

---

## Finding Trolls Under Bridges: Preliminary Work on a Motif Detector

---

**W. Victor H. Yarlott**<sup>1</sup>

**Armando Ochoa**<sup>1</sup>

**Anurag Acharya**<sup>1</sup>

**Laurel Bobrow**<sup>2</sup>

**Diego Castro Estrada**<sup>1</sup>

**Diana Gomez**<sup>1</sup>

**Joan Zheng**<sup>2</sup>

**David McDonald**<sup>2</sup>

**Chris Miller**<sup>2</sup>

**Mark A. Finlayson**<sup>1</sup>

WVYAR@CS.FIU.EDU

AOCHO032@FIU.EDU

AACHA007@FIU.EDU

LBOBROW@SIFT.NET

DCAST230@FIU.EDU

DGOME133@FIU.EDU

JZHENG@SIFT.NET

DMCDONALD@SIFT.NET

CMILLER@SIFT.NET

MARKAF@FIU.EDU

<sup>1</sup>Knight Foundation School of Computing and Information Sciences, Florida International University, 11200 SW 8th St., Miami, FL 33199 USA

<sup>2</sup>Smart Information Flow Technologies (SIFT), 114 Waltham Street, Ste. 24, Lexington, MA 02421 USA

### Abstract

Motifs are distinctive recurring elements found in folklore that have significance as communicative devices in news, literature, press releases, and propaganda. Motifs concisely imply a large constellation of culturally-relevant information, and their broad usage suggests their cognitive importance as touchstones of cultural knowledge, making their detection a worthy step toward culturally-aware natural language processing tasks. Until now, folklorists and others interested in motifs have only extracted motifs from narratives manually. We present a preliminary report on the development of a system for automatically detecting motifs. We briefly describe an annotation effort to produce data for training motif detection, which is on-going. We describe our in-progress architecture in detail, which aims to capture, in part, how people determine whether or not a motif candidate is being used in a motific way. This description includes a test of an off-the-shelf metaphor detector as a feature for motif detection, which achieves a  $F_1$  of 0.35 on motifs and a macro-average  $F_1$  of 0.21 across four categories which we assign to motif candidates.

### 1. Motifs as a Source of Cultural Information

Motifs are distinct, recurring narrative elements found in folklore and, more generally, culturally inflected materials. Motifs are interesting because they provide a compact source of cultural information: many motifs concisely communicate a related constellation of cultural ideas, associations, and assumptions. For example, “troll under a bridge” is an example of a motif common in the Western cultures that have roots in Scandinavia. To those familiar with the motif, it entails a number of

related ideas that are by no means directly communicated by the surface meaning of the words: the bridge is along the critical path of the hero, and he must cross to achieve his goal; the troll often lives under the bridge, crawling out to waylay innocent passers-by; the troll charges a toll or exacts some other payment for crossing the bridge; the troll is a squatter, not the ‘officially’ sanctioned master of the bridge; the troll enforces his illegitimate claim through threat of physical violence; and the hero often ends up battling (and defeating) the troll instead of paying the toll.

Because of this density of information, motifs are often retained as a tale is passed between cultures and down generations, and folklorists have observed that a tale’s specific composition of motifs can be used to trace the tale’s lineage (Thompson, 1977, Part 4, Chapter V). This has led folklorists to construct motif indices that identify motifs and note their presence in specific tales (usually as represented in a particular folkloristic collection). The most well-known motif index is the Thompson motif index (TMI) by Thompson (1960), which references tales from over 614 collections, indexed to 46,248 motifs and sub-motifs, 41,796 of which have references to tales or tale types. In Thompson’s index each motif is given a designating code; for example, “troll under a bridge” is referenced by the codes G304 and G475.2. In this case, “troll under a bridge” is represented by two motifs as Thompson generalizes trolls to ogres, a general class of monstrous beings; thus, the motifs are “troll as ogre” (G304) and “ogre attacks intruders on bridge” (G475.2).

Thompson informally defines a motif as items “worthy of note because of something out of the ordinary, something of sufficiently striking character to become a part of tradition, oral or literary. Commonplace experiences, such as eating and sleeping, are not traditional in this sense. But they may become so by having attached to them something remarkable or worthy of remembering” (Thompson, 1960, p. 19). He notes that motifs generally fall into one of three subcategories: an event, a character, or a prop (Thompson, 1977, pp. 415–416). Here we give an example of each (with their associated Thompson’s motif code):

A **hero rescuing a princess** (B11.11.4) is perhaps one of the most well-known event motifs in western culture. Ask a westerner the following question: “A princess has been kidnapped: who kidnapped her, who rescues her, and what does the rescuer need to do to effect a rescue?”, and common answers will be “a dragon kidnapped her, the knight must rescue her, and he must kill the dragon.” This motif may be the climax of the story, with a “happily ever after” ending just after the hero defeats the dragon, or it may just happen in the course of a story: in *Ivan Dogson and the White Polyanin*, a Russian folktale (Afanas’ev, 1957, Tale #139), Ivan slays three dragons, each with more heads than the last, rescuing a princess each time. The motif is prolific, found across the tales, literature, and movies of many cultures.

**Old Man Coyote** (A177.1) is a character motif: known in some Native American Indian tribes as Coyote, he is one of the most recognizable gods. In Native American Crow folklore, Old Man Coyote creates the earth and all the creatures on earth. He travels the world, teaching the animals how they should behave. Old Man Coyote, however, is far from a noble and elegant creator. He creates ridiculous costumes and tries to trick the Crow tribe into wearing them, only to be run off. He purposefully bungles rituals to produce food, such as transforming skin from his back to meat, in order to guilt his guests into performing the ritual correctly to get free food, later performing it correctly to discredit his former guests when they tell others he erred. Anywhere Old Man Coyote is

referenced, he calls to mind someone who has done great things, but is lazy and often far too clever for their own good, falling pray to their own cunning.

A **magic carpet** (D1155) is a prop that allows the hero to fly through the sky, and is familiar to anyone who has watched Disney’s *Aladdin*. In *One Thousand and One Nights*, Prince Husain encounters a merchant selling a carpet for an outrageous price; the merchant says: “O my lord, thinkest thou I price this carpet at too high a value? . . . Whoever sitteth on this carpet and willeth in thought to be taken up and set down upon other site will, in the twinkling of an eye, be borne thither, be that place nearhand or distant many a day’s journey and difficult to reach” (Burton, 2009, p. 496). Solomon, said to be the third king of Israel, was said to have a carpet 60 miles on each side that could transport him vast distances in a short amount of time. In Russian hero tales, magic carpets are common items that aid the hero in his quest.

While the above examples are drawn from folklore, motifs have importance beyond folktales: they occur in everyday speech, news stories, press releases, propaganda, novels, movies, plays, and anywhere that cultural materials are found. A powerful modern example is the use of the *Pharaoh* motif in modern middle eastern discourse. The Pharaoh appears in Qur’an, and comes into conflict with Moses and his attempts to free the Hebrews from Egyptian slavery. The Pharaoh is an arrogant and obstinate tyrant who defies the will of God and is punished for it. In modern Islamist extremist narratives, the Pharaoh is a symbol of struggles against anti-Islamic regimes and has been invoked against leaders such as Anwar Sadat of Egypt, Ariel Sharon of Israel, and George W. Bush, whom Osama bin Laden referred to as the “pharaoh of the century” as noted by Halverson et al. (2011).

While motifs appear in many places, it is not always clear when they are being used to communicate the greater constellation of their folkloric meaning: the raw text “pharaoh” itself could be an allusion to historical tyranny but it could easily appear in a document describing the leaders of Ancient Egypt. It is therefore also necessary to be able to clearly distinguish between different usages of specific instances of motif text—what we refer to as “motif candidates.” We do so by assigning them to one of four categories: motific, referential, eponymic, or unrelated. We describe these four categories in more detail in section §2.4.

We find that, in practice, motifs somewhat resemble figurative usage of language; however, it’s not clear-cut. Calling an athlete Finn McCool may very well be their name in Ireland—or it might be alluding to the mythical hero Finn McCool to praise the athlete’s physical prowess. Therefore, we believe that some degree of overlap between metaphor, simile, and motif exists, but that metaphor and simile detectors will not be off-the-shelf solutions for motif detection.

Because of their prominence and ubiquity, the ability to automatically detect motifs would open up a vast repository of important cultural information—something that as of yet, is unaddressed (Acharya et al., 2021)—to computational analysis, allowing systems to be built that leverage this powerful facet of human communication. Currently, motifs must be extracted by hand by trained cultural experts.

In this work, we outline preliminary efforts to build a system for automated motif detection. We briefly describe our ongoing annotation study (§2), describe our on-going work in developing a pipeline based on how humans understand stories for the automatic detection of motifs (§3), and present preliminary results exploring metaphoricity as a feature for motif detection (§4). We also

review related work to provide context (§5). We conclude by describing potential future work and summarizing our contributions (§6).

## 2. Data Production

As a first step toward developing an automated system, we have undertaken an annotation study to produce data categorizing motifs as either *motific*, *referential*, *eponymic*, or *unrelated*. This ongoing annotation will result in identifying and labeling approximately 21,000 motif candidates, for which analysis is still underway. We briefly describe the annotation study. This data will be used to validate our automatic system for detecting motifs.

### 2.1 Selection of Motifs

To begin selecting motifs, we needed to identify the cultural groups from which to draw motifs and their source folklore. This was done by checking two criteria: first, that there was a strong authoritative source of motifs (a motif index or folklore collection)—this made it easier for us to identify motif candidates. Second, a sizable population that was easily accessible—this was partially in service of an ongoing survey study as part of the same effort this work falls under, but also to ensure that we would easily be able to find annotators (although world events quickly made these concerns obsolete). Using these criteria, we selected Irish, Jewish, and Puerto Rican as the cultural groups to draw motifs from, as well as to recruit annotation candidates from.

We had several selection criteria: (1) motifs must be highly-referenced in social media or news, demonstrating they were in common use; (2) motifs must have a definitive source within the folklore of the cultural group, giving proof of relevance; and, (3) the motifs must have a high potential strength, which we subjectively estimated based on input from members of the cultural group and the frequency of usage of the motifs in social media and news.

#### 2.1.1 Examples

We identified 22 character motifs, 11 prop motifs, and 6 event motifs for use in the study, of which 13 were Irish, 10 were Jewish, and 16 were Puerto Rican. Here, we provide sample motific usages:

Finn McCool, Irish:

In fact, that was where he acquired the nickname, playing for the parish in a Co Kilkenny semi-final, a young **Finn McCool** with limbs that might bestride a bullock.

Tower of Babel, Jewish:

The first Slavic pope, John Paul II, swept joyfully through the streets of the Czechoslovak capital yesterday, proclaiming a unified Europe was at hand and communism had crumbled like the **Tower of Babel**.

Coqui, Puerto Rican:

We listen here for the first strains of the exotic, in the parrot’s squawk, the **coqui’s** clamor, the reggaeton beat - but also for the beleaguered howl of the native: the panther’s scream, the ivorybill’s rap, the Miccosukee’s lilting tongue.

## 2.2 Selection of Data

Selection of data was done using keyword searches of phrases related to motifs of interest, identified as described in §2.1, through NexisUni, a university-focused version of LexisNexis, a product that makes news articles easily searchable. The resulting articles of this basic keyword search were downloaded and converted from rich-text format to plain-text format. This data consists of news articles that LexisNexis identifies to have text matches for phrases relevant to the motif (e.g., “Finn McCool”).

These plain-text news articles were then further processed by our own Lucene-based lexical matcher, which uses a hand-crafted ruleset to identify the surface form of known motifs; these surface forms are not necessarily true motific usage (e.g., it could be an article on the origin of “Amalek”)—thus, we refer to them as motif *candidates*. We use these candidates to verify the presence of motifs and tag the spans of text containing them in a stand-off annotation. The articles are then chosen from randomly to avoid any temporal or geographic dependencies that might be order dependent, and the articles and annotations are then used for both the annotation (using the brat rapid annotation tool (Stenetorp et al., 2012)) and as input to the pipeline.

## 2.3 Annotator Training

Annotation was done by three pairs of annotators hired for their membership in the cultural groups (Irish, Jewish, and Puerto Rican) that were chosen for the accessibility of their motifs and their proximity to our researchers, as described in §2.1. Each pair consisted of annotators from a single one of these three cultural groups, tasked with annotating Irish, Jewish, or Puerto Rican motifs as appropriate. All annotators had or were in the process of obtaining college degrees.

Annotators were given a single two-hour session of training that included reviewing the annotation guide, covering any questions or concerns, and running through a small sample annotation together as a team. Further, a two-hour adjudication session was held every week to cover the week’s annotations: while these served to help reinforce the annotator’s skills, they also served to produce adjudicated data for machine learning. Annotators were not allowed to communicate with each other regarding the annotations until these adjudication meetings.

Annotators are also given access to an annotation guide that provides a high-level description of annotations, descriptions of what to annotate, tutorials of how to use the tool, and detailed descriptions of the motifs drawn from discussion with in-culture members. Annotators were allowed to consult the annotation guide at any point during annotation.

## 2.4 Annotation Procedure

Annotators used the brat rapid annotation tool with pre-marked text spans (Stenetorp et al., 2012). This was displayed as a text file with several snippets of articles containing highlighted spans representing the motif candidate that the annotators would have to change to the correct annotation.

Each span represented a motif candidate with its extent drawn from our lexical matcher and its label simply being the motif: e.g., *finn-mccool*, *amalek*, etc. Annotators were instructed to tag each of these as one of four categories:

**Motific** A usage that intends to invoke the cultural associations of a motif (e.g. referring to something large and monstrous as a “behemoth”).

**Referential** A usage that directly refers to the folklore origin of the motif or its definition (e.g., discussing the origin of the term “behemoth”).

**Eponymic** A usage that references the motif in a name—this distinction is made because while it is highly similar to motific usage, it is typically not used as such (e.g., the band “behemoth” may be referred to with no additional meaning beyond the band).

**Unrelated** A usage that is unrelated to the cultural group or cannot be established as directly related (e.g., “behemoth” as a monster in a game).

Once annotators completed a batch of annotation (a process that took roughly a week and produced anywhere from 500-1000 candidates, depending on the group and the selection of articles), an adjudication meeting was held. Inconsistent tags were reconciled via discussion and agreement on the correct tagging; this serves not only to increase the amount of gold-standard data, but also to reinforce annotator training and address any issues that arose during the annotation process.

## 2.5 Preliminary Agreement

For this annotation study, we expect annotator agreement to remain above 0.55 Fleiss’ kappa: we show preliminary results for the first two batches after training until reaching the desired Fleiss’ kappa in Table 1. These results are preliminary, but suggest that humans with a strong background in the relevant culture can reliably distinguish motific, eponymic, referential, and unrelated candidates of a motif. Table 1 also shows preliminary tag counts and number of candidates annotated with a high level of agreement—this distribution is in line with our expectations that motifs are relatively rare in clear motific usage as opposed to eponymic or referential usage, based on early exploratory work and the understanding that using motifs in a motific way requires a deeper level of understanding of the motif.

## 3. Architecture

We show our end-to-end pipeline in Figure 1. The majority of our automatic system is contained within the black square and will perform the same task as the human annotators with the same input: the raw article text with the spans containing candidate matches using a manually-created motif rule list.

The automatic system is designed as a traditional NLP pipeline (as opposed to a neural system) that caches the results of many different NLP tools, the output of which are parsed into features inspired by our understanding of humans read and understand stories, as well as intuitive constructs based on these ideas that target the three specific categories of motifs described by Thompson:

Table 1. Preliminary agreement and counts from our in-progress annotation study.

Group	Agreement ( $F_{\kappa}$ )	# Motific	# Referential	# Eponymic	# Unrelated	Batch Size
Irish	0.559	11	292	353	207	863
Irish	0.699	46	355	323	254	978
Jewish	0.579	56	232	246	18	552
Jewish	0.638	133	374	348	56	911
Puerto Rican	0.552	66	427	237	134	864
Puerto Rican	0.680	19	433	212	174	838
Total		331	2,113	1,719	843	5,006

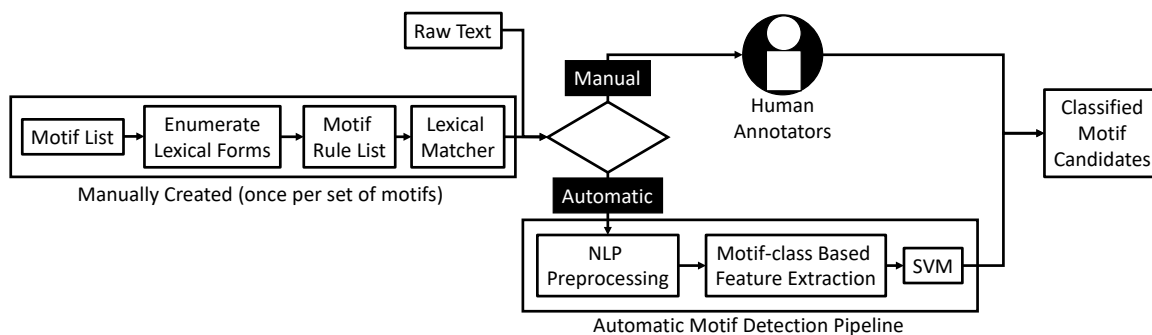


Figure 1. End-to-end Motif Detection Pipeline. Displayed are necessary manual preprocessing steps, and our automatic system, which we are validating with annotated data, performs the task assigned to our human annotators: identifying which of the candidates from our high-recall lexical matcher are motific, eponymic, referential, or unrelated.

characters (“the actors in a tale”), props (“certain items in the background of the action”), and events (“single incidents”) (Thompson, 1977, pp. 415–416). In §3.3, we discuss what features we are implementing to target these categories of motifs.

### 3.1 Manual Motif List Creation

Manual input is, at this point in time, a necessary evil—this step is required to identify a list of motifs of interest (as described in Section §2.1) and then to bootstrap initial lexical forms and write lexical matcher rules for these forms. These rules are used with a Lucene-based lexical matcher to identify the spans of text that are selected as motif candidates.

This procedure is done only once for each motif that is part of the list of motifs: once the lexical matcher’s rules are set and run over the raw text, the stand-off brat annotation files are created, and both are ready to be passed on to either human annotators, if the work is being conducted manually as in our annotation study, or to our automatic motif detection pipeline. We imagine that in the future, this step could be automated through the development of an automatic parsing engine for motif indices, as described Yarlott & Finlayson (2016, p. 7).

### 3.2 NLP Preprocessing

For preprocessing, we use many off-the-shelf NLP modules, such as Stanford’s CoreNLP (Manning et al., 2014) and OpenIE (Angeli et al., 2015), and spaCy (Honnibal et al., 2020). Additionally, we use Sparser (McDonald, 1992; Friedman et al., 2017) to extract high-precision information related to the motif candidates.

We also use modules based on figurative language: our own reimplementations of the DeepMET metaphor detector (Su et al., 2020) and a neural simile detection system (Liu et al., 2018). These modules are used, in part, to help catch the overlap between figurative language and motific usage, as well as to discriminate against categories that have a high degree of non-figurative usage, such as referential examples.

We additionally run the animacy and character systems developed by Jahan et al. (2018, 2020), our own reimplementation of a possession detector (Chinnappa & Blanco, 2018), the CAEVO event detector (Chambers et al., 2014), and the SRL included in Story Workbench (Finlayson, 2008, 2011) and the AllenNLP SRL (Gardner et al., 2017).

The end result of this preprocessing is a large cache of stand-off files containing per-document dependency parses, POS tags, NER tags, OpenIE triplets, coreference chains with accompanying animacy and character labels, events, and SRL relations; additionally, the preprocessing produces per-candidate metaphor, simile, and possession tags. All of these stand-off files are either processed into feature vectors (in the per-candidate cases) or further computed into useful feature vectors (in the per-document cases).

### 3.3 Feature Extraction

We draw substantial inspiration for our features from the theoretical description of motifs as characters, props, and events by Stith Thompson, with groups of features targetting specific classes of motifs.

**Character motifs** are very clearly related to animacy and characters in stories and so, we calculate distance, coreference, and semantic relation or grammatical participation between the candidates and animate or character coreference chains. Additionally, characters very often have a proper name, for which we target with NER features.

**Prop motifs** are also checked for semantic, grammatical, or possession relation with animate or character entities, as in motifs they are often used in relation to characters. We further check general semantic role and part-of-speech to target prop motifs.

To target **event motifs**, we use CAEVO to determine the participation in, co-occurrence with, and distance from events: these different features are because an event motif might be used *as* the event itself, but it may also be nearby as part of a comparison or as further description. We also expect the entities taking part in an **event motifs** to be animate.

The end result of this feature extraction are per-candidate vectors of features, among which are: whether or not they are animate, participation in animate chains, participation in events, semantic relation to animate entities, grammatical relation to animate entities, nearby NER tags, and whether or not they are metaphoric or involved in simile.



Table 2. Results of using the DeepMET metaphor detector by Su et al. (2020) to detect motifs. These results were obtained using part of the pipeline architecture under development as described in Section §3.

Model	Motif $F_1$	Eponym $F_1$	Referential $F_1$	Unrelated $F_1$	Macro $F_1$
Metaphor Alone, SVM	0.35	0.00	0.59	0.00	0.21

### 3.4 Motif Detection

The features being used for motif detection generally fall into one of three categories: (1) their normalized relative position to a motif candidate (windows, co-occurrence, token distance, and parse distance); (2) their semantic relation with the motif candidate (SRL, IE, possession, discourse relation, and word sense); or (3) the usage of figurative language in relation to the motif candidate (simile and metaphor). Additionally, we are developing a new metaphor detector in hopes of better capturing where metaphoric and motific usage overlaps. Our detector also attempts to predict the motif type from the features and compare that to the actual, expected motif type based on the lexical form (e.g., “Finn McCool” is a character, not a prop). These features will then be fed into a means of classifying them as motific, referential, eponymic, or unrelated.

## 4. Preliminary Results: Metaphor

As a preliminary test, we reimplemented the DeepMET metaphor detector described in Su et al. (2020) to determine if a motif candidate is in the same sentence as a metaphor. We use this test as a sanity check—intuitively, we expect motifs to primarily be figurative in use, as referring to someone as “Amalek”—an antagonist in Jewish stories—or comparing something to a “coqui”—a frog native to Puerto Rico that features prominently as its mascot—are not literally stating that they are a heinous ruler from millennia past or a frog. Interestingly, however, we find that this is not exactly the case: in Table 2, we present the results of testing the metaphor-based feature.

The results demonstrate that while there is some overlap between metaphor and motific usage, it does not achieve substantial performance ( $F_1 = 0.35$ ), with the feature performing better at detecting referential usage of motif candidates ( $F_1 = 0.59$ ).

This result is not wholly unexpected. As we discussed briefly in §1, the relation between metaphor and motif is not 1-to-1, thus resulting in relatively low performance. We theorize that the relatively high performance at detecting referential usage may be due to referential usage, which we describe as either strict discussion of the folklore origin or a definition of the motif, are much more likely to use literal language. Therefore, the a *False* metaphor tag may be an indicator of referential usage.

These results, which show that metaphor detection does not work as a drop-in motif detector, reinforce our belief that deeper systems based on story understanding and theory (as the origin of motifs *is* stories) are necessary to detect motifs.

## 5. Related Work

On the folklore side, there are many motif indices, with the Aarne-Thompson Motif-index of Folk-literature (Thompson, 1960) being the primary resource. There are numerous other indices, most primarily focused on a specific culture. Thompson also has substantial discussion on motifs and the compilation of indices in his book *The Folktale* (Thompson, 1977). While Thompson’s motif index is perhaps the primary source of motif information used today, it has been criticized because of overlapping motif subcategories, censorship (primarily of obscenity), and missing motifs (Dundes, 1997). These motif indices provide a substantial base for us to build upon and we draw heavily from both the Aarne-Thompson index as well as Tom Peete Cross’ Motif-index of of Early Irish Literature (Cross, 1952) and Dov Noy’s Motif-index of Talmudic-Midrashic literature (Noy, 1954) to select our group of motifs.

Darányi (2010) has called attention to the need for research into the automation of extraction and annotation of motifs in folklore. Further work by Darányi & Forró (2012) has determined that motifs may not be the highest level of abstraction in narrative, Darányi et al. (2012) have made substantial headway towards using motifs as sequences of “narrative DNA”, and Ofek et al. (2013) have demonstrated learning tale types based on these sequences. Declerck et al. (2012) have also done work on converting electronic representations of TMI and ATU to a format that enables multilingual, content-level indexing of folktale texts, building upon past work (Declerck & Lendvai, 2011). Currently, this work appears to be focused on the descriptions of motifs and tale types, without reference to the stories. In this work, we take the first steps towards automatically detecting motifs in folklore as well as verifying this system with gold-standard data created by in-culture annotators.

With regard to analyzing motif annotation schemes, Karsdorp et al. (2012) present an analysis of the degree of abstraction present in the ATU catalog and the methods used to note what motifs belong to a given tale type. They find the ATU annotation insufficient for analyzing recurring motifs across types, in that it the ATU scheme fails to capture commonalities across closely related types while also failing to provide sufficient detail. We aim to rectify this in our own annotation by providing detailed descriptions of motifs and providing a much clearer, more formal definition drawn from the work of Yarlott & Finlayson (2016).

Additionally, Yarlott & Finlayson (2016) have done work towards a more formal definition of motif as well as providing a general framework for a potential motif extraction and detection system. Our system continues this work by begining the development of the proposed framework.

## 6. Future Work & Contributions

Currently, we aim to find features to continue developing the motif detector with the aim of substantially outperforming the preliminary metaphor results. We are also implementing a neural baseline to compare to the result of the pipeline in both performance and generalizability. We plan to complete the annotation study and finalize the analysis to further demonstrate the importance of motifs.

The contributions of this short paper are as follows: we have shown preliminary work on the annotation of motifs alongside preliminary agreement measures that suggest humans with a strong background in the relevant cultural group can reliably distinguish motific usage of motifs. We have

described, in detail, our pipeline architecture, currently under development. We have also presented preliminary results on the use of metaphoricity as a feature for distinguishing motific usages of motif surface forms, which serve to demonstrate the gap between metaphoric and motific language with respect to automatic detectors, as well as the need for a detector specifically designed for motifs.

## Acknowledgements

This work was supported by DARPA via SBIR Phase II Prime contract FA8650-19-C-6017. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the DARPA.

## References

- Acharya, A., Talamadupula, K., & Finlayson, M. A. (2021). Toward an atlas of cultural commonsense for machine reasoning. *Proceedings of the Workshop on Common Sense Knowledge Graphs (CSKGs)*. Online. From [https://usc-isi-i2.github.io/AAAI21workshop/papers/Acharya\\_CSKGsAAAI-21.pdf](https://usc-isi-i2.github.io/AAAI21workshop/papers/Acharya_CSKGsAAAI-21.pdf). Held in conjunction with the 35th AAAI Conference on Artificial Intelligence (AAAI 2021).
- Afanas'ev, A. N. (1957). *Narodnye russkie skazki*. Moscow: Gos. Izd-vo Khudozh Lit-ry.
- Angeli, G., Premkumar, M. J., & Manning, C. D. (2015). Leveraging linguistic structure for open domain information extraction. *Proceedings of the Association of Computational Linguistics (ACL)*. From <https://nlp.stanford.edu/pubs/2015angeli-openie.pdf>.
- Burton, R. F. (2009). *The arabian nights*. Barnes & Noble.
- Chambers, N., Cassidy, T., McDowell, B., & Bethard, S. (2014). Dense event ordering with a multi-pass architecture. *Transactions of the Association for Computational Linguistics*, 2, 273–284.
- Chinnappa, D., & Blanco, E. (2018). Mining possessions: Existence, type and temporal anchors. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 496–505).
- Cross, T. P. (1952). *Motif-index of early irish literature*. Indiana University.
- Darányi, S. (2010). Examples of Formulaity in Narratives and Scientific Communication. *Proceedings of the First International AMICUS Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts* (pp. 29–35). From [http://ilk.uvt.nl/amicus/amicus{\\\_}ws2010{\\\_}proceedings.html](http://ilk.uvt.nl/amicus/amicus{\_}ws2010{\_}proceedings.html).
- Darányi, S., & Forró, L. (2012). Detecting Multiple Motif Co-occurrences in the Aarne-Thompson-Uther Tale Type Catalog: A Preliminary Survey. *Anales de Documentación*, 15. From <http://revistas.um.es/analesdoc/article/view/analesdoc.15.1.134691/131801>.
- Darányi, S., Wittek, P., & Forró, L. (2012). Toward Sequencing “Narrative DNA”: Tale Types, Motif Strings and Memetic Pathways. *Third Workshop on Computational Models of Narrative (CMN)* (pp. 2–10). Istanbul, Turkey: European Language Resources Association (ELRA).

- Declerck, T., & Lendvai, P. (2011). Linguistic and semantic representation of the thompson's motif-index of folk-literature. In *Research and advanced technology for digital libraries*, 151–158. Springer.
- Declerck, T., Lendvai, P., & Darányi, S. (2012). Multilingual and Semantic Extension of Folk Tale Categories. *Proceedings of the 2012 Digital Humanities Conference (DH 2012)*.
- Dundes, A. (1997). The motif-index and the tale type index: A critique. *Journal of Folklore Research*, (pp. 195–202).
- Finlayson, M. A. (2008). Collecting semantics in the wild: The story workbench. *AAAI Fall Symposium: Naturally-Inspired Artificial Intelligence* (pp. 46–53).
- Finlayson, M. A. (2011). The story workbench: An extensible semi-automatic text annotation tool. *Intelligent Narrative Technologies*.
- Friedman, S., Burstein, M., McDonald, D., Paullada, A., Plotnick, A., Bobrow, R., Cochran, B., Pustejovsky, J., & Anick, P. (2017). Learning by reading: Extending & localizing against a model. *Originally presented at the 2016 Cognitive Systems Conference (April, 2016), extended and published in the Cognitive Systems Journal (December 2017) as Advances in Cognitive Systems 5:77-96*.
- Gardner, M., Grus, J., Neumann, M., Tafjord, O., Dasigi, P., Liu, N. F., Peters, M., Schmitz, M., & Zettlemoyer, L. S. (2017). Allennlp: A deep semantic natural language processing platform.
- Halverson, J. R., Corman, S. R., & Goodall Jr, H. (2011). *Master narratives of islamist extremism*. Palgrave Macmillan.
- Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python. From <https://doi.org/10.5281/zenodo.1212303>.
- Jahan, L., Chauhan, G., & Finlayson, M. A. (2018). A new approach to animacy detection. *Proceedings of the 27th International Conference on Computational Linguistics*.
- Jahan, L., Mittal, R., Yarlott, W. V., & Finlayson, M. (2020). A straightforward approach to narratologically grounded character identification. *28th International Conference on Computational Linguistics (COLING 2020)*.
- Karsdorp, F., Kranenburg, P., Meder, T., Trieschnigg, D., & Bosch, A. (2012). In search of an appropriate abstraction level for motif annotations. *Proceedings of the 2012 Workshop on Computational Models of Narrative*.
- Liu, L., Hu, X., Song, W., Fu, R., Liu, T., & Hu, G. (2018). Neural multitask learning for simile recognition. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 1543–1553).
- Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. *Association for Computational Linguistics (ACL) System Demonstrations* (pp. 55–60).
- McDonald, D. D. (1992). An efficient chart-based algorithm for partial-parsing of unrestricted texts. *Third Conference on Applied Natural Language Processing* (pp. 193–200).

- Noy, D. N. (1954). *Motif-index of talmudic-midrashic literature*. Indiana University.
- Ofek, N., Darányi, S., & Rokach, L. (2013). Linking Motif Sequences with Tale Types by Machine Learning. *Proceedings of the 4th Workshop on Computational Models of Narrative (CMN'13)* (pp. 166–182). Hamburg, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., & Tsujii, J. (2012). brat: a web-based tool for nlp-assisted text annotation. *Proceedings of the Demonstrations Session at EACL 2012*.
- Su, C., Fukumoto, F., Huang, X., Li, J., Wang, R., & Chen, Z. (2020). Deepmet: A reading comprehension paradigm for token-level metaphor detection. *Proceedings of the Second Workshop on Figurative Language Processing* (pp. 30–39).
- Thompson, S. (1960). *Motif-index of folk-literature: a classification of narrative elements in folktales, ballads, myths, fables, mediaeval romances, exempla, fabliaux, jest-books and local legends*, volume 4. Indiana University Press.
- Thompson, S. (1977). *The folktale*. Univ of California Press.
- Yarlott, W. V. H., & Finlayson, M. A. (2016). Learning a better motif index: Toward automated motif extraction. *7th Workshop on Computational Models of Narrative (CMN 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.