Towards a Machine-Readable Classification System based on *Thompson's Motif-Index*: A Case Study on South American Indigenous Narratives[1]

Hacia un sistema de clasificación legible por máquina basado en Thompson's Motif-Index: un estudio de caso sobre narraciones indígenas sudamericanas

Jan Niklas JOKISCH
jan.jokisch@ae.mpg.de
Max Planck Institute for Empirical Aesthetics, Germany
https://orcid.org/0009-0003-1026-2995

ABSTRACT

Thompson's Motif-Index (TMI) is one of the most important classificatory references within folklore studies. However, leveraging its potential within contemporary digital methodologies presents us with a manifold of problems. The paper explores general issues and operational limits of the TMI, discusses possible solutions, and ultimately presents a proof of concept to reimagine motifs as labeled digraphs. It does so along a subset of 1,969 narratives from Johannes Wilbert and Karin Simoneau's *Folk Literature of the South American Indians* (1970–1992) as well as their 23,624 individual motif assignments.

KEYWORDS: Comparative mythology, Folklore, Classification, Network analysis.

RESUMEN

Thompson's Motif-Index (TMI) es una de las referencias clasificatorias más importantes dentro de los estudios folclóricos. Sin embargo, aprovechar su potencial dentro de las metodologías digitales contemporáneas nos plantea múltiples problemas. Este artículo explora los problemas generales y los límites operativos del TMI, analiza posibles soluciones y, por último, presenta una prueba de concepto para reimaginar los motivos como grafos dirigidos etiquetados. Lo hace a partir de un subconjunto de 1.969 narraciones de la obra *Folk Literature of the South American Indians* (1970–1992) de Johannes Wilbert y Karin Simoneau, así como de sus 23.624 asignaciones de motivos individuales.

PALABRAS CLAVE: Mitología comparada, Folclore, Clasificación, Análisis de redes.

1. The *Thompson's Motif Index* and comparative cross-cultural folklore studies

Stith Thompson's *Motif-Index of Folk-Literature*, frequently abbreviated as *Thompson's Motif-Index* (TMI), is a six-volume collection of over 46,000 motifs for the, as the subtitle promises, *Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jest-Books, and Local Legends*. Originally published in 1932–1936 and receiving a revision in 1955–1958, it had plenty of time to earn its place as *de facto* one of the most important references within folklore studies. Fundamentally, it tries to solve the

problem of the great unread,[2] which, in much the same way, underpins much of the digital humanities as a discipline:

> Our great libraries, enriched by the ceaseless activity of countless field workers and scholars, grow daily more difficult to explore [...] so that no man, however great his industry and skill in languages, can read the thousands of volumes in a lifetime (Thompson, 1955–1958, vol. 1, p. 9).

The TMI accomplishes this by synthesizing previous compilations and comparative studies—reportedly drawing from over 600 works[3]. Conversely, diverse projects have used it over the years to classify folk narrative traditions from all over the world, adding to the synthesis.

To understand the importance of the TMI, it is necessary to understand how motifs are conceptualized. According to Thompson, "[a] motif is the smallest element in a tale having a power to persist in tradition. In order to have this power it must have something unusual and striking about it" (1946, p. 415). Motifs are used as content elements that are remarkable enough to be recognizable and adaptable and as thus, fundamentally underpin studies of intertextuality and cultural evolution. Striking examples permeating Western culture include the Promethean episode of the "A1415. Theft of fire", the epistemological horror trope "D2065.2. Insanity from seeing strange sight", the David and Goliath constellation "L311. Weak (small) hero overcomes large fighter", and the classic propagandist justification for wealth inequality "J347.4. Rich merchant is poorer in happiness than poor man".

For the purpose of this study, it is beneficial to expound on Thompson's definition with Heda Jason's clever insight that motifs are units of content that are *context-free* and have *no variation* (Jason, 2000, p. 22). They are context-free, as they only gain meaning within a text in the interaction with other motifs, and they have no variation, as every

---

[2] "But as soon as scholars start to work on the archive of forgotten literature, techniques of close reading come up short. Problems range from the simple lack of time critics have to read closely all the texts that make up the great unread to the failure of some of these texts to signify in fashions that are meaningful using the criteria of close, formal analysis" (Cohen, 1999, p. 59).
[3] This estimate is proposed by Yarlott (2022, p. 4).

motif ultimately is atomistic. As Jason puts it "a motif has no variants; every variation is a new motif" (2000, p. 64). The interplay of these aspects makes motifs a powerful tool for classification. They are functionally flexible enough to appear in different narratives for entirely different purposes—the extreme case being a work and its parody—while being formalistically rigorous enough to distinguish between slight changes on the level of content.

Now, as Thompson claims, "the principle use of a motif-index is to display identity or similarity in the tale elements in all parts of the world so that they can be conveniently studied" (1946, p. 416). Johannes Wilbert and Karin Simoneau mirror this sentiment in the digital age when they prophesy that "the motif database can be manipulated in a variety of ways, thus opening the way to comparative cross-cultural research on a scale previously not possible" (1992, p. 51). An ideal computer-aided study based on the classificatory potential of the TMI would compare the motif content between narratives of different cultures and thereby establish similarities and differences in their narrative traditions. However, here is where the structure of the TMI fails itself. As motifs are conceptualized as context-free, every reliable measure of similarity between two narratives requires that they have at least a handful of motifs in common; however, as motifs are defined as units so elemental that they have no variations, the probability of two narratives sharing the exact same motif is slim.

We can illustrate this issue based on a concrete example. Johannes Wilbert and Karin Simoneau's 23-volume anthology series *Folk Literature of the South American Indians* (FLSAI), published from 1970 to 1992, classifies virtually every narrative it contains with motifs from the TMI. Within the context of our project the *Forager Folklore Database* (FFDB), we are working with a selection of 1,969 narratives from the FLSAI based on 19 different cultures. To this subset, 6,338 different motifs[4] are

---

[4] This number in itself is somewhat deceptive. Of the 6,338 motifs used, 3,106, i.e., almost 50%, are only assigned to a single narrative, making them useless for narrative comparison.

assigned, resulting in a total number of 23,624 individual motif assignments.

Following our idea of a cross-cultural comparative study, we can establish connections between narratives from different cultures based on their sharing of the same motifs. This results in 129,653 unique narrative pairings. This seemingly impressive number is quickly dispelled once we examine the number of motifs that underpin each pairing. Narrative pairings that rely on only one motif account for 105,320 of all pairings—that is roughly 81%. Obviously, almost no meaningful comparison can be built on a single shared motif. A more realistic, though still optimistic, assessment of at least five motifs leaves us with a meager 1,477 pairings.

If this still sounds like a workable number, we can approach the problem from another angle. The way our narratives are split up among the 19 cultures results in a total number of 1,747,646 potential narrative pairings. Figure 1 shows the percentage of actualized pairings by X or more shared motifs. As we can see, only 6% of all pairings are realized in our dataset. For our still optimistic benchmark of at least five motifs, we are down to 0.03%. That means that the overwhelming majority of our narratives are deemed incomparable—far from the best basis for cross-cultural comparative studies.
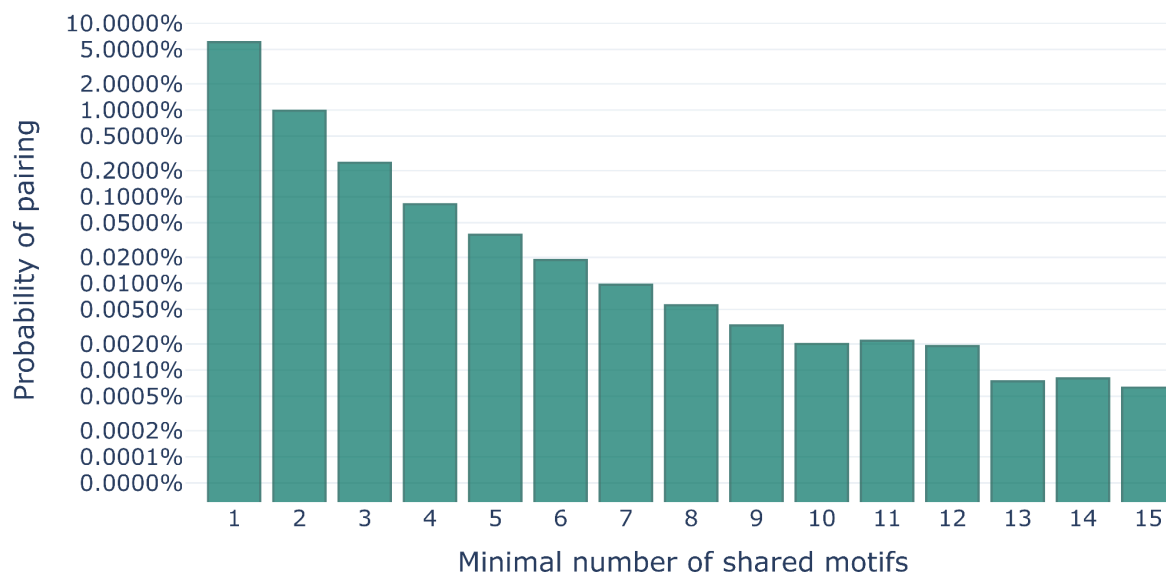
Figure 1. The probability that two randomly selected narratives of our subset have at least X motifs in common, given that they stem from different cultures (The y-axis is log-transformed to allow for easier readability). Source: own work.

We do well to keep in mind that different aspects artificially inflated these numbers. 1) The entire motif mapping was done by the same team and therefore exhibits encoder bias. We can assume that different projects would tag narratives differently, leading to lower cross-corpora agreement. 2) As the encoders frequently employed higher-level motifs, there is already an aspect of abstraction in the tagging that a corpus ideally should not exhibit. Thus, the 10 most frequent motifs alone make up 1,052 (4.5%) of all motif assignments (Table 1). One should note how general these motifs are. For instance, the higher-level motif "K810. Fatal deception into trickster's power" alone has 104 more specific sub-motifs that should have been used in its stead.

Motif	Number of mapped narratives
J652. Inattention to warnings.	186
D523. Transformation through song.	143
N440+WS. Secret learned.	117
D150. Transformation: man to bird.	103
J613. Wise fear of the weak for the strong.	90
R311. Tree refuge.	89

K910. Murder by strategy.	84
Q411. Death as punishment.	82
J580. Wisdom of caution.	80
K810. Fatal deception into trickster's power.	78

Table 1. The 10 most frequently assigned motifs in our subset. Source: own work.

3) The FLSAI expanded the TMI. This is a necessary step for most projects[5] and generally highly encouraged (Jason, 2000, pp. 61–65). However, as most projects proceed independently and no mapping between the different resulting expansions exists, individual taggings will frequently rely on motifs not available to any other project. For our subset of the FLSAI, 4,040 (17.1%) of the motif assignments rely on these added motifs and thus cannot be the basis of a pairing between a narrative from the FLSAI and one from another project. 4) The narratives all come from relatively homogenous groups, with some of them having been in cultural exchange, and all of them standing in relatively close phylogenetic relations as they all descend from the forager tribes that peopled South America about 14,000 years ago.

Thus, the TMI, as it stands, is a conceivably bad tool for the kind of comparative studies envisioned by Thompson. The question is: is there a way to improve the overlap between narratives based on their motifs without sacrificing the flexibility and rigor that make the TMI a powerful classificatory resource?

2.   Previous attempts to improve the motif overlap

I am aware of two previous approaches employed to make the TMI work for cross-cultural comparative research. While I believe that both ultimately fail in meeting the complexities of the TMI, their discussion will nevertheless help us gain a better understanding of the inherent inconsistencies and issues of the index and sound out potential pitfalls for

---

[5] Currently, our TMI has 51,710 instead of the original 46,299 motifs as we have digitized the motif extensions from Lindell (2006), Schmidt (2013), Wilbert & Simoneau (1992), and Van Deusen (1999).

our endeavor. These two approaches either 1) leverage the hierarchical structure of the TMI or 2) query the motif texts using different NLP methods.

2.1. Leveraging the hierarchical structure

The fundamental structure of the TMI is hierarchical. It is made up of 23 broad units called chapters, which split into a first order of divisions, two additional optional orders of divisions, sections, higher-order motifs, and finally the sub-motifs (Figure 2). It is this last order which one ideally should employ. This structure has its origin in a top-down approach, which aims at simplifying the process by which one finds a motif in the physical copy of the TMI. In the context of this paper, the structure also appears appealing as a way to cluster individual motifs into higher levels—be they higher motifs, sections, divisions, or chapters—increasing the overlap between narratives based on their motifs. Thus, two narratives do not need to have the same motif in common, but merely motifs that populate the same branch of the motif tree.
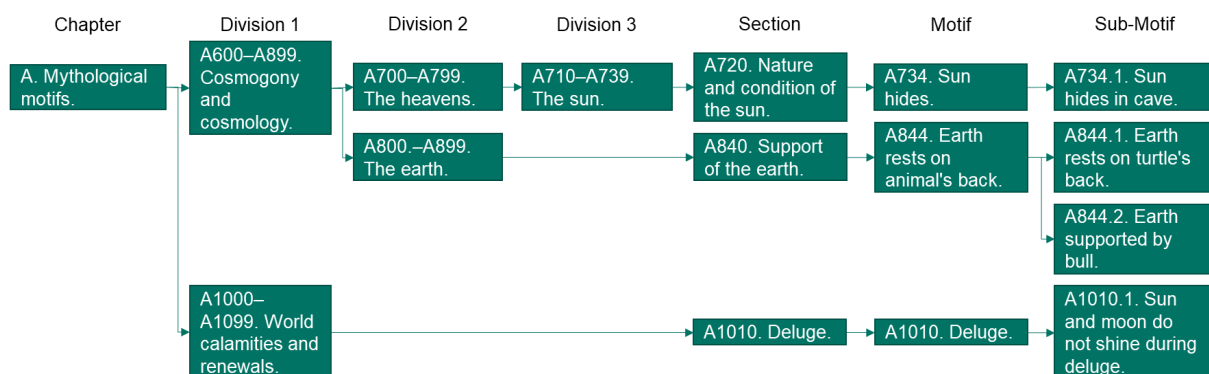

Figure 2. An illustrative excerpt of the hierarchical structure of the TMI. While Sub-sub-motifs, etc. also exist, they have been disregarded in this overview for reasons of simplicity. Source: own work.

The hierarchical structure of the TMI has been used as an aggregation method historically. Thus, the different volumes of the FLSAI generally include—in their introduction or annex—a discussion of the

distribution of motifs for the collection. For this, they aggregate the motifs on the levels of chapters and first divisions and draw conclusions about the collection based on the percentage of the different motif groupings. To give an example from the Warao volume:

> By far the most prevalent motif subgroup are *marvelous creatures* and *transformations*. The former nearly always refers to demons or spirits of which the most common are water spirits. The motif [that is the motifs of the division *marvelous creatures*] occurs 143 times (9%) in 68 (33%) of the narratives (Wilbert, 1970, pp. 31–32).[6]

I see at least three reasons why leveraging the structure of the TMI proves a poor solution. Firstly, higher levels do not present comparable steps of abstraction. While we might feel compelled to say that sections like "A1910. Creation of passeriformes" or "A2010. Creation of hymenoptera" express potentially meaningful categories as they bundle most motifs regarding the origin of a taxonomical order, a section like "A2170. Origin of miscellaneous animal forms" obviously reduces the entire endeavor to absurdity. The existence of these surplus categories, in this case a catchall containing earthworms, silkworms, lobsters, and multipedes, among others, is by its very definition—or not-definition as "miscellaneous"—asystematic. Other sections like "A120. Nature and appearance of the gods", which bundles any number of physical, psychological and accidental traits of deities, even appears counterproductive. Would we want to establish a connection between narratives if they only have in common that one possesses "A131.6. Horned god", another "A139.15. Greedy god" and yet another "A136.1.1. Deity rides boar"?

Secondly, the position of motifs within the tree structure is far from contentious. The motif "B120.0.1. Animals have second sight" is part of

---

[6] While not the topic of this paper, I don't want to leave this approach uncommented. Obviously, this kind of statistical approach on its own is somewhat meaningless. The TMI is neither exhaustive nor well balanced. The smallest chapter, "U. The nature of life", only has 172 motifs, while the largest chapter, "D. Magic", has 8,282 motifs. Thus, it should come as no surprise that the Warao volume contains significantly more motifs from chapter D (with 233 motifs) than from chapter U (with 29 motifs).

the section "B120. Wise animals" and not, as one might expect, the section "B140. Prophetic animals". The section "B670. Unusual mating between animals" is part of the division "B600–B699. Marriage of person to animal" despite the fact that the section only mentions odd pairings between non-human animals. In other cases, the classification is by modern standards downright offensive and equally questionable, for instance when the motifs "T462. Lesbian love" and "T463. Homosexual love (male)" appear next to "T465. Bestiality" and "T466. Necrophilism" under the section "T460. Sexual perversions".

The more general problem of the tree structure is augmented significantly by the approach the FLSAI took towards adding motifs to the TMI. They proceeded either by taking a general motif and making it more specific, which should generally preserve the vague logic of the tree structure, or deriving new motifs by changing the wording of an existing specific motif, which can fundamentally displace it in the index (Wilbert & Simoneau, 1992, pp. 51–52). As they acknowledge, the latter approach leads frequently to establishing a new motif in an unambiguously erroneous position within the index. To use their example, the motif "G466+WS. Lousing as task set by god" is derived from "G466. Lousing as task set by ogre" but obviously has no place within the chapter "G. Ogres" whatsoever, as it makes no reference to ogres.

Thirdly, motifs can be closely related without appearing within the same subsection. The most extreme case of this are (near) duplicates like "A1465.1. Origin of tattooing" and "A1595. Origin of tattooing" or "D491.6. Twelve stones unite to make one" and "F1009.4. Twelve stones unite to become one". Another case is motif implication—which we will discuss in greater detail later. Thus, the motif "A2012.0.1. Creation of honey-bees: transformed man" implies the motif "D182.1. Transformation: man to bee" and should lead to some linkage between narratives. However, no hierarchical approach could ever result in this,

since the earliest common ancestor that both motifs share is the root of the index.[7]

Therefore, leveraging the hierarchy of the TMI is a poor idea. The unidirectional look-up logic of the TMI simply cannot support our analytic bottom-up needs. We are forcing its structure to perform a task it was not designed for. The consequences are vague, often meaningless categories, and the loss of an unimaginable number of useful cross-references.

2.2. Querying the motif texts

As the TMI is fundamentally textual, it allows us to employ tried-and-true methods from the areas of natural language processing (NLP) and information retrieval to query the motif texts and establish connections between the motifs regardless of their position in the tree. In fact, at least three studies—Declerck & Lendvai (2011), Yarlott & Finlayson (2016), and Karsdorp et al. (2015)—did exactly that. The first two studies even saw themselves confronted with virtually the same challenge we are facing. As Declerck & Lendvai phrased it, "our objective is to pinpoint linguistic and perhaps cognitive properties and elements shared between sets of motifs which enable them to connect to one another" (2011, p. 153).

While all three projects proceed slightly differently, they ultimately employ the same broad approach. Thus, they use a barrage of different established NLP techniques—part-of-speech tagging, morphological tagging, dependency parsing, etc.—to encode the structure of each motif text. Then they semantically enrich the tokens using the WordNet

---

[7] The TMI mitigates this in parts using references—appearing in parentheses after the motif description and being formatted as a "Cf". followed by the referenced motif numbers. About 6,430 motifs in the entire index directly refer to at least one other motif. The purpose of these references is somewhat nebulous. Some are incredibly useful: "A2241. Animal characteristics: borrowing and not returning" has multiple references, among them one to "A2421.4. How partridge got voice. Borrowed from tortoise", which, in turn, refers to "A2422.4. Why tortoise has no voice". Others are of questionable purpose: "A1702. Creation of animals by creator" refers to "A0. Creator" and "A1700. Creation of animals" simply refers to the entire chapter "B. Animals". Yet others are erroneous: "E732.2. Soul in form of crow" refers to "B141.4. Dog with magic sight" and "F901. Repeated death" refers to "M341.2.4.2", which does not exist.

hyponym and hypernym structure (Fellbaum, 1998), which allows them to establish semantic relationships between different tokens. As Yarlott & Finlayson explain, one part of their pipeline attempts:

> to group motifs together into variant groups: groups of closely-related motifs that would be subtypes of a single motif in a motif index using semantic role labeling and the relations catalogued in WordNet. For example, if *cat uses the measuring trick* and *fox uses the measuring trick* were both identified, they would be grouped together as a single variant (2016, p. 7).

Declerck & Lendvai (2011) as well as Yarlott & Finlayson (2016) employ this specifically to establish connections between motifs. Thus, two motifs with the same structure, whose tokens for a specific part-of-speech are linked via a hypernym relationship —cat and fox are both hyponyms (or hyponyms of hyponyms, etc.) of the hypernym animal— can be subsumed under the same abstract motif. MOMFER, the *Meertens online motif finder* (Karsdorp et al., 2015), uses a virtually identical approach for slightly different purposes. The idea behind MOMFER is not to use the motifs for any form of computer-aided study but to offer fuzzy semantic search to simplify the process by which one might find relevant motifs to assign to a narrative manually.

It is tempting to use a similar approach to abstract from the concrete motifs in our subset. It would allow us to forgo the need for exact motif matches without sacrificing any control over the means of abstraction. Furthermore, it would work independently of the structure of the TMI, allowing us, at least in theory, to link motifs from entirely different parts of the index. However, sticking too closely to the motif text verbatim and entirely ignoring the structure has its caveats.

The TMI is frequently admired for the supposed consistency with which it expresses the same ideas syntactically with the same structures and semantically in the same terms—something unsurprisingly evoked by Declerck and Lendvai (2011, pp. 157–158). Anyone who has worked with

the TMI for an extended period will attest, however, that this is simply not the case.

Regarding the question of syntactic consistency, Torborg Lundell's study about the gender bias within the TMI is of immeasurable importance here. As she shows "female characters are victimized by the general semantic-syntactic inconsistencies throughout the index", especially with many female-related motifs "set in a passive mode with the girl being acted upon, thus promoting or supporting a socially supported tendency to regard the female as passive rather than active" (1986, p. 151). The Star-husband as well as Star-wife motifs, which are especially important for the Americas, illustrate these gender differences well. While it is the mortal man, who acts in the Star-wife motif's text, "A762.2. Mortal marries star-girl", it is the Star-husband, who is framed as the actor in the corresponding motif, "A762.1. Star-husband. Star takes mortal maiden as wife". The exact same relationship is phrased in opposite ways.[8] This extends into other areas as well. For example, the TMI employs the structures "helpful", "as helper", and "helps" seemingly synonymously. Thus, the section "B470. Helpful fish", which unsurprisingly contains motifs like "B471. Helpful shark" and "B474. Helpful salmon", also contains the motif "B470.1. Small fish as helper", which is not, as one might assume, phrased as "Helpful small fish".

Even in cases in which two motifs are syntactically identical, they can nevertheless be marred by lexical inconsistencies. For example, the TMI contains many etiological motifs. They frequently take one of three syntactically identical forms: "Origin of X", "Creation of X", and "Acquisition of X". While they appear to express different though related concepts, they ultimately seem to be used interchangeably. Thus, the motif "A1432. Acquisition of metals" has the predictable sub-motif "A1432.2. Acquisition of gold", but it also has the sub-motif "A1432.1. Origin of iron". And no, there is no other motif called "Origin of metals",

---

[8] Curiously, another motif flips the relationship again: "T111.2. Woman from sky-world marries mortal man".

which contains a sub-motif "Origin of gold". In fact, as far as I am aware, whenever the etiology of something is expressed—be this as an origin, a creation, or an acquisition—no motif exists that would describe one of the other two options. There seems to be no rhyme or reason to when these terms are used. Thus, there is "A1817. Creation of jaguar"[9] but "A2162. Origin of frog" and "A1446.2. Origin of the axe" but "A1459.1. Acquisition of weapons"..

Likewise, the TMI operates rather laxly with animal names. Frequently, it treats similar species as interchangeable. Thus, we can find the motif "A2494.5.4. Enmity between jackal and alligator". The TMI sourced it from India, despite the fact that alligators are not native to India, nor do jackals and alligators have any overlap in their global distribution. Likewise, the term fox is used for disparate species all over the world, among them the South American foxes (Lycalopex), which taxonomically are more closely related to wolves than to foxes.

Lastly, the hereditary dimension of the hierarchical structure of the TMI needs to be mentioned here. The phrasing of a single motif is not exhaustive in determining its meaning. The branches that lead to it contain contextual information that narrows down a potential interpretation. While a motif like "B62. Flying fish" allows for interpretations ranging from the banal to the fantastical, its section "B60. Mythical fish" makes clear that we are not talking about a member of the real-world fish family Exocoetidae, but of fish actually capable of flight. Likewise, the motif "A115.3. Deity arises from mist" is not describing an especially dramatic entrance, but is etiological in nature, as its section name explains, "A110. Origin of the gods"[10]. Conversely, the texts of two

---

[9] It should be noted that the bibliographic source that the TMI ties to this motif does not illuminate its choice of "creation". It refers to a summary of the mythology of the Mojo that states, "The Mojo also believed in a celestial Jaguar, father of all the terrestrial jaguars, who ate the moon" (Métraux, 1948, p. 424). This does not strictly imply an act of creation.

[10] Even MOMFER effectively suffers from this. While it is a brilliant and intuitive tool to query the TMI, its output is a list of motifs entirely bereft of any context—both regarding the heritage of the motifs and its immediate neighbors. Instead of becoming the tool through which one interfaces with the TMI, MOMFER ends up merely a tool that suggests

motifs can be verbatim the same, while describing different circumstances. The motif "D951. Tree produced by magic" belongs to the section "D950. Magic tree", and thereby describes a tree that has magic properties and was produced by magic. "D2178.8. Tree produced by magic", on the other hand, belongs to the section "D2150. Miscellaneous magical manifestations", which means that it describes a tree that although produced by magic can otherwise be unremarkable.

Due to these quirks of the TMI, I think it is highly dubious to employ a method that relies as heavily on the verbatim phrasing of a motif as any approach following Declerck & Lendvai (2011), Yarlott & Finlayson (2016), or Karsdorp et al. (2015) might do. The TMI, despite its best efforts, is not a straightforward tool. It is highly reliant on interpretation and knowledge of the folkloristic terminology and jargon, draws heavily from its complex tree structure, and is simply not consistent enough in its expressions.

3.   Motifs as labeled digraphs

The TMI is ultimately too multifaceted, wrought with biases, structurally complex, and intertwined with the discourses of folklore studies for any automated approach to capture it in a meaningful way. Therefore, what I suggest is to rewrite every single motif manually in the form of a network, more specifically a labeled digraph.

This approach is based on ideas already occurring in the previous literature on the TMI, especially in Jason's seminal study in an inconspicuous chapter of two pages entitled "How to Divide a Narrative into Motifs" (2000, pp. 60–61). Here, she expounds on an idea by Thompson (1946, pp. 415–416) by introducing a functional typography of motifs. The central insight that Jason provides is that motifs consist of characters, requisites—both frequently a living being or object coupled

---

motifs that one will have to look up in the TMI afterwards to determine their actual meaning.

with a remarkable quality[11]—deeds as well as, and here it becomes interesting, the coupling of characters with requisites and characters with deeds.

For example, we might have a character like "F445.1.1. Tobacco-spirit", a requisite like "D1212. Magic rattle" and a deed like "D2074. Attracting by magic". The tobacco-spirit is a specific kind of spirit (F400). The magic rattle is both an instance of a rattle, as it appears in other motifs, like "A1461.4. Origin of the use of the rattle" and "C830+WS6. Tabus connected with gourd rattle", as well as a more abstract instance of more specific forms of magical rattles like "D1610+WS15. Speaking rattle" and "D1646+WS. Dancing rattle". The deed of attracting someone or something by magic is just a way of phrasing a summoning, like it appears in a manifold of motifs, for example the section "D1420. Magic object draws person (thing) to it". If we combine spirits and summoning, we get "F404. Means of summoning spirits". If we specify the spirit as belonging to the domain of tobacco and specify the means of summoning as the magic rattle, we get "D1421.1+WS. Magic rattle summons tobacco spirits". In theory, we could keep going, specifying the character using the rattle, mental and physical faculties of the tobacco spirit, the provenance of the rattle, etc.

This modular structure of relationships can be taken as a model for reimagining the motifs in the TMI within a form well suited to express complex relationships in a rigorous manner: the graph.

3.1. Technical aspects of the network structure

Of the 6,338 motifs used in our subset of the FLSAI, I have transformed 1,834[12] into an RDF-inspired network structure. Every of

---

[11] For example, Jason decomposes the motif "F991.1.1. Bleeding knife" as "knife which has the ability to bleed" (2000, p. 60). Thompson himself says that "Commonplace experiences (...) may become so [i.e., worthy of note and thereby a potential motif] by having attached to them something remarkable or worthy of remembering" (1955–1958, vol. 1, p. 19).
[12] There is no greater significance to this number. It is merely what I was able to accomplish within two weeks of work.

these 1,834 motifs that is more complex than merely naming a remarkable person or object has been encoded as a set of triples in XML. I employed an associated RELAX NG schema to enforce a consistent vocabulary and structure. Figure 3 shows the encoding of a single motif—"A1724.1+WS2. Birds from body of slain ogre". Every relationship is expressed as an element r, which is made up of a subject, predicate and object (SPO) or a subject, predicate and attribute (SPA) triple. The reason behind this differentiation into two structures is purely cautionary. The SPO structure is more flexible with the same set of entities capable of populating both the subject and object position and doing so almost independently of the specific predicate used. The SPA structure, on the other hand, is significantly more limiting. There is a clear divide between the entities in subject and attribute positions and the set of attributes is systematically restricted through the relationship expressed in the predicate.

```xml
<motif xml:id="A1724.1_WS2">
  <name> A1724.1+WS2. Birds from body of slain
ogre.</name>
  <r>
    <s>bird</s>
    <p>originates_from</p>
    <o>act_of_transformation</o>
  </r>
  <r>
    <s>act_of_transformation</s>
    <p>process_has_component</p>
    <o>body</o>
  </r>
  <r>
    <s>body</s>
    <p>transforms_into</p>
    <o>bird</o>
  </r>
  <r>
    <s>ogre</s>
    <p>has_bodypart</p>
    <o>body</o>
  </r>
  <r>
    <s>ogre</s>
    <p>att_living</p>
    <a>dead</a>
  </r>
  <r>
    <s>x</s>
    <p>kills</p>
    <o>ogre</o>
  </r>
</motif>
```

Figure 3. XML representation of "A1724.1+WS2. Birds from body of slain ogre". Source: own work.

Based on our previous discussion of the limits and issues of the TMI, there are some things to note here. Regarding the motif text, we are not limited to modelling it verbatim. We can see this in the shift in the expression "slain ogre". I have opted to decompose it into its two central components: 1) the deed: someone or something kills the ogre

(<s>x</s> <p>kills</p> <o>ogre</o>) and 2) its result: the ogre is dead (<s>ogre</s> <p>att_living</p> <a>dead).

Partially, this need to rephrase the motifs is a direct consequence of the limited vocabulary that the RELAX NG schema enforces. As every new term has to be added manually before it becomes a valid choice, it encourages paraphrasing strategies based on the already defined vocabulary. This significantly reduces the risk of encoding similar concepts in highly divergent ways.

Furthermore, our representation of the motif contains significantly more information than the motif text does. The reason for that is that our approach allows us to model the context that informs the motif and to do so from the position of an informed practitioner. For instance, the etiological character of the motif, which is clearly expressed by the section title "A1710. Creation of animals through transformation" finds expression in the structure <s>bird</s> <p>originates_from</p> <o>act_of_transformation</o>.

The motif as a set of triples now functions as the basis for a network graph. Thus, every directed relationship that makes up a motif can be transformed into part of a directed labeled graph (Figure 4). The advantages of the network representation of the motifs is the unprecedented computational potential that comes with it and which we will leverage for the rest of this paper.

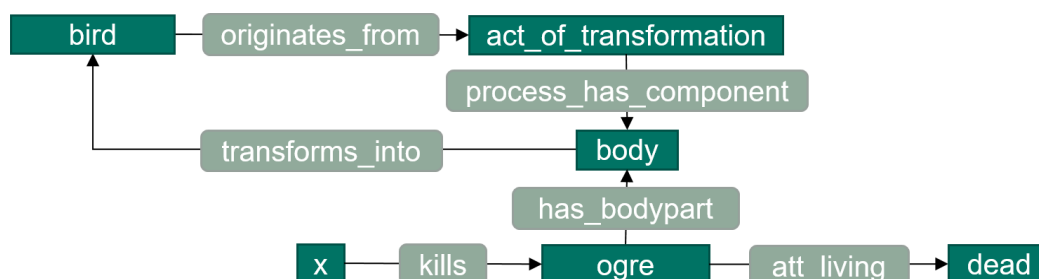A1724.1+WS2. Birds from body of slain ogre.

Figure 4. The labeled digraph representation of "A1724.1+WS2. Birds from body of slain ogre". Source: own work.

3.2. Naïve Motif Implication

Based on the network approach outlined, we can calculate connections between motifs that, in extent, could be used to establish connections between narratives even without them sharing any direct motif. That allows us to increase the overlap between narratives based on their motif content without sacrificing the complexity and richness the TMI offers. One of the easiest approaches to implement here is naïve motif implication. Naïve motif implication occurs whenever the network graph of one motif is a motif subgraph of another motif. This is a straightforward affair. If the graph of motif A is entirely contained in motif B, then every narrative that contains motif B also contains motif A. This is a direct application of the modularity of motifs as derived from Jason. As more complex motifs are made up of simpler motifs, the former imply the latter.

To illustrate this by the present example: We can easily see (Figure 5) that the motif "A1724.1+WS2. Birds from body of slain ogre" contains at least three other motifs as sub-graphs: "A1710+WS. Creation of birds through transformation", "G512. Ogre killed", and "G0. Ogres". Therefore, a narrative containing *A1724.1+WS2* can be linked with narratives that contain the same motif, the more general motifs *A1710+WS*, *G512* and *G0* or, conversely, a complex motif that implies any of the more general motifs, e.g., "K678. Cutting rope to kill ogre who is climbing the rope to reach his victim".
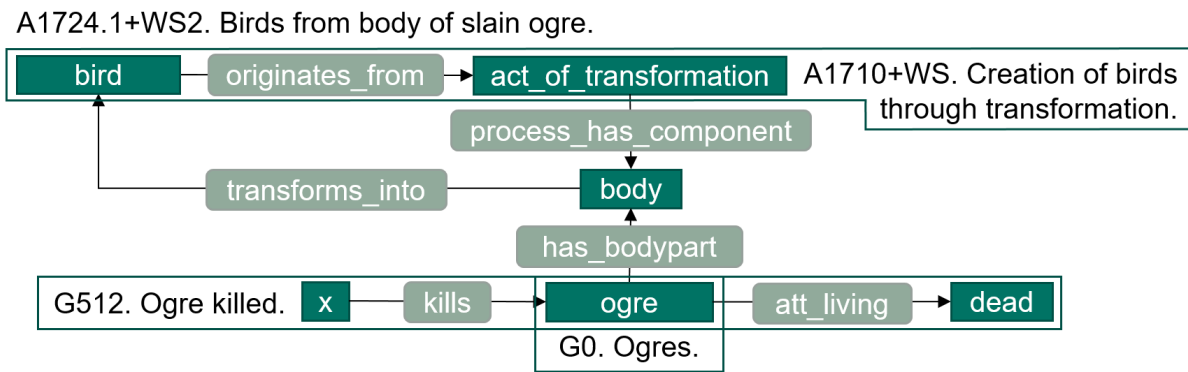
A1724.1+WS2. Birds from body of slain ogre.

Figure 5. The labeled digraph representation of "A1724.1+WS2. Birds from body of slain ogre" with three implied motifs highlighted in its subgraphs. Source: own work.

Of our 1,834 motifs, 414 (22.6%) verbatim imply at least one other motif for a total of 187 different motifs implied. Unsurprisingly, the most frequently implied motifs are broad groups of actors and requisites (Table 2).

Implied motif	Number of motifs implying it
G0. Ogres.	64
A0. Creator.	48
F420. Water-spirits.	29
G312. Cannibal ogre.	23
G10. Cannibalism.	23
D1711. Magician.	20
F500. Remarkable persons.	15
D800. Magic object.	14
P553. Weapons.	8
A220. Sun-god.	7

Table 2. The 10 most frequently implied motifs. Source: own work.

This does not mean that relatively complex motifs are not implied as well, albeit significantly less frequently (Table 3). Nevertheless, many of these implications connect motifs from entirely different parts of the TMI. The motif "N825.3. Old woman helper", for example, is implied by

the motifs "H971.1. Tasks performed with help of old woman" and "H1233.1.1. Old woman helps on quest".

Implied motif	Number of motifs implying it
A1414. Origin of fire.	5
A1710+WS. Creation of birds through transformation.	4
G512. Ogre killed.	4
D861. Magic object stolen.	3
A511.1. Birth of culture hero.	2
A736.3. Sun and moon as brothers.	2
A610. Creation of universe by creator.	2
N825.3. Old woman helper.	2
G72+WS. Mother eats her children.	1

Table 3. A selection of motifs, which are implied despite their complexity. Source: own work.

The impact of motif implication on the augmentation of the cross-cultural comparison of the TMI is hard to determine. As no comprehensive study on the practices of TMI practitioners exists, it is unclear to what extent narratives are paired not just with the most complex motifs applying to them but also with the motifs implied by those complex motifs.

We can get a general idea of this by applying our motif networks to our subset of the FLSAI. The implying motifs are matched with 908 of our narratives for a total of 2,309 assignments. Accounting for the fact that multiple motifs within one narrative can imply the same motif and that one motif can imply more than one other motif, we are left with 1,839 potential implied motif assignments. Of these, only 227 (12.3%) are among the motifs assigned by Wilbert and Simoneau. That means roughly half of our narratives are, on average, missing 1–2 motifs that they would definitely share with another narrative.

Surprisingly, the most common lacking but implied motifs for our subset are far from meaningless. They include relevant character motifs

like "G0. Ogres" (179 times), "D1711. Magician" (149 times) and "A220. Sun-god" (65 times), requisites like "D1275. Magic song" (71 times) and "F547.3. Extraordinary penis" (17 times), and fundamental yet complex motifs like "A1414. Origin of fire" (38 times), "A2610. Creation of plants by transformation" (22 times) and "D270+WS5. Transformation: man to sun" (18 times).

When we apply these missing motifs to the narratives and calculate the probability for cross-cultural pairings again, we get a noticeable increase (Figure 6). While the overall amount of pairings only increase by about 29%, the pairings based on at least three or more motifs on average increase by at least 145%. For our early soft milestone of at least five motifs, we have an increase of roughly 156% (Figure 7).
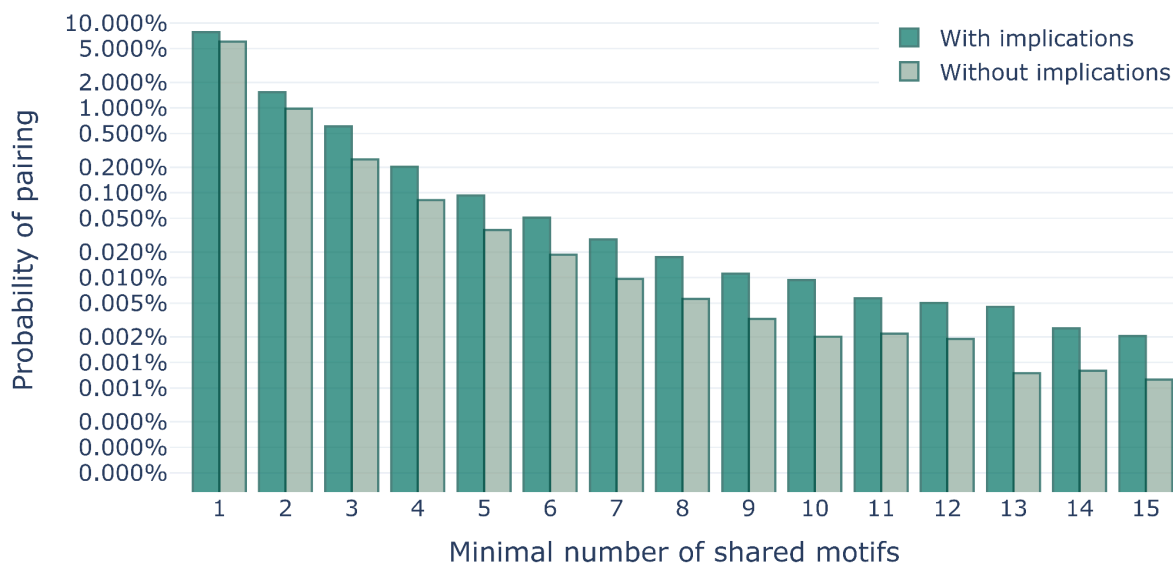

Figure 6. The probability that two randomly selected narratives of our subset have at least X motifs in common, given that they stem from two different cultures. Comparison between the probability with and without the missing implied motifs added. The log-transformation underplays the difference visually. Source:Source: own work.
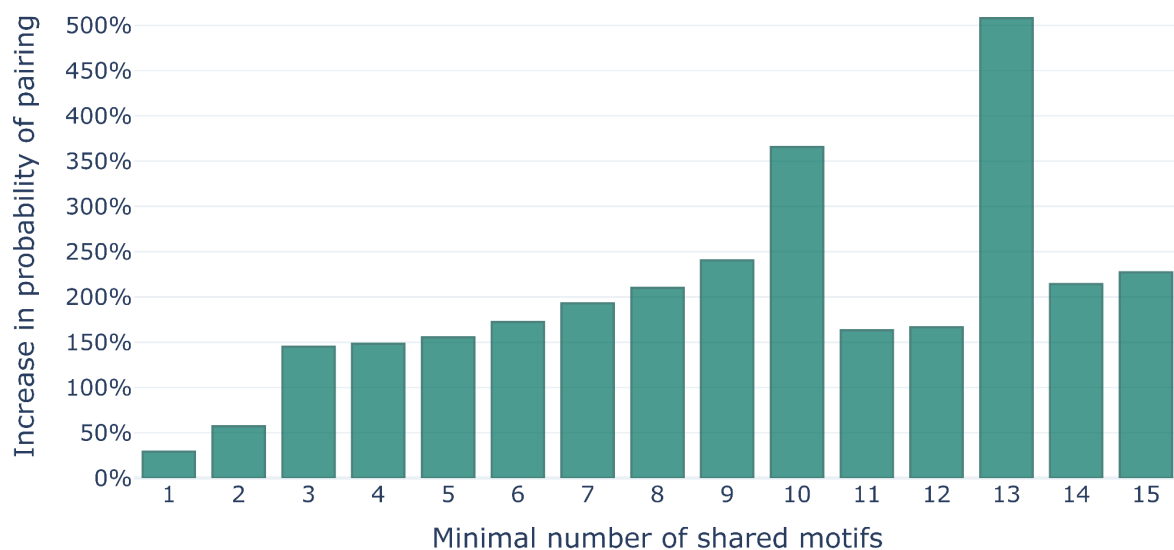
Figure 7. The increase in the probability of the cross-cultural pairing between two narratives, which share at least X motifs, after attribution of the missing implied motifs. Source: own work.

We should keep in mind that the increase approximated by our limited study is most likely significantly below its true value. For one, we are only looking at a subset of the TMI. Currently, every motif can at most imply 1,833 other motifs, i.e., all other motifs in the subset. But we are merely working with roughly 3.5% of all the motifs in the TMI. Almost 50,000 other potential implications for ever motif are not even considered right now. That means that the number of implying motifs as well as the number of motifs implied by each motif are most likely significantly higher.

Furthermore, the TMI is not exhaustive[13].While this is a trivial statement for all complex motifs and motifs specific to the flora and fauna of a region—the already huge division "A1700–A2199. Creation of animal life" does not contain an etiological motif for every single animal in existence—many implied motifs are missing from the index as well. For example, the motif "H1023.19. Task: bringing bundle of faggots without rope to tie them. Helpful snake coils himself about them" and "B524.3.

---

[13] See Alan Dundes for a biting criticism of Thompson's self-imposed prudish censorship within the TMI (1997, p. 198).

Helpful snake protects man from attack" both imply a motif "Helpful snake". As this is a common motif form with its own section, "B300. Helpful animal", we would expect to find it in the TMI, but no such motif exists. Therefore, we are lacking a not insignificant number of implied motifs that have simply never been defined in the TMI[14].

Finally, there is a theoretically and practically more interesting reason why I consider our measurements for implications below their true values. Our approach, so far, is naïve, i.e., semantically blind. It requires that one motif be verbatim the subgraph of another motif. This is extremely limiting. The implied motifs could certainly not just be subgraphs verbatim, but subgraphs, whose nodes are taxonomical abstractions from the nodes of the original graphs. Returning one last time to "A1724.1+WS2. Birds from body of slain ogre", as it implies "A1710+WS. Creation of birds through transformation", it should obviously also imply "A1710. Creation of animals through transformation". After all, birds are animals, and thus *A1710* is just a more general form of *A1710+WS*. Therefore, we are missing all implications based on the abstraction of our node labels.

As a proof-of-concept, we can introduce classes for some of the entities in our vocabulary, i.e., for animals, plants, foodstuff, humans, deities, ogres, body parts, and tools. If we use these classes to create an abstracted version of the implying motifs and check if an implying motif—in its original and/or abstracted form—implies another motif, we get greatly improved results. The number of implying motifs rises from 414 (22.6%) to 792 (43.2%)—an overall increase of 47.7%. The number of implied motifs does not rise as starkly but still changes from 187 to 218—a more modest increase of 14.2%. Thus, even the most basic of abstractions increases the number of implying and implied motifs significantly.

---

[14] Our approach is able to mitigate this. We could either generate motifs automatically by simply checking for all instances of two motifs sharing at least one triple that is not yet a motif or use the same approach to generate a list of motif candidates that we could evaluate and then add to the index manually.

4.   The need for a phenomena-based folk ontology

As we have seen, our approach is limited by its semantic blindness. We have not overcome the atomism of the motif, but merely moved it from the level of the motif as a whole to the level of its components. We will need a way to express the lexical relationships between our graph nodes to sufficiently abstract from the individual motifs. This is an issue of significant age within the motif discourse. Vladímir Propp has addressed it as far back as 1912 in his criticism of Alexander Veselóvskij's idea of the motif as indivisible:[15]

> However, the motifs which he [Veselóvskij] cites as examples do decompose. If a motif is something logically whole, then each sentence of a tale gives a motif. (A father has three sons: a motif; a stepdaughter leaves home: a motif; Iván fights with a dragon: a motif; and so on.) This would not be so bad if motifs were really indivisible; an index of motifs would then be made possible. But let us take the motif "a dragon kidnaps the tsar's daughter" (this example is not Veselóvskij's). This motif decomposes into four elements, each of which, in its own right, can vary. The dragon may be replaced by Koščéj, a whirlwind, a devil, a falcon, or a sorcerer. Abduction can be replaced by vampirism or various other acts by which disappearance is effected in tales. The daughter may be replaced by a sister, a bride, a wife, or a mother. The tsar can be replaced by a tsar's son, a peasant, or a priest. In this way, contrary to Veselóvskij, we must affirm that a motif is not monomial or indivisible. The final divisible unit, as such, does not represent a logical whole (Propp, 1928/1968, pp. 12–13).

As previously discussed, Declerck & Lendva (2011),[16] Yarlott & Finlayson (2016), and Karsdorp et al. (2015) approach this problem using WordNet. Given the scopes of these projects—which aim at Western texts and Western users—the use of WordNet is broadly speaking permissible. Admittedly, even this has its caveats. While the lexical relationship

---

[15] "Indivisibility" is largely a result of what Jason called "no variation". As every variant is a new motif, a motif cannot be decomposed and its basic units exchanged—not even with functionally similar units—without this resulting in the creation of new motifs.

[16] As far as their paper suggests, Declerck and Lendvai further streamlined the lexical classification by employing seven rather generic classes for nouns: "Hum (people and human-like creatures); Inst (groups and institutions); Abstr (abstracta); Concr (concreta); Geogr (geographical names); Tier (animals); Anim (life forms, non-human and non-animal)" (2011, p. 156).

between terms like "god" and "deity" or "kill" and "slay" is modelled sufficiently in WordNet, the same is not true for the technical terminology that the TMI employs. The use of "ogre" to classify every monster is not in accordance with its vernacular meaning. Thompson, following Pierre Saintyves (1923, pp. 299–306), employs the term "ogre" to denote any kind of monstrous being, be it a giant, a dragon, a cannibal, an evil spirit, or an animal of enormous size. The WordNet entry for "ogre", *ogre.n.02*, on the other hand, has the definition "(folklore) a giant who likes to eat human beings". Consequently, ogres are part of the class "giant" and do not subsume any of the other previous entities. In the case of South American forager cultures, then, it seems inappropriate to assume that WordNet—a taxonomy developed in a psychology department on the English language—should be the best way to model their world-picture.

Thompson already saw this issue when constructing the TMI, which, due to its hierarchical structure, fundamentally takes the form of a taxonomy. There are certain animal-related motifs that are not placed in the section where we would expect to find them based on their position in a biological taxonomy. Thus, "A2135. Origin of whale" is classed under "A2110. Creation of particular fishes" and "A2091. Origin of spider" under "A2070. Creation of miscellaneous insects". In both cases, the TMI explains itself in a footnote, informing us that the whale is, "[i]n folk thought, a fish" and that "[i]n folk thought the spider is classed as an insect and not with the other arachnida".

There is something odd about this decision. For one, the structure of the TMI is designed primarily to support the manual look-up of motifs. As thus, it is aimed at a western audience who will struggle with these bewildering choices in the classification. These are also some of the only instances of this kind of non-scientific classification. After all, does "A1837. Origin of seal" belong to "A1830. Creation of canidae and other carnivora", because seals, in folk thought, are generally classed as "other carnivora"? Furthermore, what does "folk thought" exactly mean? Certainly, not every culture considered in the TMI would class whales with

fishes and spiders with insects. Thompson's suggestion that folk thought is consistent across cultures and can be generalized sweepingly does, in my opinion, more harm than good.

In general, the hierarchical structure of taxonomies seems ill-suited for our goals. As every entity can only be classed in exactly one class, and this class again in another higher-order class, etc., it forces every entity into a single system of linear abstraction. An entity cannot belong to two branches at the same time. If we distinguish between fish-like and mammal-like animals, we are forced to assign the whale to one of these classes. In a taxonomy, no good way exists to highlight the fish-like and mammal-like qualities of the whale. A good alternative to taxonomies are ontologies and knowledge graphs. They do not rely on a hierarchical structure but employ classes, relationships, and attributes to link data in a network structure. These relationships can be class-based (e.g., "instance of") or connect two entities directly (e.g., "eats"). This affords the flexibility needed for the complex relationships we should account for.

What I would like to suggest is a phenomena-based ontology. To some extent, the TMI already follows this approach. The reason for classing whales as fish, for example, seems to stem directly from purely observable factors regarding their physiology, locomotion, and habitat. In other cases, the TMI goes so far as to assign different motifs to the same entity if it differs fundamentally in the way it is experienced. This is illustrated most clearly in the treatment of the planet Venus, which repeatedly appears in the TMI under phenomenologically distinct synonyms, e.g., "A781. Origin of Venus (planet)", "A781.1. Origin of Morning Star", and "A781.2. Origin of Evening Star".

Employing phenomena-based ontologies, we can class actors, requisites, and deeds based on external phenomenological factors that most likely led to these entities playing their specific role in a narrative. After all, as a narrative moves from one culture to another, a group changes, or a culture is exposed to new objects and concepts, we would assume that the narratives broadly preserve their structure, i.e., their

abstract line of thinking, but adapt the entities in their actor-, requisite-, and deed-positions to their current living situation—which are some of the ways we can arrive at the kind of variations that Propp discussed. We can see this in the way that foreign aspects like firearms are seamlessly included in the etiological tales about the origin of weapons or in the way trickster cycles of different cultures rely on many almost identical episodes but differ in their characters.

I would like to illustrate the most basic form of this idea with just one extended example of animal paramour tales from central and northern South America. Among the Chamacoco (Wilbert & Simoneau, 1987, pp. 252–272), Makka (Wilbert & Simoneau, 1991, pp. 210–212), and Sanema (Colchester, 1981, pp. 64–65), we find almost identical tales of women falling in love with animal paramours. The stories always play out the same way: the woman finds an animal, keeps it in a pot or a hole in the ground, and visits it periodically to have intercourse with it, until her husband or other family members finds and kills it.[17] These animals, interchangeably, are snakes, worms, and eels.[18] It is trivial to state that these animals resemble each other and that their phallic shape makes them ideal candidates for the physical and narrative function they perform in these narratives.[19]

Obviously, not all aspects that might lead to a specific actor, requisite, or deed filling a given position in a narrative can be reduced to information that an outsider might glean from mere observation of natural phenomena. As narratives are cultural products, they encode culture-specific information. This includes metaphysical properties of beings, objects, and actions —ritualistic qualities, roles within the

---

[17] The Ka'apor have a similar tale, which, with an etiological twist, explains the origin of penises from the murder of the rankuaĭ-ang—the ghost penis—a snake-like being that originally lived under the ground and satisfied women, but was killed by a foolish young man (Huxley, 1957, pp. 150–151).

[18] For the Chamacoco and Makka tales, the paramour is an eel. In the case of the Sanema tale, the issue is a bit more complicated. While the paramour is a worm, the Sanema class worms with snakes (Colchester, 1981, p. 115).

[19] In a similar vein, alligators/caimans appear as paramours frequently due to their permanently erect penises (for an overview see Zeller, 1983, pp. 58–82).

cosmology, function within shamanistic practice, etc.— but also something as simple as edibility. Which substances are considered fit for consumption is not so much a quality of the substance, but relies on cultural, social, ethical, and religious aspects. Thus, the addition of emic aspects to the general ontology seems worthwhile, as it would allow us to systematically consider entities that, although physically different, perform similar functions within a culture.

The creation of these emic, phenomena-based ontologies would greatly improve our abilities to abstract from the node labels and allow us to query the TMI for specific motifs in more complex ways. Regarding the former, we could abstract from our nodes based on classes as well as based on phenomenological and cultural factors, creating broad motif categories like "flying animal as ogre" or "origin of inedible plant with ritualistic usage". Regarding the latter, we could search for specific narratives based on their motif content and linked knowledge systems, finding narratives about tabu animals or the transformation of a person into an edible plant. This would contribute to the abstract goal that Thompson envisioned for the TMI: "The end of this study [i.e., the TMI] will have been attained if the multiform materials that it treats become thereby easier of investigation and more convenient for reference" (1955–1958, vol. 1, p. 11).

5.   Conclusion

This study obviously is nothing more than a proof of concept. It relies on a subset of a subset of the mythology of a single part of the world, as well as of less than 3% of all motifs in our extended TMI. As that, it cannot do more than give us a rough idea of what might be possible if we adapted the TMI into a machine-readable form based on labeled digraphs and emic ontologies. Regardless of whether one finds this approach particularly compelling, I hope I succeeded in making one thing unmistakably clear: the TMI —just like many other foundational scholarly works of the past— is too complex to be operationalizable with a

bit of quick and dirty NLP. In its clunkiness and size, it requires commitment. Either we spend the time and effort to translate the TMI into an adequate machine-readable form, or we simply do not use it for the kind of studies sketched out in this paper. Not many things in scholarship are as treacherous as a bad classification.

Obviously, the approach has operational limitations. In fact, some meta-qualities like tabu, punishment, task, and reward might not be expressible as graphs but necessitate hypergraphs to be modeled accurately. For instance, the step from the motif "D1822. Loss of magic sight" to "Q590+WS6. Punishment: loss of magic vision" might best be expressed by ascribing the meta-quality "punishment" not to any single node in *D1822*, but to the entirety of the motif graph—a form of "nesting" that hypergraphs introduce. Admittedly, this switch would have virtually no impact on the approach suggested in this paper, and I have opted to forgo it solely on the basis of simplicity. Even then, however, it is not certain if every true motif[20] is meaningfully expressible in the form of a (hyper)graph.

Furthermore, I need to mention that our own graph representations of motifs would obviously not eradicate the inconsistencies of the TMI altogether, but ultimately just suffer from other inconsistencies and biases—though hopefully greatly reduced in frequency and severity. That is the curse of any attempt at classification, as Borges so brilliantly illustrates in his *Emporio celestial de conocimientos benévolos*.

The biggest argument against following my approach is time. Considering that it took me about two weeks to encode 1,800 motifs, it would take years to turn the entire TMI into a set of graphs—keeping into consideration the creation of a coherent vocabulary, multiple encodings to

---

[20] The TMI contains many entries that I—following many others, e.g., Jason (2000, p. 62)—do not consider a motif. The reason for that is their complexity. They do not express a unit of content but a plot summary, e.g., "J2318. Numskull convinced that he is a thief. An article is stolen. Declared that whoever does not eat all his cheese is the thief. A stone is put in the numskull's cheese so that he cannot eat it all. He admits the theft". Therefore, they are not motifs but what is generally called a tale type—which has its own index (Uther, 2004). Thompson was aware of this, but apparently did not perceive it as a problem (1946, p. 416).

increase consistency, frequent revisions, etc.—not to mention the creation of even the most basic of phenomena-based ontologies. As the ultimate benefits of this kind of research are hard to gauge, I struggle to advocate outspokenly for the founding of a TMI consortium tasked with the creation and curation of a graph-based motif index. However, Berezkin and Duvakin's draft for an alternative motif index of folklore (2015-2018)[21] has been used for intriguing evolutionary and phylogenetic studies in recent years (Thuillard et al., 2018; d'Huy et al., 2023), and the advances in artificial intelligence and machine learning might make the dream of automatic motif attribution a reality.[22] Thus, an invigoration of the field of folklore studies seems possible, and would benefit tremendously from a machine-readable motif index of folklore.

REFERENCES

Berezkin, Y. E., & Duvakin, E. N. (2015-2018). *World Mythology and Folklore: Thematic Classification and Areal Distribution of Motifs*. *Analytical   Catalog*. https://www.ruthenia.ru/folklore/berezkin/eng.html

Cohen, M. (1999). *The Sentimental Education of the Novel*. Princeton University Press.

Colchester, M. (1981). Myths and Legends of the Sanema. *Antropológica*, *56*, 25–127.

d'Huy, J., Le Quellec, J., Thuillard, M., Berezkin, Y., Lajoye, P., & Oda, J. (2023). Little statisticians in the forest of tales: Towards a new comparative mythology. *Fabula*, *64*(1–2), 44–63. https://doi.org/10.1515/fabula-2023-0013

Declerck, T., & Lendvai, P. (2011). Linguistic and semantic representation of the Thompson's motif-index of folk-literature. In S. Gradmann, F. Borri, C. Meghini, & H. Schuldt (Eds.), *Research and Advanced*

---

[21] Accesible desde: http://www.ruthenia.ru/folklore/berezkin.
[22] Yarlott's dissertation (2022), which partially chases the goal of automatic motif attribution, while intriguing, ultimately falls short of the complexities of the TMI. What he manages to achieve is the attribution of 35 character and requisite motifs like "leprechaun", "chupacabra", and "babel" (2022, pp. 68–69).

Technology for Digital Libraries (vol. 6966, pp. 151–158). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24469-8_17

Dundes, A. (1997). The motif-index and the tale type index: A critique. *Journal of Folklore Research*, *34*(3), 195–202. https://www.jstor.org/stable/3814885

Fellbaum, C. (Ed). (1998). *WordNet: An Electronic Lexical Database*. Bradford Books. https://mitpress.mit.edu/9780262561167/

Huxley, F. (1957). *Affable savages: An anthropologist among the Urubu Indians of Brazil*. Hart-Davis.

Jason, H. (2000). *Motif, type and genre: A manual for compilation of indices & a bibliography of indices and indexing*. Suomalainen Tiedeakatemia.

Karsdorp, F., Van Der Meulen, M., Meder, T., & Van Den Bosch, A. (2015). MOMFER: A search engine of Thompson's motif-index of folk literature. *Folklore*, *126*(1), 37–52. https://doi.org/10.1080/0015587X.2015.1006954

Lindell, K. (2006). *Motif-index of Southeast Asian folk-literature: According to the system of Stith Thompson*, Lund University.

Lundell, T. (1986). Gender-related biases in the type and motif indexes of Aarne and Thompson. In R. B. Bottigheimer (Ed.), *Fairy tales and society: Illusion, allusion, and paradigm* (pp. 149–164). University of Pennsylvania Press.

Métraux, A. (1948). Tribes of eastern Bolivia and the Madeira headwaters. In J. H. Steward (Ed.), *Handbook of South American Indians* (vol. 3, pp. 381–454). Government Publishing Office. https://repository.si.edu/handle/10088/34599

Propp, V. (1968). *Morphology of the folktale.* (Trans. L. Scott, 2nd ed.). University of Texas Press. (Original work published in 1928). https://doi.org/10.7560/783911

Saintyves, P. (1923). *Les contes de Perrault et les récits parallèles: Leurs origines (coutumes primitives et liturgies populaires)*. E. Nourry.

Schmidt, S. (2013). *A catalogue of Khoisan folktales of Southern Africa*. (2nd ed.). Köppe. (Original work published in 1989).

Thompson, S. (1946). *The folktale*. The Dryden Press.

Thompson, S. (1955–1958). *Motif-index of folk-literature: A classification of narrative elements in folktales, ballads, myths, fables, mediaeval romances, exempla, fabliaux, jest-books, and local legends*. (Vols. 1–6.) Indiana University Press.

Thuillard, M., Le Quellec, J.-L., d'Huy, J., & Berezkin, Y. (2018). A large-scale study of world myths. *Trames: Journal of the Humanities and Social Sciences*, *22*(4), 407–424. https://doi.org/10.3176/tr.2018.4.05

Uther, H.-J. (2004). *The types of international folktales: A classification and bibliography, based on the system of Antti Aarne and Stith Thompson*. (Vols. 1–3). Suomalainen Tiedeakatemia.

Van Deusen, K. (1999). *Raven and the rock: Storytelling in Chukotka*. University of Washington Press.

Wilbert, J. (1970). *Folk literature of the Warao Indians: Narrative material and motif content*. UCLA Latin American Center Publications.

Wilbert, J., & Simoneau, K. (1987). *Folk literature of the Chamacoco Indians*. UCLA Latin American Center Publications.

Wilbert, J., & Simoneau, K. (1991). *Folk literature of the Makka Indians*. UCLA Latin American Center Publications.

Wilbert, J., & Simoneau, K. (1992). *Folk literature of South American Indians: General index*. UCLA Latin American Center Publications.

Yarlott, W. V. H. (2022). *Communicating with cultures: How humans and machines detect narrative elements*. *FIU Electronic Theses and Dissertations*, 4983. https://digitalcommons.fiu.edu/etd/4983

Yarlott, W. V. H., & Finlayson, M. A. (2016). Learning a better motif index: Toward automated motif extraction. *In 7th Workshop on Computational Models of Narrative (CMN 2016)* (vol. 53, pp. 7:1–7:10)*.* Schloss Dagstuhl–Leibniz-Zentrum für Informatik. https://doi.org/10.4230/OASICS.CMN.2016.7

Zeller, K. R. (1983). *Das Krokodil / der Kaiman in Vorstellung und Darstellung südamerikanischer Indianer*. Renner.