# Learning Folktale Plot Representations

Anonymous ACL submission

## Abstract

Folktales possess both historical and literary significance as they offer a glimpse into the culture and traditions of the communities that created and passed them down through the generations. Folklore scholars have been meticulously analyzing and classifying folktales based on their type and motifs, however, the automatic identification and discovery of folktale types remains an area of ongoing study. We propose a computational approach for identifying folktale types by utilizing an online library of folklore texts and a neural embedding model. Our method semantically encodes the texts, resulting in tale representations that capture the similarities between tale plots. We validate this by visualizing the representations using t-SNE and applying K-means clustering, along with human evaluation.

## 1 Introduction

Folktales are fictional narratives that originate in a culture's oral tradition (Ashliman, 2004), serving various purposes such as educating, disciplining, or entertaining. These narratives can take on different forms, ranging from tales and proverbs to jokes, and are considered an important subject of study in literature and history because they play a crucial role in preserving cultural heritage and traditions. Almost all fairy tales fall under the category of folktales, which often blend elements of fantasy and reality, reflecting themes of daily life interweaved with magical creatures, miraculous events, and impossible feats.

The field of study dedicated to folktales is called Folklore Studies, which involves the collection, preservation, and examination of these stories. With the advancements of technology and internet connectivity, researchers and enthusiasts have compiled and made available digital folktale collections such as the Dutch Folktale Database (Meder et al., 2016), the Multilingual Folk Tale Database (MFTD)[1], the Archive of Portuguese Legends (APL)[2], SurLaLune[3], and the Folklore and Mythology Electronic Texts (Folktexts)[4]. Some of these collections are aimed at facilitating scholarly research in humanities and sociology, while others are meant for readers to browse a wide range of tales.

A variety of indexing systems are used to classify and organize folktales. One prominent example is the Aarne-Thompson-Uther (ATU) tale type classification system, which features 2400 distinct index nodes arranged in a hierarchical structure of types, sub-types, etc., and aims to highlight the similarities between tales. The classification system was initially developed by Antti Aarne in 1910, later updated and expanded by Stith Thompson in 1928 and 1961, and further revised and expanded by Hans-Jörg Uther in 2004.

Despite the seemingly endless variety of stories that mankind can tell, folktales seem to be constructed from a limited set of patterns and elements. The ATU folktale type index is based on these recurrent motifs, narrative concepts and plots, grouping different versions of the same tale under a single ATU category, making it a useful tool for analysis in the field of folkloristics (Dundes, 1997). Having said that, it's important to note that the ATU classification system has faced criticism for censorship, as Thompson excluded a significant amount of material deemed sexual or 'obscene' (Goodwin, 1995). Furthermore, it has been criticized for disproportionately featuring tales from Eurasia and North America, and not giving enough representation to Central Asia, where new forms of folktales may emerge.

Computational Folkloristics has been the sub-

[1]MTFD: www.mftd.org
[2]APL: www.lendarium.org
[3]SurLaLune: www.surlalunefairytales.com
[4]Folktexts: https://sites.pitt.edu/~dash/folktexts.html

ject of research on automatic classification of folktales. Studies have employed various machine learning techniques such as Support Vector Machines (SVM) to classify works from the Dutch Folktale Database (Nguyen et al., 2012), and deep neural networks, such as Hierarchical Attention Networks (HAN) to detect ATU types using the MFTD collection (Pompeu et al., 2019).

Previous research in this field has operated under the assumption that the classes of folktales are fixed, preventing the discovery of new types. Our goal is to learn representations of folktales based on their plot and generate numerical vectors that encode the story's plot and character actions, allowing us to detect and quantify the similarities between tales. We achieve this by training a RoBERTa (Liu et al., 2019) model using a contrastive loss function. This can aid in comparative folktale analysis, helping folklorists to classify newly obtained tales or revise existing classifications, as identifying tale motifs is a time-consuming and labor-intensive task. Additionally, our approach can assist in the evaluation of automatically generated stories by providing a fast and transparent method of determining whether the generated text conforms to a known class of folktales.

We utilize the Folktexts collection of myths and tales, which is publicly available and is considered one of the most comprehensive and well-organized collections of online folklore currently available (Rewis, 2020). We collect the texts following the methodology of Hagedorn and Darányi 2022, however, in our approach, instead of only utilizing the single ATU type assigned to each tale, we also take advantage of Dr. Ashliman's notes and annotations and include any additional ATUs that are mentioned. This is significant as it allows us to capture the multiple plots that may be present in a single text, as a tale can weave one or more ATU types in its narrative.

## 2 Related Work

Nguyen et al. 2012 developed a classifier for folktales using data from the Dutch Folktale Database which were organized according to the following narrative genres: Fairy tales, Legends, Saint's legends, Urban legends, Personal narratives, Riddles, Situation puzzles, Jokes, and Songs. The authors employed a Support Vector Machine (SVM) (Boser et al., 1992) to perform this task. Another classification approach was proposed by Pompeu et al. 2019 who used the MFTD folktale collection and a modified Hierarchical Attention Network (HAN) (Yang et al., 2016) extended with a K-Nearest Neighbors (KNN) (Wang et al., 2017) component that jointly predicts the tale's first-level ATU type and its second-level sub-type. Both of these works operate on the assumption that the classes of folktales can't change, and hence, don't allow the discovery of new types. In contrast, our work aims to learn meaningful folktale representations that will enable the discovery of new types when our model is used as a text encoder.

In their 2020 thesis, Rewis presented an approach that can fulfill our need for new tale type discovery by using a document-to-vector model (Doc2Vec) (Le and Mikolov, 2014) to encode tales. Although they did not aim to construct a classifier, their methods have served as inspiration for our work, as we take it one step further by applying supervised learning to create a more robust method of learning tale representations.

The work of Hagedorn and Darányi 2022 focuses on faithfully extracting the folktales and their ATU type indices from D. L. Ashliman's Folktexts collection. We build upon their work by including all ATU types for each tale, as they are presented on Ashliman's website, making use of Ashliman's full expertise and effort. This enables us to capture the multiple plots that may be present in a single text, as tales often incorporate more than one ATU type in their narratives.

## 3 Method

The ATU folktale classification system groups closely related folktales within a type and is hierarchically organized. Figure 1 shows the 7 first-level and the 43 second-level ATU types and their index range, while Figure 2 shows how the first-level ATU type of Animal Tales expands into multiple sub-types.

In Table 1, we present a comparison of tales from different ATU types. The left-most column contains snippets from tales in the 510A ATU index, which share a common theme of a heroine who is mistreated by her stepmother and stepsisters, but ultimately, by the test of the slipper, achieves happiness by marrying into royalty. In contrast, the right-most column contains tales from the 275 ATU index, which feature the theme of a competition between two contestants, where the characters are drawn from the animal domain and are typically

| **Cinderella** (ATU 510A) | **The Hare and the Tortoise** (ATU 275A) |
|---|---|
| *... So they called Cinderella, and when she heard that the prince was there, she quickly washed her hands and face. She stepped into the best room and bowed. The prince handed her the golden slipper, and said, "Try it on. If it fits you, you shall be my wife". She pulled the heavy shoe from her left foot, then put her foot into the slipper, pushing ever so slightly. It fit as if it had been poured over her foot. As she straightened herself up, she looked into the prince's face, and he recognized her as the beautiful princess.*<br><br>Jacob and Wilhelm Grimm, 1812. | *... "Let us make a match" replied the tortoise. "I will run with you five miles for five pounds, and the fox yonder shall be umpire of the race." The hare agreed, and away they both started together. But the hare, by reason of her exceeding swiftness, outran the tortoise to such a degree, that she made a jest of the matter; and finding herself a little tired squatted in a tuft of fern, that grew by the way, and took a nap; thinking that if the tortoise went by, she could at any time fetch him up, with all the ease imaginable.*<br><br>Aesop, translated by S. Croxall, 1831. |
| **The Hearth-Cat** (ATU 510A & 480) | **Why Does the Buffalo Walk Slowly and Tread Gently?** (ATU 275) |
| *... The king inquired who was the next to try on the slipper, and asked the mistress if there was any other lady left in her house who could fit on the slipper. The schoolmistress then said that there only remained a hearth-cat in her house, but that she had never worn such a slipper. The king ordered the girl to be brought to the palace, and the mistress had no alternative but to do so. The king himself insisted on trying the slipper on the girl's foot, and the moment she put her little foot into the slipper and drew it on, it fitted exactly. The king then arranged that she should remain in the palace and married her.*<br><br>C. Pedroso, translated by H. Monteir, 1882. | *... One day the hare said to the buffalo, "Let us try a race together and settle this quarrel once for all."*<br><br>*The buffalo was well contented with the proposal, and they agreed to race one another. When the day came, the hare, putting his ears back, started the race. He ran so fast that you might have said he was flying upon the ground.*<br><br>*But the buffalo was a match for him. He went thundering away, his hoofs splashing the mud and raising seas of mire.*<br><br>M. Gaster, 1915. |

Table 1: Comparison of tales from different ATU tale type indices. Left column: tales from the 510 ATU index. Right column: tales from the 275 ATU index.

unequal in terms of strength or ability.

### 3.1 Data

We gather 2400 tales from the Folktexts collection, of which only 1671 are tagged with one or more ATU types. Our collection of folktales includes 296 unique leaf-level ATU types, 262 ATU types immediately preceding the leaf-level (parent nodes), 42 second-level, and 7 first-level. We divide the annotated data into a training set and a test set, with the training set comprising 1510 texts with 298 unique ATU types and the test set including 161 texts, with 128 unique ATU types. Among the tales in the test set, we handpicked 11 well-known tales to function as our landmark tales, such as "The Hare and the Tortoise", "Cinderella", "Puss in Boots", etc.

### 3.2 Approach

We use sentence embedding techniques, such as SBERT (Reimers and Gurevych, 2019), to represent folktales as vectors that capture the semantic information of entire sentences. These models are designed to capture a range of semantic relationships between sentences, such as similarity, contradiction, and entailment. Sentence embeddings can easily capture the style but do not necessarily reflect the cultural background or thematic family of each tale. Fortunately, these algorithms can be trained for various objectives, and one such way is by using supervised learning with a contrastive loss function, which can produce sentence embeddings that are more semantically rich.

In our experiments, we use four ATU granularity

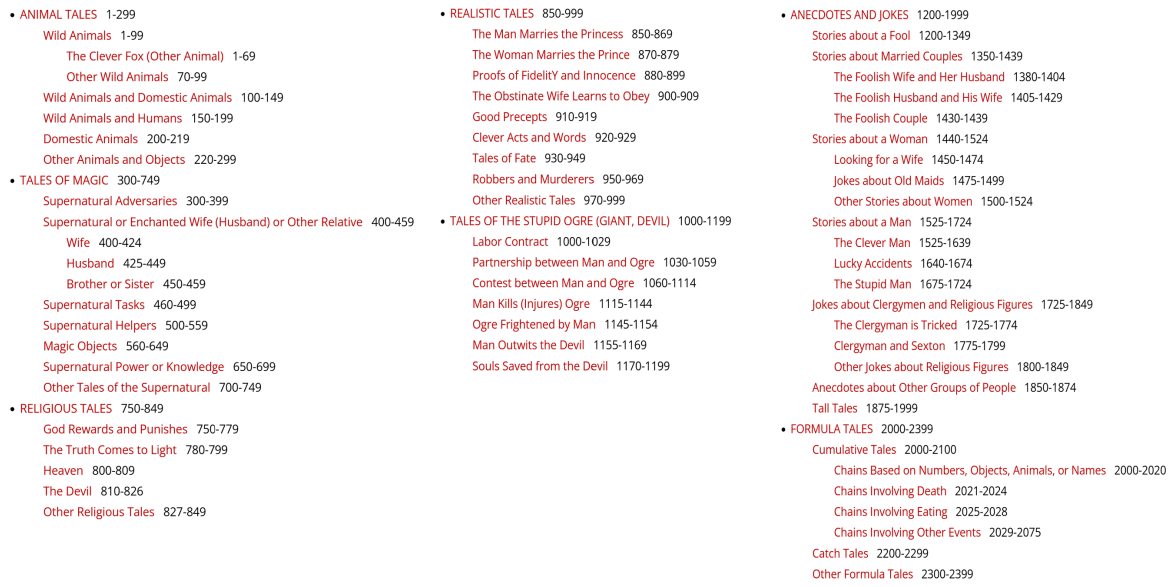


Figure 1: The top-most three levels of the ATU folktale type index hierarchy. We are focusing on the first level (Animal Tales, Tales of Magic, etc.) and the second level (Wild Animals, Wild Animals and Domestic Animals, etc.). Image from http://www.mftd.org/index.php?action=atu

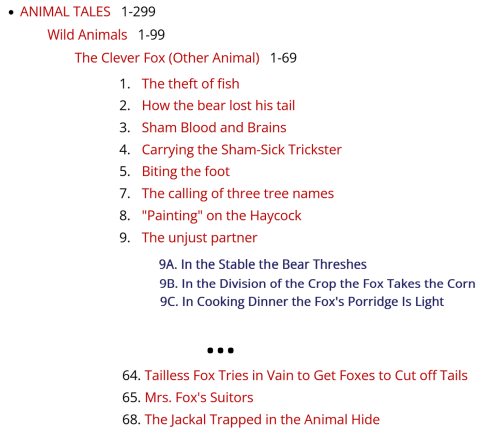


Figure 2: Expansion of the Wild Animals ATU folktale type. Besides the two first levels of the ATU hierarchy, we are also interested in the leaf level (e.g. 9A, 9B, and 9C), as well as the level immediately preceding the leaf-level (e.g. 9 The Unjust Partner). Image from http://www.mftd.org/index.php?action=atu

levels to better understand how our selected model captures plot information within its embeddings. The first-level (coarse), being the most general level of the ATU hierarchy, points to a tale's genre rather than a specific plot. The second-level (mid) narrows down the topic range to specific character types, but the plots still remain quite abstract. The level immediately preceding the leaf-level (simple), where plots and characters are sufficiently defined. Lastly, the leaf-level (fine) where plot variations are more nuanced. We believe that this level actually splits the same underlying plot into multiple types based on the presence of specific characters and actions, therefore the ideal granularity to learn folktale plot representations is the simple level.

We perform an additional training step using the ATU indices as tale labels on top of the *all-roberta-large-v1* SBERT model. The most appropriate optimization objective for our task is Triplet Loss (Hermans et al., 2017), and we train the model on our training set for 5 more epochs.

#### 3.2.1 Tale Splitting

In order to work within the limitations of the maximum input sequence length of SBERT's RoBERTa model, which is 256 tokens, we divide each folktale into sequences of roughly 256 tokens long, including the tale's title. We prioritize preserving quotes without breaking them in the middle of a sentence, so the chunk size may vary but will not exceed 256 tokens. During the training phase, each tale chunk will be associated with the ATU index of the original tale. For testing, we encode each tale chunk individually, and then combine the embeddings by taking the mean to obtain a single representation of the tale.

### 3.3 Baselines

To evaluate the effectiveness of our model in encoding folktales into meaningful vectors, we de-

sign several baselines. Our first baseline is the simple yet powerful TFIDF vectorization of the tales. Our second baseline is the approach deployed by Rewis 2020, using Doc2Vec (Le and Mikolov, 2014). Both these approaches were fitted directly on the test set.

As a third baseline, drawing inspiration from the work of Hay et al. 2020, where they use a BERT (Devlin et al., 2018) model trained for classification to learn author writing style representations, we train for 5 epochs a RoBERTa (Liu et al., 2019) (340 million parameters) model for classification using the ATU indices as labels. This serves as a comparison to the approach proposed by Pompeu et al. 2019, as we anticipate to obtain similar results if we train Pompeu's model on our data and derive tale embeddings. To investigate the effect of splitting a tale into smaller parts on the model's ability to capture the tale's plot, we also train a Longformer (Beltagy et al., 2020) (150 million parameters) model. Since the maximum input sequence length of the longformer is 4096, we set an upper limit of 2560 tokens for the tale chunk size, which is 10 times longer than our other model's chunk size limit. In addition, we compare all trainable models with their untrained variants to understand the effects of training on our tale dataset.

#### 3.3.1 Pre-training with MLM Objective

The Folktexts collection of tales also contains several tales without an ATU index. During our experiments, we also explored using these unlabeled tales to perform an unsupervised pre-training step using the Masked Language Modeling (MLM) objective for RoBERTa and Longformer. Unfortunately, we found that this approach did not consistently improve performance, so we decided not to include these results in our paper.

## 4 Results

We evaluate all models on their ability to encode folktales into meaningful representations using our fixed testing set. When visualizing the embeddings of folktales, meaningful representations will place tales with similar plots close together and tales with dissimilar plots further apart. It's natural to assume that any unseen tale belonging to an unknown type will have a story plot that is similar to the plots of a subset of known types, and that another tale of the same unknown type is likely to have a story plot that is similar to the plots of this same subset of known types. We also expect that our model will be able to capture subtle nuances in the plot that were previously unnoticed, potentially placing some tales that originally belong to different classes close together.

### 4.1 Automatic Evaluation

To evaluate the effectiveness of our model, we encode the test tales using our trained models and then reduce the resulting embeddings to 2 dimensions using t-SNE (Van der Maaten and Hinton, 2008). We then apply K-Means Clustering (Arthur and Vassilvitskii, 2006) and measure the quality of the clustering using the Fowlkes-Mallows index (FMI) (Fowlkes and Mallows, 1983), computed as the geometric mean of the pairwise precision and recall using the ATU type classification of the tales. This metric is ideal to measure our method's clustering capabilities, as a value close to zero indicates poor agreement between our tale representations and their original ATU type classification and a value close to one indicates good agreement. In Table 2, we present the FMI score for K = 20, which is calculated by evaluating on the 20 most frequent ATU tale type indices found in the entire dataset.

We conduct this test for all granularity evaluation settings for each available model, in order to see whether a model trained on a finer (or coarser) granularity can predict a coarser (or finer) class split. Our results show that the best-performing model for the fine and simple ATU granularities, which are the hierarchies of the ATU tale type index that better classify the plot of a story, is *sent-transformers-atu-fine*. This model, trained on the leaf-level ATU indices, captures details of the tale's plot accurately. We found that using longer sequences of text did not improve the ability of a model to encode tales in a more meaningful way. As seen in the results table, RoBERTa performs relatively equally with Longformer, and frequently outperforms it. We were surprised to see that TFIDF performed well in the fine evaluation setting, but it did not surpass the performance of *sent-transformers-atu-fine*.

We suspect that the success of TFIDF in the fine-grained categorization is due to a combination of its capability to capture semantic differences between words and the manner in which the ATU system partitions tale types. The fine-grained classification leverages the use of specialized terminology, as characters and actions are described in more detail. For instance, "The Hare and the Tortoise" (ATU

| Model Family | Model Granularity | Evaluation Granularity | | | |
|---|---|---|---|---|---|
| | | Fine | Simple | Mid | Coarse |
| **TFIDF** | N/A | 0.43 | 0.36 | 0.19 | 0.37 |
| **Doc2Vec** | N/A | 0.37 | 0.46 | 0.19 | 0.35 |
| **RoBERTa** | N/A | 0.06 | 0.07 | 0.08 | 0.23 |
| **RoBERTa-atu** | Fine | 0.15 | 0.15 | 0.11 | 0.24 |
| | Simple | 0.07 | 0.11 | 0.15 | 0.25 |
| | Mid | 0.30 | 0.39 | **0.43** | 0.53 |
| | Coarse | 0.26 | 0.29 | 0.37 | **0.64** |
| **Longformer** | N/A | 0.17 | 0.14 | 0.13 | 0.28 |
| **Longformer-atu** | Fine | 0.11 | 0.12 | 0.12 | 0.31 |
| | Simple | 0.13 | 0.11 | 0.12 | 0.29 |
| | Mid | 0.11 | 0.09 | 0.13 | 0.29 |
| | Coarse | 0.28 | 0.35 | 0.34 | **0.64** |
| **Sent-Transformer** | N/A | 0.28 | 0.35 | 0.19 | 0.33 |
| **Sent-Transformer-atu** | Fine | **0.44** | **0.55** | 0.22 | 0.36 |
| | Simple | 0.32 | 0.43 | 0.20 | 0.32 |
| | Mid | 0.28 | 0.50 | 0.34 | 0.48 |
| | Coarse | 0.32 | 0.27 | 0.20 | 0.35 |

Table 2: FMI score for K = 20 computed on the 20 most frequent ATU indices found in the entire dataset. Models that are trained on our tale training set have the "-atu" ending in their names, while untrained models (applied 0-shot to our test set) do not have a postfix. Model Granularity refers to the level of ATU hierarchy that was used to train the model, while Evaluation Granularity refers to the granularity used to compute the FMI score. The granularities are: first-level (coarse), second-level (mid), level immediately preceding the leaf-level (simple) and leaf-level (fine).

275A) belongs to a different fine granularity level class than "The Fox and the Snail" (ATU 275B) even though their simple granularity level class would be the same (ATU 275). TFIDF is capable of differentiating between "hare" and "fox" and "tortoise" and "snail", and therefore performs better when these classes are separate. This is not the case with our sent-transformers-atu-fine model which, despite being trained on the fine granularity level, performs optimally at the simple level. This is because the model encodes entities such as "hare" and "fox" as similar.

Figure 3 shows a snapshot of our visualization tool, displaying the 2D points representing the tales of the test set encoded with our best-performing model. Besides the FMI scores, the effectiveness of the method is demonstrated through visual inspection, where tales within a given ATU index tend to cluster closely around the corresponding landmark tale. This can be observed in the 510 ATU index tales, which are closely grouped near the Cinderella landmark tale, tales of the 275 ATU index cluster around The Hare and the Tortoise landmark. Similar observations can be made for tales of the ATU indices 980, 1540 and 1645, which also cluster closely around their respective landmark tales.

### 4.2 Human Verification

We use the best-performing model in automatic evaluation (*sent-transformers-atu-fine*) to manually analyze the tales of the test set that belong to different classes but are represented closely together in the embeddings. This allows us to investigate which elements in the texts led to this phenomenon and understand the model's behavior.

To facilitate this analysis, we have created an interactive visualization app, which currently includes the most necessary features but has the potential for future expansion into a more comprehensive folktale analysis tool. Figure 3 also shows 4 circled pairs of points, which are the tales selected for the human verification, due to their proximity in the plot:

1. "The Adventures of Juan"[5] and "Andres the Trapper"[6]

2. "The Crocodile the Brahman and the Fox"[7]

[5] The Adventures of Juan link
[6] Andres the Trapper link
[7] The Crocodile the Brahman and the Fox link

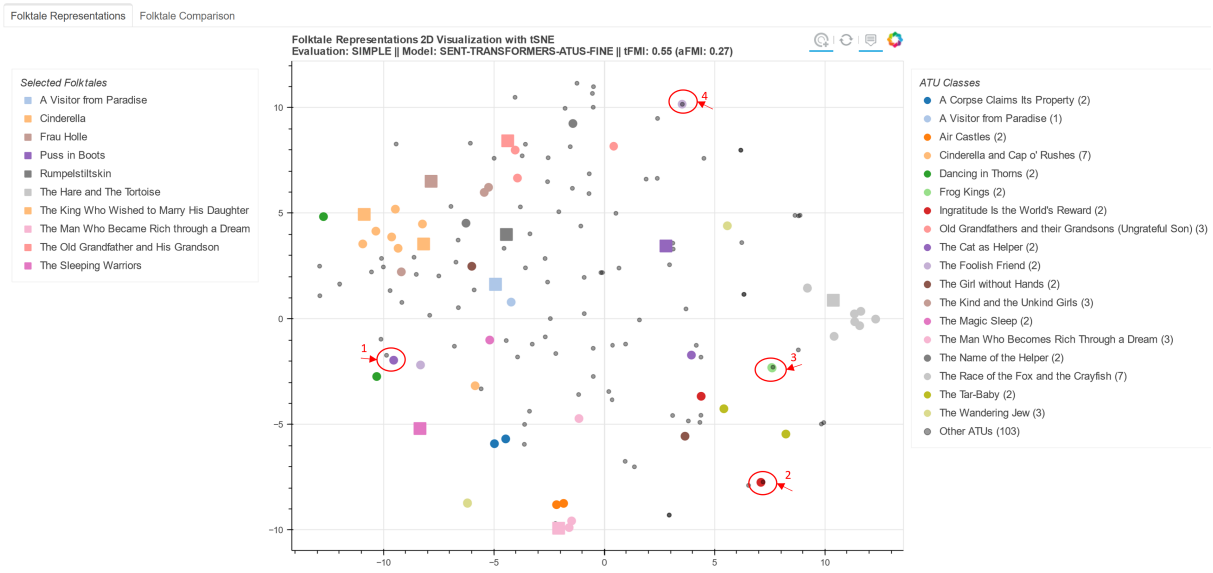


Figure 3: Snapshot of our visualization tool, displaying the 2D points representing tales. The 4 circled pairs of points are the tales selected for human evaluation.

and "The Monkey and the Crocodile"[8]

3. "The Two Frogs who were Neighbors"[9] and "The Princess and the Frog"[10]

4. "The Gardener and the Bear"[11] and "The Kobold and the Polar Bear"[12]

#### 4.2.1 Selected Tales Analysis

Both tales in the first pair on our list originate from the Philippines. They have a common theme of a poor boy as the main character who starts with nothing. Throughout the tales, the characters encounter a helper who aids them in achieving wealth and finding a suitable partner, ultimately leading to a joyful ending.

The tales in the second pair feature a crocodile as the antagonist. The tales depict animals as cunning and clever, and in the end, the protagonist outwits the crocodile and emerges victorious. These tales have the same cultural background and they originate from India.

The third pair of tales share a common theme of a main character who is initially hesitant to take the right course of action. Additionally, the tales share similar vocabulary, both featuring frogs in a pond.

The last pair of tales feature characters who exhibit foolish behavior, with their lack of wisdom ultimately revealed at the conclusion of the story. Additionally, both tales contain elements of physical violence, with the bear playing a role in the action. Furthermore, the human characters in the stories are in a state of slumber when the pivotal event takes place.

These examples demonstrate that our model is capable of identifying and grouping together folktales that share similar themes, plot structures, and character development arcs, regardless of their origin or cultural background. Through this limited human verification, we conclude that the pairs of tales represented near each other were in fact closely related in terms of plot. As non-experts in the field of folklore, we believe that this approach to folktale classification has the potential to greatly enhance the efficiency and effectiveness of future tale classification efforts, by reducing the time and manual labor required for this task.

## 5 Conclusion

Recent advancements in the field of Natural Language Processing have made it possible to develop neural models that can automatically cluster folktales based on their plot. Categorizing folklore based on the plot and type of tale is an area that has received extensive scholarly inquiry and examination in the field of Folklore Studies. By utilizing the popular tale and myths collection of D. L. Ashliman, we trained a sentence transformer model to encode tales into numerical vectors whose prox-

[8] The Monkey and the Crocodile link
[9] The Two Frogs who were Neighbors link
[10] The Princess and the Frog link
[11] The Gardener and the Bear link
[12] The Kobold and the Polar Bear link

imity in a lower-dimensional space represents similarity in tale type. To verify this, we conducted an automated evaluation using K-means clustering and a human analysis of selected pairs of tales, the latter indicating that the model can reveal subtle nuances in the plot that were previously undetected, making it an invaluable tool for comparative folktale analysis.

While Ashliman's Folktexts represents a thorough curation of folktales, it is not a comprehensive representation of the semantic landscape of folklore. In our future work, we intend to augment this corpus by incorporating other folktale collections such as the MFTD, which includes tales in various languages. This expansion will enhance the robustness of our models with respect to the narrative structure of tales, as different languages may convey similar concepts through different phrasings and lexical choices.

Additionally, we aim to harness the capabilities of our models to rethink the ATU classification system. Our computational approach has demonstrated its ability to identify commonalities in tales, such as their origin, thematic elements and situational motifs, while simultaneously accounting for their differences. By training this approach to all available data, we can use these powerful tale representations to reshape the task of tale classification, by merging previously distinct tale types, dividing a type into multiple new types, or even create new types from previously unclassified tales when our model suggests that seemingly dissimilar tales have more in common than what is apparent to the naked eye.

## References


David Arthur and Sergei Vassilvitskii. 2006. k-means++: The advantages of careful seeding. Technical report, Stanford.

Dee L Ashliman. 2004. *Folk and Fairy Tales: A Handbook: A Handbook*. ABC-CLIO.

Iz Beltagy, Matthew E Peters, and Arman Cohan. 2020. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150*.

Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. 1992. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Alan Dundes. 1997. The motif-index and the tale type index: A critique. *Journal of Folklore Research*, pages 195–202.

Edward B Fowlkes and Colin L Mallows. 1983. A method for comparing two hierarchical clusterings. *Journal of the American statistical association*, 78(383):553–569.

Joseph P Goodwin. 1995. If ignorance is bliss,'tis folly to be wise: What we don't know can hurt us. *Journal of Folklore Research*, pages 155–164.

Joshua Hagedorn and Sándor Darányi. 2022. Bearing a bag-of-tales: An open corpus of annotated folktales for reproducible research. *Journal of Open Humanities Data*, 8(16).

Julien Hay, Bich-Lien Doan, Fabrice Popineau, and Ouassim Ait Elhara. 2020. Representation learning of writing style. In *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, pages 232–243.

Alexander Hermans, Lucas Beyer, and Bastian Leibe. 2017. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*.

Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196. PMLR.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Theo Meder, Folgert Karsdorp, Dong Nguyen, Mariët Theune, Dolf Trieschnigg, and Iwe Everhardus Christiaan Muiser. 2016. Automatic enrichment and classification of folktales in the dutch folktale database. *Journal of American Folklore*, 129(511):78–96.

Dong Nguyen, Dolf Trieschnigg, Theo Meder, and Mariët Theune. 2012. Automatic classification of folk narrative genres. In *Proceedings of the Workshop on Language Technology for Historical Text (s) at KONVENS 2012*, pages 378–382.

Duarte Pompeu, Bruno Martins, and David Martins de Matos. 2019. Interpretable deep learning methods for classifying folktales according to the aarne-thompson-uther scheme.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Benjamin Rewis. 2020. espace of folklore: Mapping folkloric texts semantically with document embeddings.

Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of machine learning research*, 9(11).

Zhiguo Wang, Wael Hamza, and Linfeng Song. 2017. $k$-nearest neighbor augmented neural networks for text classification. *arXiv preprint arXiv:1708.07863*.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 1480–1489.