic and stylistic contrast between the Hobbits—plain and modern—and other more archaic peoples (Turner, 2007; Turner, 2014: 392). Additionally, on a more lexical level perhaps, Tolkien plays with words for colours, nature, emotions and moods to express opposed states. He even coined a word himself to express opposed states: *eucatastrophe*. He states (*Letters 131*):

> In the scene where all the hosts of the West unite to do honour and praise to the two humble Hobbits, Frodo and Sam, we reach the 'eucatastrophe' of the whole romance: that is the sudden joyous 'turn' and fullfilment of hope, the opposite of tragedy, that should be the hallmark of a 'fairy-story' of higher or lower tone, the resolution and justification of all that has gone before.

Along the varying voices that discuss the style in *the Lord of the Rings*, one has given a quite intriguing view on binary oppositions. It is Le Guin's (2001) appraisal of the rhythmic quality of the work that inspired me to explore if and how binary oppositions in *the Lord of the Rings* function as rhythm. She claims that they appear in a pattern of stress and relief, much like a contextual trochee on a multitude of levels (from chapter to book to volume). Other scholarly works on binaries in *the Lord of the Rings* is from Pretorius (2002) focusing on a few specific binaries through a literary analysis approach, Doyle's (2020) cultural and moral approach on Tolkien's good and evil, and Khan's (2024) approach focusing on linguistic and narrative structures on binary oppositions. Verlyn Flieger, has also explored oppositions in *the Splintered Light* (1983), stating:

> No careful reader of Tolkien's fiction can fail to be aware of the polarities that give it form and fiction.

This paper focuses on a primary quantitative approach to identifying binary oppositions through the creation of a Tolkien embeddings model and its incorporation in topic-modelling techniques. The model consists solely of Tolkien's writings that are stylistically and contextually relevant to *the Lord of the Rings*. The works and manuscripts that are included in the embeddings model are listed in Tables 1 and 2.

Table 1: Published works by Tolkien included in the model

No.	Date	Title	Classification	Word-tokens
1	1935	*The Hobbit*	Legendarium	95,968
2	1945	*Leaf by Niggle*	Short Story	7,467
3	1949	*Farmer Gilles of Ham*	Short Story	16,476
4	1954-1955	*The Lord of the Rings*	Legendarium	496,764
5	1962	*The Adventures of Tom Bombadil*	Legendarium	8,738
6	1967	*Smith of Wootton Major*	Short Story	10,237
			Sum of word-tokens: 635,650	

1.2. Tolkien's Textual Background

The contextually relevant writings along with *the Lord of the Rings* are part of Tolkien's legendarium. They start as early as 1914 and continue on well into the 1960s, yet throughout his life only three relative works were published (see Table 1). From the numerous unpublished works that are related to the legendarium some were complete yet abruptly ended (i.e. *the Children of Húrin*), but most were either incomplete, revised a multitude of times or existing in a few different versions of varying discrepancy (i.e. works published in *The Fall of Gondolin* see Table 2). Along his work on his legendarium, he also wrote a few different short stories and poems or verses, a small number of them is also included in the model based on their relevance.

Inspired by the Finnish and Celtic language when he was still a teen, he started creating his own languages, and they needed a world to exist in—the legendarium (*Letters 219-220*). It could actually be considered a 'side project' and his languages a 'secret vice' as he called it (*Letters , Fimi 20*), as he was a professor of English Language and Literature at the University of Oxford—not that his academic interests and his writings do not have common roots. His translations of Old and Middle English epics or romances like *Beowulf* and *Sir Gawain and the Green Knight*, influencial and analytical lectures on said works (*Beowulf: The Monsters and the Critics*), and his background as a lexicographer for the Oxford English Dictionary mirror his own romances and epics and language creation. Yet it is because of this that he did not have the luxury of time to dedicate to his creations, which leaves us with numerous unfinished manuscripts, hastily and infamously illegible writings on loose pieces of paper, or poems written on the margins of books.

After Tolkien's death the vast majority of his writings has been edited and published by his son Christopher Tolkien (henceforth Christopher). This process entailed many challenges, as it would often be the case where Tolkien would cross out words from manuscripts and make emendations, or create a whole new version of that writing. Sometimes the style was different, some other times the story would change, or some others the character names would change—more drastic as Tevildo to Sauron, or simpler like Melko to Melkor. The problem that arises here is that due to this multiplicity of versions and emendations, and manuscript completeness the process of identifying the appropriate texts for the model became significantly more time consuming than expected.

Table 2: Posthumously published works by Tolkien included in the model

No.	Date Approx. (Pub. Date)	Title Published In	Category	Word-tokens
1	1912-1914 (2015)	*The Story of Kullervo* *The Story of Kullervo*	Finnish Tale	8,593
2	1916-1917 (2018)	*The Tale of the Fall of Gondolin* *The Fall of Gondolin*	Legendarium	23,459
3	1917-1950s (1977)	*The Silmarillion* *The Silmarillion*	Legendarium	130,757
4	1917 (2017)	*The Tale of Tinúviel* *Beren and Lúthien*	Legendarium	15,230
5	1920-1943 (1976)	*Letters from Father Christmas* *Letters from Father Christmas*	Letters	15,966
6	1920s-1937 (2007)	*The Childen of Húrin* *The Children of Húrin*	Legendarium	60,124
7	1925-1931 (2017)	*Lay of Leithian* *Beren and Lúthien*	Legendarium	18,163
8	1926 (2018)	*The Fall of Gondolin (Sketch)* *The Fall of Gondolin*	Legendarium	1,976
9	1930 (2017)	*Beren and Lúthien (Quenta Noldorinwa)* *Beren and Lúthien*	Legendarium	6,164
10	1930 (2018)	*The Fall of Gondolin (Quenta Noldorinwa)* *The Fall of Gondolin*	Legendarium	4,226
11	1930 (2017)	*Lost Tale of Nauglafring* *Beren and Lúthien*	Legendarium	2,505
12	1937 (1998)	*Roverandom* *Roverandom*	Short-Story	25,091
13	late 1950s (1993)	*Athrabeth Finrod ah Andreth* *History of Middle-Earth: Morgoth's Ring*	Legendarium	8,708
			Sum of word-tokens: 320,962	

Note: Dates reflect the estimated time of original composition. Publication dates refer to editions edited by Christopher Tolkien or later editors. The works are represented with the original composition followed by the book it was published in beneath it.

Some years after the initial publication of *the Silmarillion*, a work intended by Tolkien himself to be published along with *the Lord of the Rings*, Christopher went on and published a series of 12 volumes called *the History of Middle Earth* including as much of his father's unedited writing as he could with ample commentary and discussion. Primarily *the History of Middle-Earth* follows closely the contents of *the Silmarillion*, the latter having received considerable editing to be able to be published as a complete narrative. However, due to the sheer size of the volumes and the overall time required for separating editor from author, these writings are not included in this paper's model construction, apart from *Athrabeth Finrod ah Andreth* finished some time around 1959. *The Silmarillion* in this case is included, albeit with slight caution.

2. Quantitative Methods for Identifying Oppositions

As discussed in 1.2. the model was constructed with the texts listed in Tables 1 and 2. The model was trained on sentence level granularity. This is due to the fact that there is no one way

the oppositions appear, there are cases where they appear in between a couple of pages (5.08.011-5.08.020)[1] or just within a paragraph (4.08.004)[1] or on chapter titles (i.e. chapters 1 and 2)[2]. So I chose finer granularity for this paper (Ding et al., 2021). Furthermore, I maintained the hyphenated words, i.e. when I removed the punctuation after the sentence breaks. This is to ensure Tolkien's characteristic hyphenation (Alden, 2021; Tauber 2025) as part of his stylistic choice. Similarly, accents to words belonging to Tolkien's languages are also maintained, i.e. Lúthien. Finally the sentences were broken on periods, exclamation marks and question marks. Only point of note with breaking the texts into sentences is the poems and songs found throughout his texts. I chose to maintain sentence level breaking and not break them per lyric, so that each line of the corpus would be a linguistically sound sentence.

2.1. Training the Tolkien Model

To train the word embdeddings for the model on around 950,000 tokens of my Tolkien corpus, I used Word2Vec's skip-gram architecture. Even though it is primarily directed towards considerably larger datasets than mine, it suited mine because of its ability to capture nuances. Skip-gram has been shown to perform better with infrequent words (Mikolov et al, 2013: Golberg et al, 2014), which is particularly relevant in a literary corpus compared to other architectures like CBOW (Continous Bag of Words). This choice was also influenced by prior comparative work in NLP textbooks (Albrecht et al., 2020), as well as my own preliminary tests, where skip-gram produced more coherent clusters for the analytical goals of this study.

The skip-gram model was trained using a vector size of 100 and a window of 5. Words occurring fewer than five times were excluded from training to reduce noise. The model was trained over 5 epochs, providing sufficient convergence given the corpus' size and thematic consistency. Additionally, following the approach introduced by Mikolov et al. (2013), I employed negative sampling (negative=5) to improve efficiency.

I would like here to note that at an early stage I tested GloVe's applicability in identifying oppositions in *the Lord of the Rings*. GloVe—gigaword 50d variant—is trained on Wikipedia and newswire text, which can be considered as non-literary English usage (Pennington et al., 2014; Parker et al., 2011). It has a vocabulary size of approximately 400,000 words. This preliminary evaluation was conducted by measuring high-similarity word pairs using cosine similarity. Yet the results were quite unsatisfactory. It was with this incentive that I created the Tolkien model. While the Tolkien model proved to be more useful, I needed to confirm whether it was Tolkien's language that made it *better* or if this could be contributed to it being a literary model. With this in mind I constructed two comparative Word2Vec models: one trained on British literary texts

---

[1] Citation format taken from the Digital Tolkien Project. *The Lord of the Rings* is broken down into 6 books throughout 3 volumes, 2 books per volume. The main text format is <book>.<chapter>.<paragraph>.

[2] Chapter 1: *A Long-Expected Party*
Chapter 2: *The Shadow from the Past*

from Tolkien's period (1900s-1950s), and another on texts by authors known to have inspired Tolkien. While both literary models performed better than GloVe, the model based on Tolkien's inspirations yielded the highest opposition-to-vocabulary ratio (see Fig.1). Simply put, aligning style and theme in training embdeddings for literary analysis is highly important. Although the Tolkien-trained model did not produce the absolute highest ratio, it significantly outperformed GloVe, suggesting that domain-specific embeddings offer more precise semantic representation when applied to texts of similar register and lexicon. Refer to the Appendix for a high-similarity lexical comparison between GloVe and Tolkien models, based on a few keywords relative to the *the Lord of the Rings*.
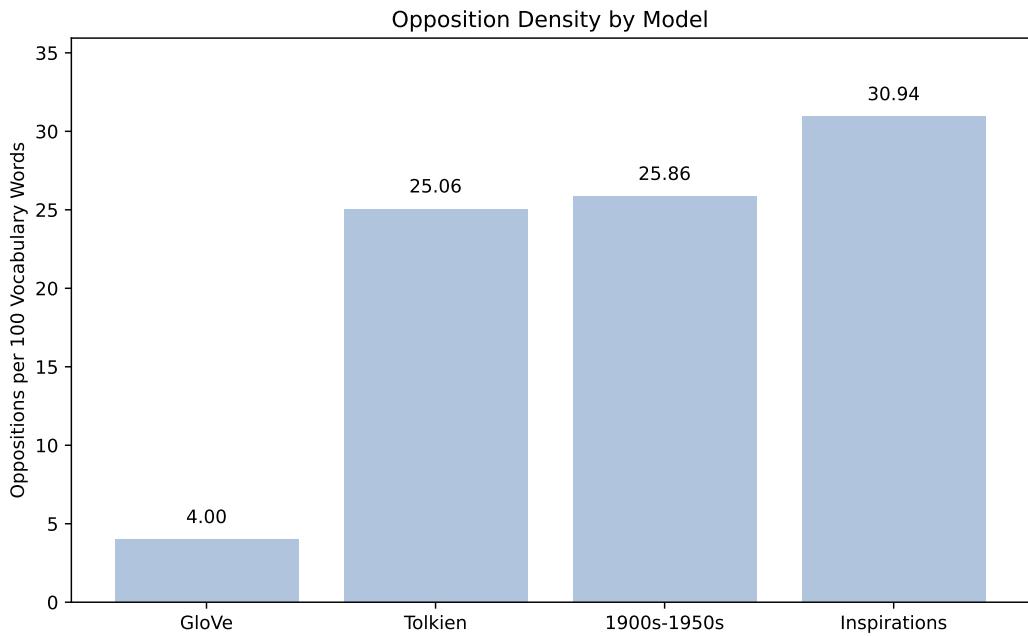

Figure 1: Opposition-to-vocabulary ratios for different word embedding models.

2.2. Tolkien Model Implementation in BERTopic
2.2.1. Data Preprocessing

As the aim of the methods in this paper is to identify binary oppositions in *the Lord of the Rings* with as little qualitative intervention as possible, I opted to explore topic-modelling as means of identifying opposed words or themes. Preprocessing of the plain text consisted of: stopword removal and document segmentation. *The Lord of the Rings* contains very high frequencies of many proper nouns (see Table 3 for examples) that significantly influence the quality and coherence of the topics, so I included them in the stopword list. To maintain thematic cohesion within all the documents processed by BERTopic I segmented the text on paragraph level and in cases of dialogue I kept the dialogue from start to finish as part of the same segment. There was a need of re-arranging the segments after stopword removal as some segments' final token count was reduced to very low values such as 3. This resulted to a document length average of 105.13, with

Table 3: Examples of proper nouns and their frequencies in *the Lord of the Rings*

Frodo	1982
Sam	1290
Gandalf	1118
Aragorn	721
Shire	277
Gondor	261

the total tokens of the documents post stopword removal being 142,874—a little over one fourth of the original. While it is the typical fashion in topic-modelling to segment in bigger chunks of text (Misra et al., 2011; Riedl & Biemann, 2012; Tabata, 2020) by setting a set amount of tokens per chunk, I chose to maintain the semantic cohesion of a paragraph—whether long or short—as the work changes places and perspectives often after volume 1 *the Fellowship of the Ring*, where the narrative breaks into three major fronts. Similarly, keeping dialogues intact ensures that the intricacies of character speech (also discussed in sections 1.1. and 3.) remain untouched within each segment.

2.2.2.   Cosine Similarity in BERTopic and Word2Vec

BERTopic is a neural topic-modelling technique that leverages progress done in model embeddings and clustering techniques to achieve more coherent and representantive topics than traditional models like LDA(Latent Dirichlet Allocation) or NMF(Non-Negative Matrix Factorization)(Grootendorst, 2022). For this paper, I utilised the normal BERTopic function with HDB-SCAN(Hierarchical Density-Based Spatial Clustering of Applications with Noise) for document clustering and UMAP(Uniform Manifold Approximation and Projection for Dimension Reduction) for dimension reduction. I opted for BERTopic as it comparatively produced more coherent and nuanced results than LDA, but this will not be further discussed in this paper. As BERTopic typically converts documents into embeddings using pre-trained transfomer based models such as BERT(Bidirectional Encoder Representations from Transformers) by default, it was also useful in measuring inter-topic distance by applying my Tolkien model as a model for measuring oppositions using cosine distance. For the computation of inter-topic cosine distance, I averaged the Word2Vec embeddings trained on Tolkien's corpus for each topic. In other words, the mean of the word vectors for each topic. I checked how the word similarities are distributed, and it is a normal distribution (see Fig. 2), so taking the mean instead of the median leads to the same results.
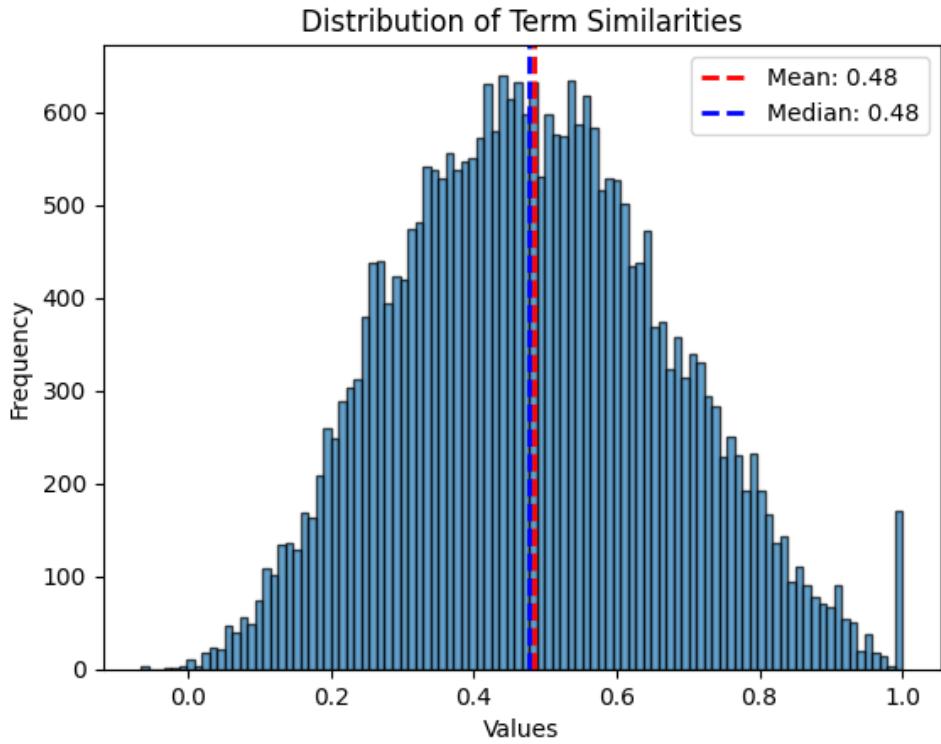
Figure 2: Distribution of term similarities based on the Word2Vec Tolkien model.

Now, BERTopic has its own cosine similarity function, however this similarity is calculated differently. The default function in BERTopic utilises the embedding's model used to extract the topics (this is based on the language set as an argument at the extraction/learning point of the topic-modelling process, in this case the all-MiniLM-L6-v2 (Sentence Transformers all-minlm-L6-V2, n.d.;Wang et al., 2020) is used). While the topics were highly coherent in their contextual nuances, this was not the same for the default inter-topic distances. A clearer representation of the differences between the default inter-topic distance matrix and the Tolkien Word2Vec model can be seen in Figure 3 below.
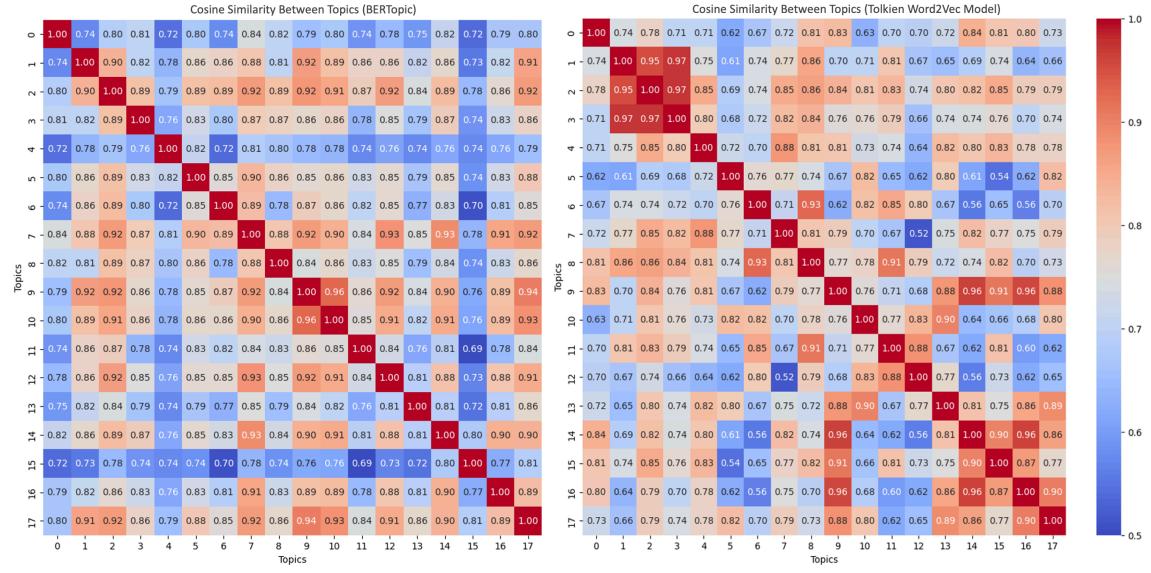
Figure 3: Default BERTopic (left) and Tolkien model (right) cosine similarity matrices

The figure on the left depicts the inter-topic cosine similarity of just the BERTopic emdeddings based on the model I mentioned above, while the figure on the right is based on my Tolkien Word2Vec model. As it can be observed, there are distinct differences between the two. The discussion on the topic meanings itself can be found later on in the paper, so I will not go into in depth detail as to what each topic represents, however, I would like to draw attention to a few specific differences that are striking. Topic 12 dealing with war and 7 dealing with trees have a difference of 0.41 between the two graphs, with the Tolkien model showing low-similarity. As Word2Vec embdeddings capture the meaning of a word in the context of the corpus that it is trained in, this difference goes to show that to Tolkien trees and war are very far apart in contextual usage. Additionally, while both graphs' cosine threshold was set to 0.5 or higher, the Tolkien graph has more significant fluctuations between high and low values. The Tolkien Word2Vec model overall provides more accurate and educated distances than the default distance.

2.2.3. Most Representative Chapters per Topic

As this paper aims to identify oppositions in *the Lord of the Rings*, and as a rhythm as well, I utilised BERTopic's most representative document function to identify which chapters were more representantive of which topic (similar to LDA's alpha value, (Blei et al., 2003)). Based on each chapter's narrative note—a chapter's negative or positive note—as well as the terms themselves, I classified the topics into positives and negatives. The argument here is that, since this function aids in understanding in what intervals the topics are appearing throughout the work, by classifying them into these two categories can quantitatively showcase the contextual trochaic rhythm. For the visualisation of the topics throughout the chapters I used ChatGPT and a table containing the top three chapters per topic (see Figure 4). This is further discussed in section 3.

51
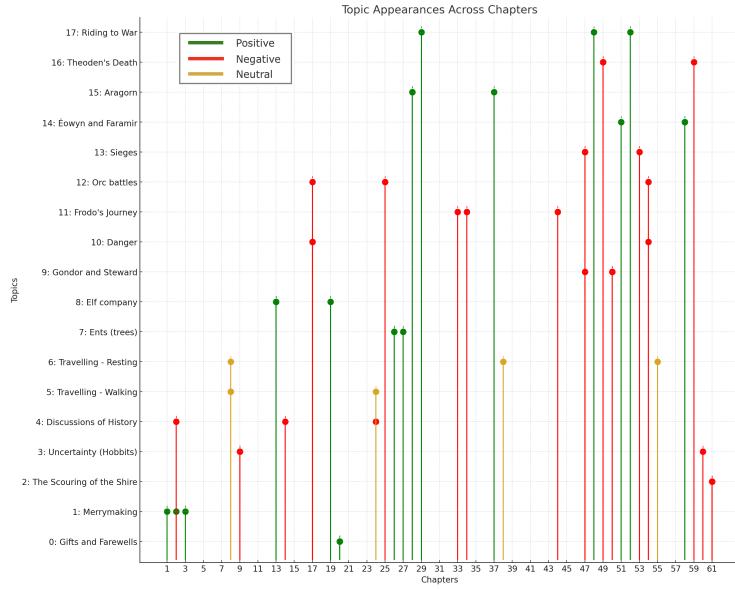
Figure 4: Topic occurrence throughout *the Lord of the Rings*

3. Results

The implementation of topic-modelling and semantic vector analysis yielded several notable patterns in the textual structure of *The Lord of the Rings*. This section is divided into three parts. First, I present the extracted topics, including their distribution and content (Table 4). Second, I examine the semantic proximities between topics by applying cosine similarity to averaged Word2Vec embeddings as introduced above. Finally, I focus on identifying recurring thematic binary oppositions throughout the work by analysing the most representative chapters across thematically distinct topics.

Apart from the contents of each topic in the table below, we can also see the positive (+) or negative (-) classification, and the most representantive chapters for each topic. Under the umbrella of positive I consider topics that deal with Tolkien's forces of good (Elves, Hobbits, kings, nature) as well as positive concepts as seen in Topic 14 that deals with love and healing. Topic 17's positive classification despite its connotations of war is derived from the context of valiant and honourable warriors. For the negative classification, concepts such as death, war, suffering, battles, danger and uncertainty are marked as negative. Topic 9, which at a first glance cannot outright be discerned as negative by looking at its terms, deals with the state of Gondor at the first stages of the work. A city whose Steward was manipulated by evil forces and later on tragically ended his life while trying to kill his unconsious son in a moment of despair.

Table 4: Topic overview with terms

Topic	Labels and Classification	Chapters	Top 15 Terms
0	Gifts and Farewells .... (+)	20, 20, 20	silver, gave, set, virtue, gifts, wrought, lady, spoons, parting, gift, garments, gems, fair, leaves, small
1	Merrymaking .... (+)	1, 2, 3	hobbits, good, began, birthday, guests, supper, food, fire, party, small, table, hobbit, people, hobbiton, eat
2	The Scouring of the Shire . (-)	61, 61, 61	ruffians, farmer, hobbits, rosie, lads, village, leader, chief, weapons, score, hobbiton, waggons, gang, hundred, sheds
3	Uncertainty (Hobbits) ..... (-)	9, 9, 60	hobbits, inn, chief, folk, travellers, gate, bob, time, good, hob, times, farmer, ponies, big, pony
4	Discussions of History .... (-)	14, 2, 24	ring, time, men, power, orcs, fear, thought, enemy, dark, found, find, elves, ents, hope, night
5	Travelling - Walking ...... ( )	8, 8, 24	trees, road, light, land, dark, mountains, sun, river, side, hills, path, high, looked, sky, night
6	Travelling - Resting .... ( )	38, 55, 8	sleep, asleep, water, night, light, heard, woke, lay, sat, dream, wind, fell, soft, eyes, awake
7	Ents (trees) .... (+)	27, 27, 26	ents, elves, hobbits, trees, forest, folk, call, heard, days, time, woods, land, speak, thought, men
8	Elf Company .... (+)	19, 19, 13	sleep, thought, looked, face, food, night, end, lady, hear, bright, head, words, deep, fire, elvish
9	Gondor and Steward ...... (-)	50, 50, 47	lord, city, door, house, servants, houses, stood, burn, turned, healing, death, kings, steward, men, face
10	Danger .... (-)	54, 54, 17	door, doom, passage, ran, stairs, orc, orcs, light, side, heard, hall, feet, gate, steps, sprang
11	Frodo's Journey .... (-)	33, 33, 34	eyes, precious, master, dark, felt, hands, looked, thought, head, hand, hobbits, lay, face, hobbit, moment
12	Orc Battles .... (-)	54, 34, 17	orcs, orc, hand, legs, head, fell, eyes, black, sword, sprang, cried, ground, cry, knife, red
13	Sieges .... (-)	47, 47, 53	men, wall, retreat, city, walls, gate, cried, enemy, orcs, black, horsemen, army, lord, fire, rock
14	Love and Healing .... (+)	58, 58, 51	lady, warden, lord, house, king, healing, city, care, days, battle, healer, houses, looked, women, war
15	Aragorn .... (+)	37, 37, 28	sword, lord, man, speak, hand, master, son, men, spoke, ring, eyes, thing, words, leave, days
16	Theoden's Death .... (-)	59, 59, 49	king, city, prince, victory, men, knights, mark, lord, arms, bear, death, rode, ride, honour, lay
17	Riding to War .... (+)	52, 29, 48	king, rode, men, ride, riders, lord, night, rider, host, road, turned, white, days, wind, passed

I would like to separately discuss topics 5 and 6 as they deal with travelling. While they can

be marked as negative and positive, walking and resting respectively, the classification would not follow the same principles as above. There are cases in the work when walking is not a perilous thing even though it is tiring, and there are cases when resting does not provide rest. Thus I classify them as neutral and kept the brackets blank.

Consequently, in the above Figure 4, topics 5 and 6 are coloured yellow, positives green and negatives red. While this figure needs refining, we can observe that at a primary stage there are a few areas showcasing a reversal between positive and negative. Quite intriguing is the back and forth between positive and negative topics from chapters 47 to 54, with 7 different topics actively coming in and out of the narrative.

3.1.   Inter-Topic Distance

In this section I will closely examine the high and low similarities as depicted in the graph below, while investigating the more appropriate cosine threshold for each.
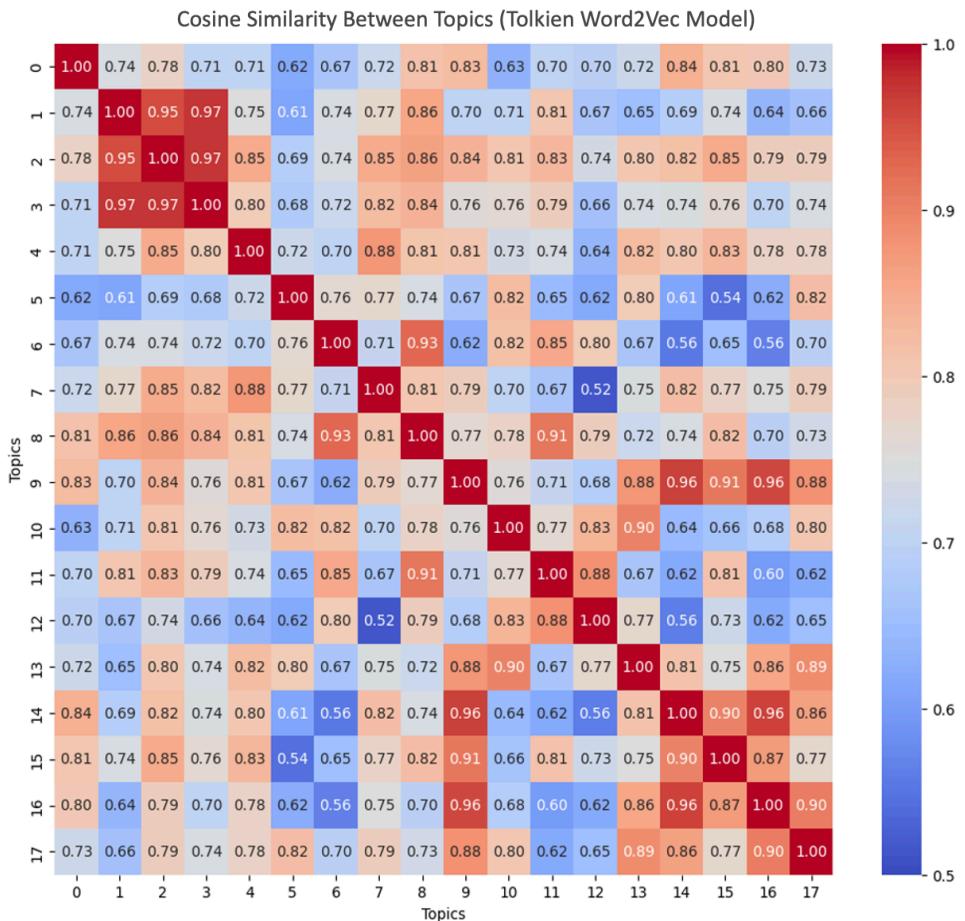

Figure 5: Cosine similarity matrix based on the Tolkien model.

Firstly, when evaluating cosine distance the rule is that the higher the value the closer the distance and the opposite (Sahoo & Maiti, 2025), but what would be the appropriate threshold? What

value or higher and what value or lower works the best in this case? To evaluate this, I firstly crossed out the middle values, from 0.70 to 0.80, with the output range being from 0.5 to 1. I would also like to note that this calculation was also run with range of 0 to 1, but the results were still above 0.5 which did not produce an easily readable graph, hence why in this case it is set to start from 0.5.

Starting with the first 10 decimals on either side—0.60 to 0.70 and 0.80 to 0.90—there are few things that can be easily seen. For example topics scoring at 0.85 like 2 and 15, or 2 and 7, are irrelevant, but not necessarily opposed either. These two pairs deal with the Scouring of the Shire—Hobbits returning home at the end of the work to find it overrun by evil forces and reclaiming it—trees, and Aragorn—the second protagonist reclaiming his position as king. Consequently, anything below 0.85 gives the same results of thematic irrelevance between topics. Now, we can redirect our attention to topic distance of 0.88. Topics that score 0.88 do show some relevance, for example topic 9 and topic 13. Topic 9's context of Gondor and Steward does occur during the siege of Gondor, but topic 13 that deals with sieges is not limited to Gondor. At the same time, topics 11 and 12 also score 0.88, this score can be attributed to Frodo's journey taking a negative turn and Frodo suffering for the majority of his travels because of the same forces that mobilise the orcs into battle. However, topics 9 and 17 have much less in common. As previously explained in the above section, topic 9 deals with Gondor's steward and topic 17 with warriors, these two also do not have thematic connections through the work. As this score is not consistently applicable to function as a threshold for high-similarity, I move one up to 0.89. At 0.89 there are topic pairs like 13 and 17, both dealing with war on the same fronts which confirms their similarity. Interestingly enough, 0.89 appears only in this topic pair. From this threshold and higher the topics along with their scores and similarity characteristics are portrayed in Table 5.

Table 5: High-similarity topic pairs with threshold 0.89 or higher

Similar Themes	Topic Pairs with Scores
Frodo and Resting	8 & 11 ..... 0.91
Elves and Resting	8 & 6 ...... 0.93
Hobbits	1 & 2 ...... 0.95 2 & 3 ...... 0.97 3 & 1 ...... 0.97
Kingdoms of Men	9 & 14 ..... 0.96 9 & 15 ..... 0.91 14 & 15 ..... 0.90 14 & 16 ..... 0.96 16 & 17 ..... 0.90
Sieges and War	9 & 16 ..... 0.96 13 & 17 ..... 0.89

Now focusing on the low-similarity thresholds, the scores are evaluated in a similar fashion. In this scenario however, I am evaluating the opposition between the topics. Topics scoring at 0.68 like 9 and 12, are again irrelevant. Similarly topics 10 and 16 score 0.68, they deal with dangerous situations, and one of the characters' death, respectively. These two are again irrelevant even though they are negative, because topic 16 deals more with a character's death and discussions about his death rather than danger. From here, let us lower the threshold to 0.66. Topics 10 and 15 are up to an extend opposed, as the topic concerning Aragorn deals with heroism and kings in contrast to topic 10 that deals with danger. Yet I cannot call them thematic oppositions with certainty, due to their contextual differences (one deals with a character and the other with a state of things). Topics 3 and 12 are quite irrelevant, as the former deals with uncertain situations around the Shire (where Hobbits live), which could be described as typical British countryside (Curry, 1997; Weidner, 2002). The uncertainty that is occurring around the Shire in the work, and the orc battles in topic 12 have some similarities based on the forces that cause both of them in the narrative, but they are not opposed nor precisely similar. Consequently, we move to the next value of 0.65. Here, topics like 1 and 13 start to show more clear oppositions. Topic 1 deals with merrymaking, especially Hobbit feasts and celebrations, while topic 13 deals with sieges and of course has very strong connotations to war. These two are more thematically opposed than the previously discussed, as this distinctly shows times of peace versus times of war. Similarly, topics 12 and 17 dealing with orc battles and valiant warriors riding to war, are opposed as the two sides of good and evil. From 0.65 and lower the topics along with their scores and opposing characteristics are as found in the table below. Note here that travelling has low-similarity with a variety of different topics and it is difficult to locate the reason. It is possible that the terms in topic 5 are significantly more nature oriented than all the other topic terms which can lead to low-similarity by default. However there are a couple of topics that pair with travelling that show oppositions and are highlighted in Table 6.

Table 6: Low-similarity topic pairs with threshold 0.65 or lower

Opposing Themes	Topic Pairs with Scores
Travelling and Merrymaking	5 & 0 . . . . . . 0.62
	5 & 1 . . . . . . 0.61
	5 & 11 . . . . . 0.65
	5 & 12 . . . . . 0.62
Travelling and Healing	5 & 14 . . . . . 0.61
	5 & 15 . . . . . 0.54
	5 & 16 . . . . . 0.62
Resting and War	6 & 9 . . . . . . 0.62
	6 & 14 . . . . . 0.56
	6 & 15 . . . . . 0.65
	6 & 16 . . . . . 0.56
Danger and Healing	0 & 10 . . . . . 0.63
	10 & 14 . . . . . 0.64
Frodo(Hobbits) and Men	11 & 14 . . . . . 0.62
	11 & 16 . . . . . 0.60
	11 & 17 . . . . . 0.62
Orcs and Men	7 & 12 . . . . . 0.52
	12 & 14 . . . . . 0.56
	12 & 16 . . . . . 0.62
	12 & 17 . . . . . 0.65

Note: Travelling and Merrymaking's respective score and topics are marked with blue. Travelling and Healing's are marked with purple.

4. Discussion

This section evaluates the reliability and limitations of the methods used to extract and measure oppositions in *The Lord of the Rings*. Rather than using the topics as thematic representations, the analysis assesses whether topic-modelling combined with semantic vector similarity can serve as an effective means for identifying binary structures. The subsections that follow are organized around broad thematic clusters—not to showcase oppositions directly, but to examine how well these methods capture meaningful groupings and separations within the text. These reflections aim to determine whether the patterns observed can be trusted as indicators of semantic opposition or merely statistical noise. This critical approach is particularly important given the interpretive risks of applying unsupervised machine learning methods to literary texts, where meaning is deeply context-bound and not always aligned with statistical patterns.

4.1. Exploring Topics Dealing with Hobbits and Men

In his legendarium Tolkien has created a variety of different races with their own cultures and intricacies (*History of Middle-Earth: The peoples of Middle-Earth* (Tolkien, 1996)). In this paper I touch upon three of them: Men, Hobbits and Ents. Starting with the case of Hobbits and Men in Table 6, it can be observed that topics that deal with Frodo (the main character) are opposed to

Men. Now, Men in *the Lord of the Rings* and Tolkien's legendarium while they physiologically correspond to the average human, their culture does not correspond in the same way. Tolkien's influece from Anglo-Saxon romances and the Classics is projected onto his depiction of Men— heroic, valiant and kingly, yet easily corrupted (Fimi, 2008; Fernández Camacho, 2023). Hobbits on the other hand represent an idealised and non-technological England (Weidner, 2002; Turner, 2014; *Letters 230*). Turner (2014) further states:

> Hobbits are stolidly unheroic, but they have a streak of deep-seated courage which enables them to take on responsibilities greater than their physical stature would suggest, allowing them to lead the modern reader into contact with traditional heroic attitudes and situations which might otherwise seem completely alien.

This is can be also observed in the language the two races use when they speak. Figure 6 below is taken from the Digital Tolkien Project, an online scholarly source that provides a plethora of different quantitative information on Tolkien's works. This particular figure is a rug plot of where each not-contraction occurs in the three volumes of *the Lord of the Rings*, separated by chapters. Hobbits appear consistently from the start of *the Fellowship of the Ring* (henceforth FotR) until its end, however as the narratives break from there onwards there is a very noticeable difference in the usage of not-contractions whenever Hobbits appear. In *the Two Towers* the contractions appear from chapter 3 to 5 and from 9 to 11 when two of our four Hobbits—Merry and Pippin— appear. The chapters before and in between only deal with Men, and have 0 not-contractions. Following chapter 11, the narrative shifts to only Frodo and Sam—our other two Hobbits—and Sméagol—a creature that used to be a Hobbit. For the remainder of *the Two Towers* (henceforth TT) we follow these three characters' journey. An interesting point of note here is in chapters 16 to 18, as they encounter Men and the frequency of the contractions becomes much lower as it is visible by the amount of occurrences. In *the Return of the King* (henceforth RotK) for the first 10 chapters where Merry and Pippin appear, the contractions are much sparser as the characters are separated and meet again in chapter 8 and the contractions increase. Again Hobbits appear in chapters 11 to 14, and from chapter 16 to the end. This stylistic difference further supports the opposed topics of Table 6, with evidence that is usually part of stopword removal and not included in the topic-modelling process.
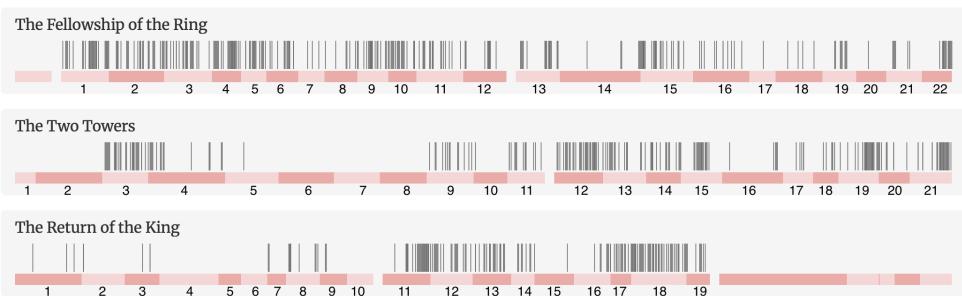
Figure 6: Not-contractions throughout *the Lord of the Rings* as taken from The Digital Tolkien Project's Glossary

To add to this, topic 11 that particularly deals with Frodo's journey (TT chapters: 12-21 and RotK chapters: 11-19), and is opposed to topics that deal with Men, has one more dimension of opposition within it that is irrelevant to style. Frodo's quest objective is to destroy the One Ring. He proves to be resilient against the corruption of the One Ring in contrast to Men that throughout the history of Tolkien's legendarium were easily influenced by it, and failed in its destruction. Now, a knowledgeable reader of the work might argue with me here, that Frodo in the end did not actually succeed in destroying the Ring, but it was rather, a coincidence, or the result of his pity for Sméagol that led to the destruction of the Ring. To which, I agree, yet even though this is indeed the case, Frodo withstanding the will of the Ring up until the last moment, while other characters fail (Boromir), showcase the opposition in resilience between them.

4.2.   Lexical Distances in Similar and Opposing Topic Pairs

A part of this overall study is to also identify binary oppositions at a lexical level. In this subsection I examine the relationship of terms in opposed and similar topics. While I cannot examine every single pair in this paper, I select two from each. The goal of this examination is to evaluate if terms semantically follow their topic distances, by showing opposing words.

Starting from high-similarity topic pairs, we have the topic pair of 1 and 2 also seen in Table 5 dealing with Hobbits. Topic 1 that deals with Hobbits and Merrymaking contains terms like: bithday, supper, food, good, party, eat. This topic is mostly respresented by chapters at the start of the work (see Table 4) and is classified as positive. On the other hand, Topic 2 is only represented by chapter 61 (or RotK 18) *the Scouring of the Shire*. This topic as briefly discussed in section 3., is negatively classified. The interesting point between these two topics is that they portray departure (topic 1) and return (topic 2) of the Hobbits from and to the Shire. There is an opposition that appears here, as when the Hobbits leave the Shire, they leave it in a peacefull state, untainted by the evils of the world, yet when they return back the find it in the exact opposite state. The high-similarity appearing here can be attributed to the fact that both topics deal with Hobbits containing relatively similar vocabulary, and as discussed in the previous subsection language and style are changed. Taking a closer look at the terms in Figure 7 below, the term 'weapons' in Topic 2 is opposed or dissimilar to most of the words in Topic 1, especially words like 'hobbits' and 'good'. From an evaluating standpoint this shows that the Word2Vec model functions well in measuring semantic distances. Similarly, other words, like 'village' and 'hobbiton' show high-similarity which is also to be expected. A few key points can be concluded here, high-similarity can mean similar or semantically close themes, but it can also lead to opposing ones which questions the primary argument that low-similarity means oppositions.
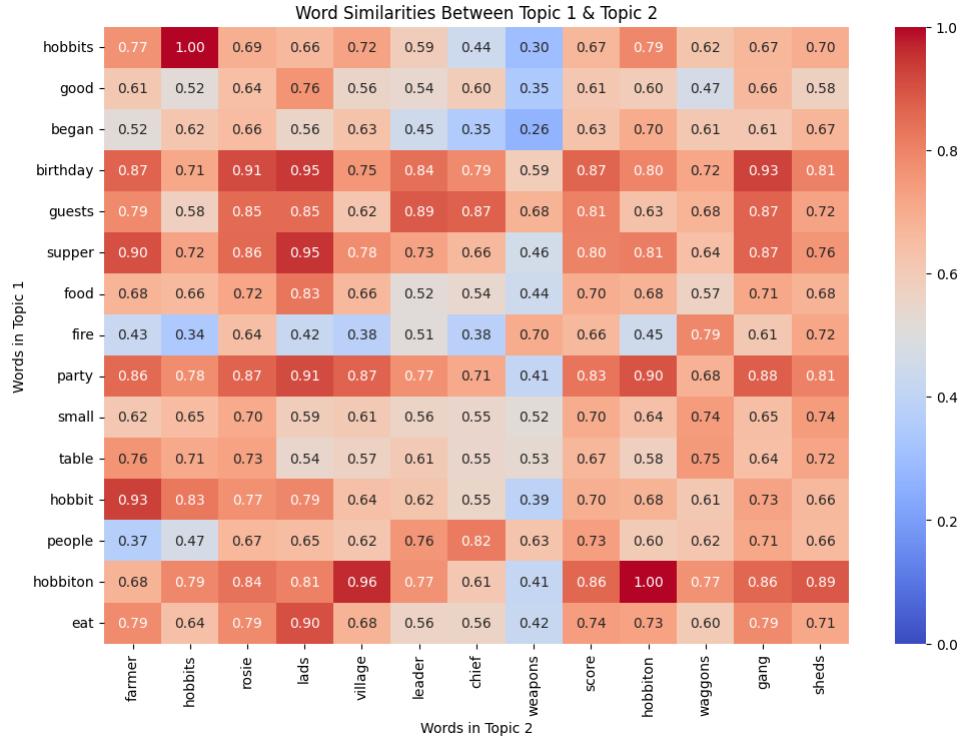
Figure 7: Top 15 term distances of topics 1 and 2 with similarity score of 0.95

Continuing to another high-similarity pair, topic 9 and 14 (Figure 8). Now topics 9 and 14 both deal with Men, but again we have a negative and a positive one. Topic 9 labeled Gondor and Steward, is most represented by chapters 50 (Rotk: 7) *the Pyre of Denethor* and 47 *the Siege of Gondor* (Rotk: 4). As briefly introduced in section 3., topic 9 deals heavily with negative concepts like death, suicide and war. This is visible in terms like: 'burn', 'death', 'steward' and 'city'. Topic 14 on the other hand labeled Love and Healing, deals precisely with these two concepts and it can be seen in the terms: 'healing', 'lady', 'lord', 'care'. Topic 14 is mostly represented by chapter 58 (RotK 15) *the Steward and the King* and chapter 51 (RotK 8) *the Houses of Healing*. Both topics connect to the character of Faramir. In the documents included in topic 9 he is nearly killed by his father in a moment of despair before killing himself, and in the documents included in topic 14 he is healed, falls in love and takes his father's place as steward when the city is healed. This case, is very similar to the one above, high-similarity topics leading to opposing themes. The high-similarity that arises in this topic pair again could be attributed to similar vocabulary and register. Of course, it can also be said that, both topics deal with Faramir and Men in general and thus they are similar. However, the more nuanced analysis would suggest otherwise.
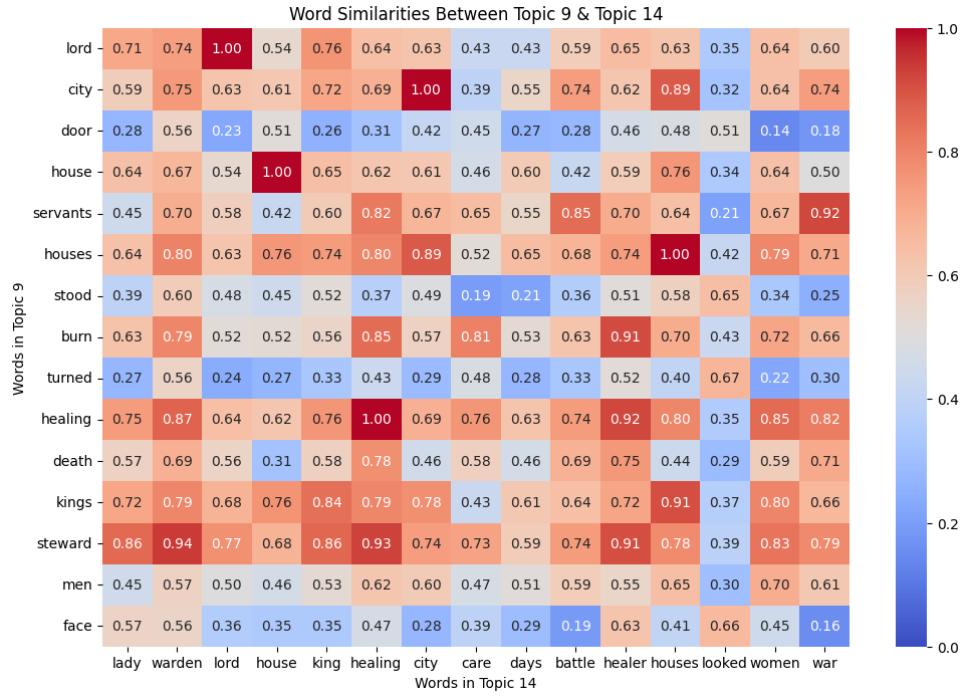
Figure 8: Top 15 term distances of topics 9 and 14 with similarity score of 0.96

Moving on to low-similatity topics, I first examine topics 11 and 14 (Figure 9). Topic 11 deals with Hobbits, but particularly Frodo's quest to destroy the Ring. The most representantive chapters are 33 (TT: 12) *the Taming of Sméagol* and 34 (TT: 13) *the Passage of the Marshes*. Frodo's quest is that of suffering, physical and more importantly psychological, and as such this topic is classified as negative. However, the terms do not so clearly express this. The terms 'precious' and 'master' relate to Frodo's relationship with Sméagol, but the rest are not that easily distinguished. Nevertheless, there is a clear opposition between these two topics. Topic 14, as discussed above, deals with healing and characters being able to mend their physical and psychological wounds. Frodo even after the quest was done and the Ring destroyed, could not be psychologically healed. The trauma induced by the Ring was too much, as opposed to Faramir or other characters being healed in the documents that represent topic 14 (Éowyn and Merry). The overall discussion of Frodo or Hobbits being opposed to Men as above, is still applicable in this case, however if we look at isolated instances we can delve deeper into specific oppositions that appear through measuring similarity. Yet, and this is an issue that arises when looking at the relationship between terms, as can be observed by the heatmap below, the terms are not opposed with each other, nor are they similar. They are vastly irrelevant, especially the cases where cosine similarity is the lowest: felt and house, felt and city, head and war. The words 'hobbits' and 'lord' do show the opposition discussed above, but this is not the norm of the relationship between the terms.
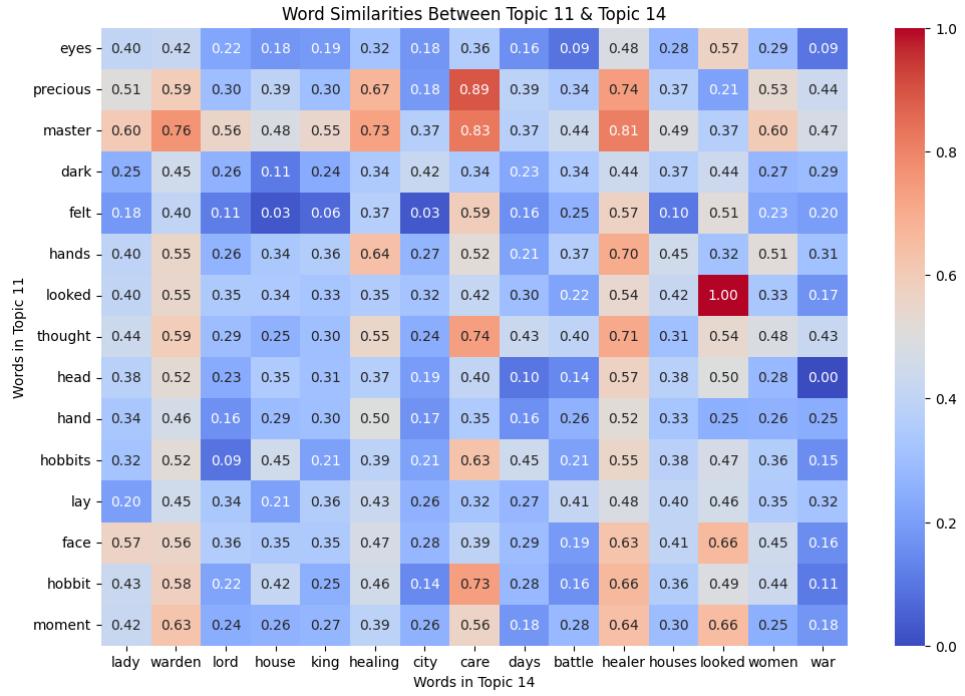
Figure 9: Top 15 term distances of topics 11 and 14 with similarity score of 0.62

Finally, the last pair I examine is topics 7 and 12. Topic 7 deals predominantly with Ents. They are sentient trees that shepherd the forests and guard them. There has been extensive research on Tolkien's portrayal of nature and his personal love for it (Saguaro & Thacker, 2013; Curry, n.d.; Tolkien, 2022), with this in mind as well as the context of the topic itself I classified it as positive. The most representative documents for this topic are in chapters 26 (TT: 4) *Treebeard* and 27 (TT: 5) *the White Rider*. Trees and nature in general in *the Lord of the Rings* suffer and is easily affected by evil forces like orcs, as they cut trees for war which very well opposes topic 12. Topic 12 deals with orc battles, with most representative chapters being 54 (RotK: 11) *the Tower of Cirth Ungol*, 34 (TT: 13) *the Passage of the Marshes* and 17 (FotR: 17) *the Bridge of Khazad-Dûm*. All three of the chapters mentioned deal with battles or armies, orcs and evil forces—Ringwraiths, goblins, Balrogs. Now, taking a closer look at the term distances a similar pattern as the previous topic pair can be observed. Terms with very low-similarity scores are irrelevant, e.g. 'trees' and 'cried', or 'heard' and 'sword'. The terms overall seem to have almost no semantic relevance either, yet nature versus evil is part of Tolkien's major oppositions in *the Lord of the Rings*.
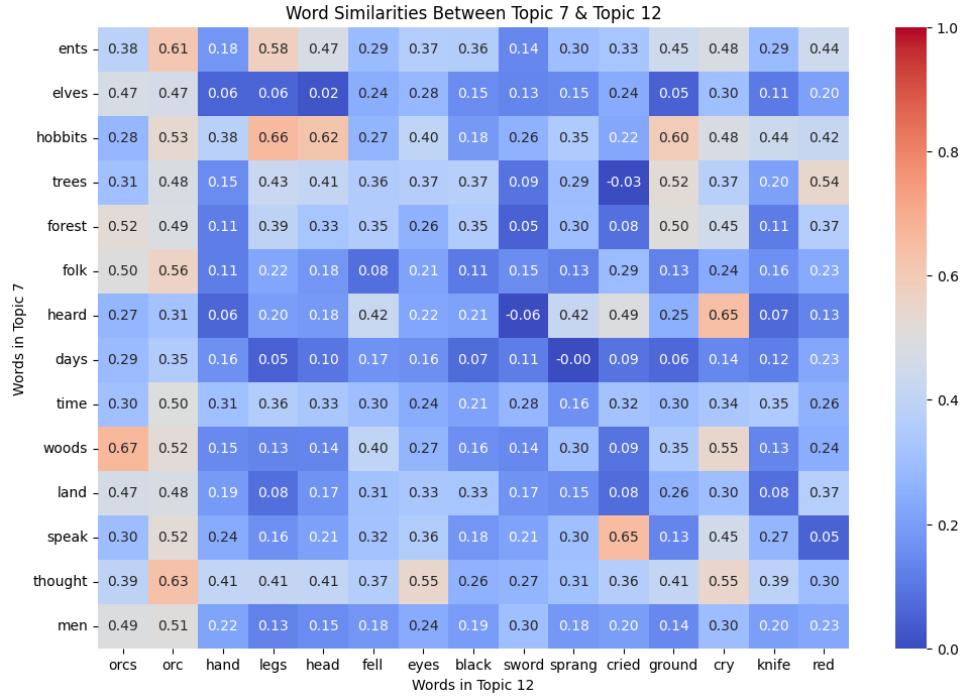
Figure 10: Top 15 term distances of topics 7 and 12 with similarity score of 0.52

There are two major points that can be concluded from the above examinations. First, oppositions can be found in both ends of the spectrum. Highly similar topics (as were grouped in 3.) score high-similarity because they are indeed similar, i.e. Hobbits being in the Shire (topics 1,2 and 3). However, oppositions can be found in high-similarity topic pairs. The difference from the oppositions of low-similarity pairs, is that these oppositions are found within the same concepts—positive Hobbits and negative Hobbits. Whereas in low-similarity pairs, it is about two separate concepts that are opposed by default—trauma and healing. Measuring inter-topic distance through Tolkien's Word2Vec model seems to be providing significant insight into these conceptual relations. However, distance on a lexical level cannot be measured well with this method as of this moment. Secondly, Tolkien's Word2Vec model can also catch differences in style even when the style appears more pronounced in words that are by default in stopwords, which further solidifies its usability in identifying thematic oppositions.

5. Conclusion

This study presented a primary quantitative method for identifying binary oppositions in *the Lord of the Rings*, exploring how they contribute to the work's structural and thematic rhythm. By creating a Tolkien-specific Word2Vec model and combining it with BERTopic topic-modelling technique, I was able to identify distinct thematic oppositions, such as healing and trauma, as well as more subtle oppositions within Tolkien's narrative. The model, trained exclusively on Tolkien's

texts, enabled a deeper semantic alignment, revealing patterns that non-literary models could not capture.

The integration with BERTopic allowed for a detailed examination of inter-topic distances, showcasing that binary oppositions do not always appear through low-similarity scores. In fact, high-similarity topic pairs often revealed oppositions within the same theme, for instance positive and negative situations concerning Hobbits. Additionally, examining low-similarity pairs revealed more distinct oppositions, such as nature versus war or simple Hobbits against higher Men. These results suggest that Tolkien's use of oppositions is intricately woven into both the structure and semantics of the narrative, and it can be further examined as a form of rhythm.

While the findings offer insight into the possible contextual rhythm of oppositions, this paper also points to areas for further exploration. Analysing the recurrence and distribution of these oppositions across other works like *the Silmarillion* could provide stronger evidence for their role in Tolkien's larger legacy. This would help solidify the idea that these thematic oppositions are not isolated to *the Lord of the Rings* but are a deliberate stylistic choice in his entire legendarium.

Bibliography

[1] Albrecht, J., Ramachandran, S. & Winkler C. (2020) *Blueprints for Text Analytics Using Python: Machine Learning Based Solutions for Common Real World Applications*. O'Reilly Media, Incorporated.

[2] Alden, S. (2021). *Hyphens as Sub-Lexical Morphemes in The Hobbit*. Mallorn: The Journal of the Tolkien Society, 62, 36-40. https://www.jstor.org/stable/48650605

[3] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). *Latent dirichlet allocation*. Journal of Machine Learning Research, 3, 993-1022. https://doi.org/10.5555/944919.944937

[4] Curry, P. (1997). *Defending Middle-earth: Tolkien, myth and modernity*. St. Martin's Press.

[5] Curry, P. (n.d.). *Tolkien and Nature*. The Tolkien Estate. https://www.tolkienestate.com/writing/patrick-curry-tolkien-and-nature/

[6] Ding, L., Wang, L., Liu, X., Wong, D. F., Tao, D., & Tu, Z. (2021). *Progressive Multi-Granularity Training for Non-Autoregressive Translation*. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.findings-acl.247

[7] Doyle, M. (2019). *Utopian and Dystopian Themes in Tolkien's Legendarium*. Lexington Books. https://doi.org/10.5040/9781978738928

[8] Fernández Camacho, P. (2023). *Elven-Latin and Semitic Adûnaic: Linguistic, Religious, and Political Strife in Tolkien's Island of Númenor*. Journal of Inklings Studies, 13(1), 67-86. https://doi.org/10.3366/ink.2023.0176

[9] Flieger, V. (2002). *Splintered light: Logos and language in Tolkien's world* (2nd ed.). Kent State University Press.

[10] Fimi, D. (2022). *Tolkien, Race, and Racism in Middle-earth* (2022) by Robert Stuart. Journal of Tolkien Research, 15(1), 3.Available at: https://scholar.valpo.edu/journaloftolkienresearch/vol15/iss1/3

[11] Goldberg, Y., & Levy, O. (2014). *word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method*. arXiv preprint arXiv:1402.3722.

[12] Grootendorst, M. (2022). *BERTopic: Neural topic-modelling with a class-based TF-IDF procedure*. arXiv preprint arXiv:2203.05794.

[13] Holmes J.R. (2014). *The Lord of the Rings*. In Lee, S.D. (Ed.), *A Companion to J. R. R. Tolkien* (2020) (pp. 116-126) John Wiley & Sons

[14] Le Guin, U. (2001). *Rhythmic pattern in The Lord of the Rings*. In K. Haber (Ed.) Meditations on Middle-earth

[15] Lee, S. D. and John Wiley & Sons (eds). (2022). *A Companion to J.R.R. Tolkien*. Second edition. Hoboken, NJ:John Wiley & Sons.

[16] Mikolov, T., Chen, K., Corrado, G. and Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. arXiv preprint arXiv:1301.3781.

[17] Misra, H., Yvon, F., Cappé, O., & Jose, J. (2011). *Text segmentation: A topic-modelling perspective*. Information Processing & Management, 47(4), 528-544. https://doi.org/10.1016/j.ipm.2010.11.008

[18] Parker, Robert, Graff, David, Kong, Junbo, Chen, Ke, and Maeda, Kazuaki (2011). English Gigaword Fifth Edition Linguistic Data Consortium doi:10.35111/WK4F-QT80. https://catalog.ldc.upenn.edu/LDC2011T07 .

[19] Pennington, J., Socher, R. & Manning, C.D. (2014). *GloVe: Global Vectors for Word Representation*. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages1532-1543, Doha, Qatar. Association for Computational Linguistics.

[20] Riedl, M., & Biemann, C. (2012). *Text Segmentation with Topic Models*. Journal for Language Technology and Computational Linguistics, 27(1), 47-69. https://doi.org/10.21248/jlcl.27.2012.158

[21] Sahoo, S., & Maiti, J. (2025). *Variance-Adjusted Cosine Distance as Similarity Metric*. arXiv preprint arXiv:2502.02233.

[22] Saguaro, S., & Thacker, D. C. (2013). *Tolkien and Trees*. In J. R. R. Tolkien (pp. 139-155). Macmillan Education UK. https://doi.org/10.1007/978-1-137-26401-5_10

[23] Sentence transformers all-minilm-L6-V2 (n.d.) Available at: https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

[24] Tabata, T. (2020) *Latent topics in British classic fiction: using LDA to classify texts into meaningful groups*. Text Mining and Digital Humanities, 2019, 47-58

[25] Tauber, J. (2025a), Digital Tolkien Project: A scholarly project focused on Tolkien from both a corpus linguistic and digital humanities perspective. https://digitaltolkien.com/

[26] Tauber, J. (2025b), Tolkien Glossary. https://glossary.digitaltolkien.com/

[27] Tolkien, J. R. R. (1936). *Beowulf: The monsters and critics*. Humphrey Milford for the British Academy.

[28] Tolkien, J. R. R. (1937). *The Hobbit* (1st ed.). London: George Allen & Unwin.

[29] Tolkien, J. R. R. (1945). *Leaf by Niggle*. In The Dublin Review, no. 432, Jan., pp. 46-61. Dublin: Burns Oates & Washbourne.

[30] Tolkien, J. R. R. (1949). *Farmer Giles of Ham*. London: George Allen and Unwin.

[31] Tolkien, J. R. R. (1954). *The Fellowship of the Ring*. The Lord of the Rings. London: George Allen & Unwin.

[32] Tolkien, J. R. R. (1954). *The Two Towers*. The Lord of the Rings. London: George Allen & Unwin.

[33] Tolkien, J. R. R. (1955). *The Return of the King*. The Lord of the Rings. London: George Allen & Unwin.

[34] Tolkien, J. R. R. (1962). *The Adventures of Tom Bombadil*. London: George Allen & Unwin.

[35] Tolkien, J. R. R. (1967). *Smith of Wootton Major*. Houghton Mifflin

[36] Tolkien, J. R. R. (1976). *The Father Christmas Letters*. London: George Allen & Unwin.

[37] Tolkien, J. R. R. (1993) *Morgoth's Ring, in The History of Middle Earth* (Tolkien C., Ed.). HarperCollins

[38] Tolkien, J. R. R. (1996) *The Peoples of Middle-earth, in The History of Middle Earth* (Tolkien C., Ed.). Harper-Collins

[39] Tolkien, J. R. R. (1998). *Roverandom* (Scully, C. & Hammond, W. G, Eds.). HarperCollins.

[40] Tolkien, J. R. R. (2015). *The Story of Kullervo*. (C. Tolkien, Ed.). HarperCollins.

[41] Tolkien, J. R. R. (2017).*The Children of Húrin* (Tolkien, C. Ed.). HarperCollins.

[42] Tolkien, J. R. R. (2017). *Beren and Lúthien* (Tolkien, C. Ed.). HarperCollins.

[43] Tolkien, J. R. R. (2018). *The Fall of Gondolin*. (C. Tolkien, Ed.). HarperCollins.

[44] Tolkien, J. R. R. (2022). *Nature of Middle-Earth*. (Hostetter, C. F., Ed.). HarperCollins Publishers.

[45] Tolkien, J. R. R. & Carpenter, H. (2014). *Letters of J. R. R. Tolkien*. Houghton Mifflin Harcourt Publishing Company.

[46] Tolkien, J. R. R. & Gordon, E. V. (1925). *Sir Gawain & the Green Knight*. Clarendon Press.

[47] Tolkien, J. R. R. & Tolkien C. (1977). *The Silmarillion* (Tolkien, C. Ed.). HarperCollins.

[48] Tolkien, J. R. R., Higgins, A., & Fimi, D. (2020). *Secret Vice*. HarperCollins Publishers Limited.

[49] Turner, A. (2007). Prose Style. In Drout, M. D. C. (Ed.), *J. R. R. Tolkien Encyclopedia* (2013)(pp. 545-546) Routledge

[50] Turner, A. (2014). *Style and Intertextual Echoes*. In Lee, S.D. (Ed.), A Companion to J. R. R. Tolkien (2020) (pp. 374-385) John Wiley & Sons

[51] Wang, W., Wei, F., Dong, L., Bao, H., Yang, N., & Zhou, M. (2020). *Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers*. Advances in neural information processing systems, 33, 5776-5788.

[52] Weidner, Brian N. (2002). *Middle-earth: The Real World of J.R.R. Tolkien*, Mythlore: A Journal of J.R.R. Tolkien, C.S. Lewis, Charles Williams, and Mythopoeic Literature. 23(4), 7. Available at: https://dc.swosu.edu/mythlore/vol23/iss4/7

6. Appendix

Table 7: Top 10 similar words for example words in Tolkien and GloVe Models

Example	Tolkien Model	GloVe Model
light	gleam ....(0.77) bright ....(0.75) star ....(0.75) dimmed ....(0.75) shine ....(0.74) pale ....(0.74) sunlight ....(0.73) shining ....(0.72) brighter ....(0.71) visible ....(0.71)	air ....(0.78) lights ....(0.76) heavy ....(0.75) lighter ....(0.75) surface ....(0.75) display ....(0.75) bright ....(0.75) visible ....(0.74) ground ....(0.74) color ....(0.73)
ring	finger ....(0.71) ruling ....(0.69) sauron ....(0.69) burden ....(0.67) wield ....(0.64) weight ....(0.62) chain ....(0.62) stroke ....(0.62) dreadful ....(0.62) enemy ....(0.62)	rings ....(0.89) hanging ....(0.69) triangle ....(0.69) gate ....(0.68) sword ....(0.68) wire ....(0.67) hidden ....(0.66) mask ....(0.66) inside ....(0.65) into ....(0.65)
party	finished ....(0.96) bag-end ....(0.96) arrived ....(0.96) attention ....(0.94) polite ....(0.93) week ....(0.93) eaten ....(0.93) tea ....(0.93) actually ....(0.92) important ....(0.92)	democratic ....(0.88) opposition ....(0.88) parties ....(0.84) coalition ....(0.83) leader ....(0.83) election ....(0.83) candidate ....(0.83) elections ....(0.82) socialist ....(0.82) liberal ....(0.81)

Table 8: Top 10 similar words for example words in Tolkien and GloVe Models

Example	Tolkien Model	GloVe Model
destroy	suffer . . . . (0.97)	destroying . . . . (0.84)
	perish . . . . (0.96)	destruction . . . . (0.82)
	attempt . . . . (0.96)	protect . . . . (0.78)
	fulfilment . . . . (0.95)	terrorists . . . . (0.76)
	abide . . . . (0.95)	enemies . . . . (0.76)
	restrain . . . . (0.95)	rid . . . . (0.75)
	vengeance . . . . (0.95)	invade . . . . (0.75)
	rule . . . . (0.95)	kill . . . . (0.74)
	fulfilled . . . . (0.95)	seize . . . . (0.74)
	achieved . . . . (0.95)	dismantle . . . . (0.73)
quest	betray . . . . (0.96)	ultimate . . . . (0.77)
	folly . . . . (0.96)	dream . . . . (0.76)
	release . . . . (0.96)	destiny . . . . (0.73)
	test . . . . (0.96)	dreams . . . . (0.71)
	desires . . . . (0.96)	survival . . . . (0.69)
	choice . . . . (0.95)	success . . . . (0.69)
	vain . . . . (0.95)	pursuit . . . . (0.69)
	fulfilled . . . . (0.95)	glory . . . . (0.69)
	doubtless . . . . (0.95)	opportunity . . . . (0.68)
	nonetheless . . . . (0.95)	struggle . . . . (0.68)