Of Lions and Yakshis

Ontology-based Narrative Structure Modelling for Culturally Diverse Folktales

Franziska Pannach [a,*], Caroline Sporleder [a], Wolfgang May [a], Aravind Krishnan [b], and Anusharani Sewchurran [c]

[a] *Institute of Computer Science, University of Göttingen, Germany*
E-mails: franziska.pannach@uni-goettingen.de, caroline.sporleder@cs.uni-goettingen.de, may@informatik.uni-goettingen.de
[b] *Department of Electronics and Communication Engineering, College of Engineering, Trivandrum, Kerala, India*
E-mail: aravindhp1999@gmail.com
[c] *Centre for General Education, Durban University of Technology, Midlands, South Africa*
E-mail: AnusharaniS@dut.ac.za

Editors: Antonis Bikakis, University College London, UK; Beatrice Markhoff, University of Tours, France; Alessandro Mosca, Free University of Bozen-Bolzano, Italy; Stephane Jean, University of Poitiers - ENSMA, France; Eero Hyvönen, University of Helsinki (HELDIG) and Aalto University, Finland
Solicited reviews: Kalliopi Kontiza, The National Gallery, London WC2N 5DN, UK; Albert Meroño, Vrije Universiteit Amsterdam, Netherlands; Three anonymous reviewers

Abstract. Vladimir Propp's theory *Morphology of the Folktale* identifies 31 invariant functions, subfunctions, and seven classes of folktale characters to describe the narrative structure of the Russian magic tale. Since it was first published in 1928, Propp's approach has been used on various folktales of different cultural backgrounds. ProppOntology models Propp's theory by describing narrative functions using a combination of a function class hierarchy and characteristic relationships between the *Dramatis Personae* for each function. A special focus lies on the restrictions Propp defined regarding which *Dramatis Personae* fulfill a certain function. This paper investigates how an ontology can assist traditional Humanities research in examining how well Propp's theory fits for folktales outside of the Russian-European folktale culture. For this purpose, a lightweight query system has been implemented. To determine how well both the annotation schema and the query system works, twenty African tales and fifteen tales from the Kerala region in India were annotated. The system is evaluated by examining two case studies regarding the representation of characters and the use of Proppian functions in African and Indian tales. The findings are in line with traditional analogous Humanities research. This project shows how carefully modelled ontologies can be utilized as a knowledge base for comparative folklore research.

Keywords: Computational Folkloristics, Ontologies, Folktales, Narrative Structure, Vladimir Propp

1. Introduction

Folk and Fairy tales are a substantial part of oral folklore and an intangible element of cultural heritage. They play an important role in the cultural heritage of regions, nations or cultural minorities. While German fairy tales have been collected and edited by the Grimm brothers at the beginning of the 19th century [1], their Russian counterpart Alexander Afanasyev collected more than 450 folktales of Russian and Slavic origin [2]. Afanasyev's tale collection later became the foundation of Vladimir Propp's theory *Morphology of the Folktale*, which was published in 1928, but only gained international momentum after being translated into English in 1958 [3].

*Corresponding author. E-mail: franziska.pannach@uni-goettingen.de.

1570-0844/0-1900/$35.00 © 0 – IOS Press and the authors. All rights reserved

This project aims to construct an ontology, ProppOntology, and a lightweight query system[1] for multicultural folktales following the morphological approach of the Russian folklorist Vladimir Propp.

The goal of this project is to investigate how a carefully modelled ontology can help assessing the intercultural differences between folktales with regard to their narrative structure. For this purpose, an OWL (Web Ontology Language) [2] ontology was created, and subsequently annotated a small corpus of African and Indian tales according to their Proppian functions and character representations.

The system permits storing the metadata about folktales and their publications, e.g. authors and editors, sources, or publishers, as well as the annotations[3] following Propp's *Morphology of the Folktale*, according to a well-designed ontology. An annotation contains the verbalisation of a Proppian function, e.g. the function *β Absentation* might be verbalised as "Each morning, Sibanda would sneak off to his food tree" [4]. With this, it allows users to query Proppian functions and their verbalisations, and to compare tales that share functions or characters with common traits, such as the the *Villain* who is an evil witch. Hence, ProppOntology provides folklorists with the necessary information to form an opinion on which of two or more divergent analyses is more suitable for the given tale.

The ontology is open for additions of tales from these or other cultural backgrounds. The number of potential comparisons provided through the lightweight query system, and the conclusions folklorists can draw from those, naturally grows with the number of annotations available. The ProppOntology can assist different approaches on intercultural folktale comparison, e.g. how a certain Proppian function is verbalised in different tales, or how the verbalisation of a function in the same tale changes with translation. To this end, 20 mostly sub-Saharan African tales in their English translation, and 15 Indian tales both in Malayalam (the native language of the investigated tales from Kerala/India) and English were annotated. This small corpus is used to illustrate the application of the system for comparative analysis.

[1] ProppOntology and the query system is available at https://teaching.gcdh.de/ontology/index.

[2] https://www.w3.org/TR/owl2-overview/

[3] Note that we distinguish between *annotations* of the narrative patterns of the folktale in the domain, and the technical concept of *OWL AnnotationProperties*.

Comparing Proppian analyses is tedious work and requires a lot of insight into his approach. As of yet and to our knowledge, there is no online system that allows to compare Proppian analyses of folktales to those of other tales or variants of the same tale. Therefore, contextualisation of Proppian analyses remains a manual task.

A description of the project with a focus on African tales can be found in [5].

The paper is structured as follows: Related work is discussed in Section 2. Section 3 introduces the application domain, i.e., Propp's *Morphology of the Folktale* and a short introduction to the Thompson-Motif-Index (TMI) and the Aarne-Thompson-Uther (ATU) types. In Section 4, an overview of the modelling approach is given. Section 5 describes the used sources of folktales. Section 6 presents the design approaches of the query system, giving more insights into the implementation in Section 7. Section 8 discusses the ontology-aided information extraction from tale texts, which were implemented as a proof-of-concept. Section 9 presents results and describes some use cases from the application point of view. Limitations and future work are discussed in Section 10. This paper ends with a conclusion in Section 11.

2. Related Work

Federico Peinado et al. [6] modelled a description logic ontology, ProppOnto, based on Proppian functions. Their work is probably the closest to the project presented in this paper. They implemented Proppian functions and some additional subconcepts regarding persons, places and objects as ontology classes. Their description logic foundation was used to generate folktale plots using Knowledge-Intensive Case-Based Reasoning (KI-CBR). Similar to ProppOntology, a significant amount of additional domain knowledge was modelled to achieve their goal of plot generation.

In order to achieve a temporally sound story line, they used relations and concepts from the CBROnto case representation structure [7]. In a two-stage approach, they first generated a raw plot from Proppian functions which were then filled with a textual representation.

Furthermore, Peinado et al. [6] made some design choices that were not pragmatic for the intercultural comparison ProppOntology aims to achieve, e.g., the annotation properties of their classes do not include the literal that represents a function in a function se-

quence, such as Sequence 1 in Section 3.2. As the users of their generator software do not directly interact with the ontology, this is not a drawback of their work. Since the purpose of ProppOntology is assisting scholarly Proppian annotations, the identification of functions through their literal representation is critical.

Most importantly, while the generator application ProtoPropp is still mentioned on the author's website, Peinado et al̇do not provide its source code, and ProppOnto itself is not publicly available.[4]

Thierry Declerck et al. [8] created an ontology that modelled Proppian functions as classes with a vast set of rdfs:labels in English, German, and Russian. The Internationalized Resource Identifiers (IRI) of their classes contain a short description of the function according to the corresponding literal, e.g., *Delta1*. The functions in Declerck's work are not grouped into the five categories defined by Propp. Furthermore, they did not provide object properties for functions, such as those modelling sequential order or those that connect a function with the corresponding tale it appears in. The extensive labels and rdfs:comments provided by [8] were found a very useful addition to the existing ProppOntology. Therefore, they were added to ProppOntology and converted to skos:prefLabels for the English labels, and skos:altLabels for the German and Russian labels.

Declerck et al. [9] also built an ontology on the Thompson-Motif-Index in combination with Aarne-Thompson-Uther types. These two indices are core instruments for traditional folktale research, as described in Section 3.3.

Declerck et al. provide rdfs:labels in English and German, motifs and types appear both as classes in the ontology and as individuals [10]. Their ontology also includes a set of additional motifs as defined by the ETrap project[5]. As for now, only the core ATU classes and TMI motifs have been imported into the ProppOntology. While the population of the ontology with motifs and tale types as individuals technically does not follow the approach of the ProppOntology, they were nevertheless imported to allow possible reuse at a later stage in the project. Following the approach of the project presented in this paper, motifs and tale classes should have only been added as individuals if they occur in a tale that is annotated.

The MOMFER project [11] provides an online search interface for TMI motifs. They also provided a quantitative analysis on the distribution of motifs within their broader categories, finding that motifs or magic, mythological motifs, and motifs of marvel are the most prevalent within the TMI. In a number of case studies, they evaluated the use of motifs in geographical contexts, or the distribution of motifs with respect to different genders of folktale characters [11], [12]. While following a motif-based approach that is different from the structural approach followed by ProppOntology, their work is an important showcase of the possibilities that the use of computational methods can bring to the field of folkloristics. Especially their spatial analysis on motif reuse is similar to what ProppOntology aims to achieve with regard to intercultural comparison folktales.

Nikolina Koleva [13] built the folktale ontology *Monnet* that modelled family relations of characters. She used SWRL[6] rules for deriving property instances, such as *uncle(x,y)* and property values, such as *isEvil = true*. These rules were then used to automatically populate the ontology by iterating through a text, searching for semantic cues that introduce fairy tale characters. Koleva used NooJ grammars to detect the entities for populating the ontology. The automatic extraction of characters and role attribution from fairy tale texts seemed to work comparatively well, albeit on a small number of folktales.

Additionally, Koleva provided labels, although mistakingly annotated as dc:language fields, for the character classes in German, English, Russian, and Bulgarian. Those fields were carefully transformed into skos:prefLabel and skos:altLabel.

However, her characters lack certain features such as their verbalisation, or information about the tale they appear in. Therefore, in this project approached semi-automatic population in a different manner.

Nonetheless, the class hierarchy of the characters, including the SWRL rules, and the individuals resulting from the automatic population were imported to the ProppOntology as they are a valuable addition to the ontology beyond Propp's definition of character roles, especially with regard to the prominent motifs of family relations in folktales from different origins. Parts of the ontology were excluded, including *Body of Water*, *Event*, or *Body Part*, as they were found to

[4] In personal communication with the author of [6], the authors were able to have a look into the ontology.

[5] https://www.etrap.eu/

[6] Semantic Web Rule Language, https://www.w3.org/Submission/SWRL/

be insufficiently represented, e.g. *Body Parts* only included the classes *Legs* and *Wings*. The class hierarchy was restructured in order to resolve slight inconsistencies, e.g. *Parent* was not originally a subclass of *Relative*.

Since *Monnet* was specifically created to be used on Proppian folktales and it provides multi-lingual labels for classes, Koleva's work was preferred over other genealogical ontologies despite its shortcomings. Furthermore, *Monnet* was populated with individuals from the Russian tale *The Magical Swan-Geese* [2] which were a valuable addition to ProppOntology.

With regard to the structure of African tales, the work of Uta Reuster-Jahn [14] provides insights into the endings of tales of the Merwa people in Tanzania. She elaborates on the lack of a reward situation at the end of a tale, and argues that endings are more a moral resolution, that is represented either by the punishment of the main character or a gain in knowledge and morality of a community. Chukwuma Azuonye [15] investigates the fitness of Propp's theory for African Tales. His choice of the *Transfiguration* function at the end of his function sequence follows the same argumentation. He claims that the result of the tale is an increase in morality for the entire community, i.e. the children now follow their parents' rules because they learned about the destiny of the victim. While technically, the *Transfiguration* function is used to represent physical changes, e.g. in clothing of the hero, it shows that Proppian functions might not be well-suited to represent *Dénouement* situations in African tales. This hypothesis is investigated in the Results and Evaluation Section 9.

3. Description of the Domain

3.1. Morphology of the Folktale

The Russian folklorist Vladimir Propp introduced 31 invariant functions, shown in Fig. 1 describing the morphology of the Russian magic folktale.

In his work *Morphology of the Folktale*, he introduces seven classes of Dramatis Personae, i.e. agents, within a story: *the hero, the donor*, who provides the *hero* with means to overcome the *villain*, *the dispatcher, the helper, the false hero, and the princess/her father*.

He argues that narratives of (russian) folktales always follow a pattern that can be derived from his set of functions. Narrative functions, such as *XXXI Wedding W*, are strictly defined and specify recurrent units from which the tales are constructed. They follow a theory-inherent order, indicated by roman numerals. Furthermore, they are identified by a symbol, either a latin or greek letter, or an abbreviation which represents the function, e.g. *W* (*Wedding*), ↑ (*Departure*) or ↓ (*Return*). A function is always tied to specific Proppian characters, e.g. the Wedding function only applies if the *hero* character marries the *princess* (or a character that fulfills the narrative role of a princess). If a wedding between two other characters takes place, or if it appears at any other point than the end of the plot, the function does not apply. Hence, a story line is constructed around a subset of Proppian functions.

A sequence of functions represents the plot of a tale and is encoded in a string of function literals, as shown in Sequences 1 and 2 in Section 3.2. A function sequence is constructed from a subset of Proppian functions in their specific order. According to Propp, a tale can consist of one or multiple sequences, in the latter case, he calls them *moves*. A tale can consist of serial moves, e.g., two subsequent stories in a tale, or embedded, where one move interrupts the story line of another move, or even in an interleaved way, being told in parallel.

Propp defined four axioms regarding the narrative structure of folktales [3, p.21-23]:

1. Functions of characters serve as stable, constant elements in a tale, independent of how and by whom they are fulfilled. They constitute the fundamental components of a tale.
2. The number of functions known to the fairy tale is limited.
3. The sequence of functions is always identical.
4. All fairy tales are of one type in regard to their structure.

Furthermore, he grouped his functions into five categories: *Preparation*, i.e. the initial functions and first appearances of the main characters, *Complication*, in which the act of misfortune or villainy takes place, *Functions of the Donor*, where a helpful figure provides the hero with means to overcome the villain, the *Struggle* between hero and villain, and *Dénouement* in which the heroes are rewarded for their action. Functions belonging to the preparation category are represented by Greek letters. An overview about the main functions is given in Fig. 1 [16, p.4].

Propp's function categories and specific functions from the point of view of the application domain are

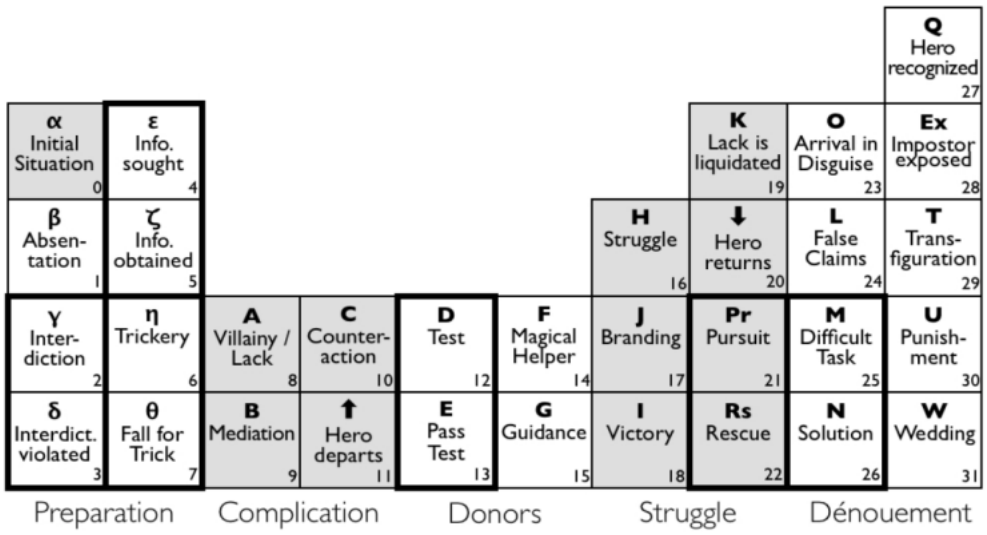


Figure 1. Schematic overview of Proppian functions. Bold lines represent function pairs, numbers on the lower right represent the order in which the functions appear. Contrasting colors indicate function grouping. Source: [16, p.4]

discussed in the results in Section 9 at the end of this paper. In the next section, some of them are illustrated exemplarily in a concrete use case.

3.2. Propp and African Folktales

Structural analysis of African folktales and the applicability of Proppian functions to them has not been introduced without critical side-eyeing. Daniel J. Crowley was sceptical whether Propp's approach was applicable to folklore studies, because in his opinion it was "doing too much violence to the variant nature of tales" [17, p.130].

Since then, studies that investigate the fitness of Propp's theory for African Tales have been conducted. A prominent example is the "Morphology of the Igbo Folktale" by Azuonye [15].

Azuonye published a morphological analysis of the *Obaraedo* tale in 1990 [15]. The same tale has been analysed by Ikechukwu Okodo in 2012 [18].

In the tale, the girl *Obaraedo* is left alone at home by her mother, who gives her specific instructions on how to prepare her food and orders her not to go outside at a specific time of the day, as an evil spirit will come and steal her nose. The girl disobeys her mother, which results in her nose being stolen. After the parents return, the father learns about his daughter's misfortune, and goes out to consult a wise herbalist, the *dibia*. The latter accompanies the father home and defeats the evil spirit. The girl receives her nose back and the kids of the community learn to obey their parents.

Both analyses explain how Propp's functions are represented in the *Obaraedo* tale. While Azuonye did not provide the full text of the tale, Okodo included the text translated to English in his article. From the explanations Azuonye gave on how the Proppian functions appear in the untranslated text, it becomes apparent that both of them worked with the same version of the tale.

However, Azuonye's findings regarding the structure of the tale are significantly different from Okodo's analysis. Due to Propp's formalistic approach, both findings are easily comparable. Azuonye defines the function sequence of the *Obaraedo* tale as

$$\alpha\beta\gamma\delta\theta aBC \uparrow HIK \downarrow T. \tag{1}$$

While Okodo defines the sequence as

$$\beta\gamma\delta\epsilon\eta\theta aBC \uparrow FGHIK \downarrow \tag{2}$$

The first difference that the reader encounters between both analyses is that the *initial situation* α does not appear in Sequence 2. The initial situation is often omitted since according to Propp it should not be regarded as an own function, but as "an important morphological element", which rather introduces the hero and the circumstances in which the tale takes place [3, p.25].

In Sequence 2, the preparatory functions include the tuple *Reconnaissance* ϵ, and *Trickery* η while in Se-

quence 1, *Violation δ* is directly followed by *Complicity θ*.

Following the *Departure* ↑ function, Okodo identifies the function *Provision or Receipt of Magical Agent F*. Propp's defines this function as the *Hero* acquiring the use of a magical agent [3, p.43]. Therefore, it can be assumed that Okodo sees the herbalist as the hero of the tale. In that case, Okodo's analysis of the tale is inconsistent, since he uses the *Departure* function when Obaraedo's father leaves the village to summon the herbalist. The *Departure* function, however, is specified for the departure of the *hero*. Both Azuonye and Okodo use the functions *Struggle H* and *Victory I*, when the herbalist/dibia fights the spirit, indicating again that he fulfills the role of the hero in the tale.

Additionally, they both define the *Departure* function as departure of the girl's father, but the *Return* function as a function of the herbalist/dibia.

In that sense, they are both separating the action described in the functions from the Dramatis Personae who fulfill them. This shows a rather free interpretation of what Propp clearly defined as "The Functions of the Dramatis Personae" [3].

These differences between two folkloric analyses show that the interpretation of Propp's functions is not universal, nor is there only one correct sequence of functions per tale. Even for Russian magic tales that were annotated by Propp, trained and untrained annotators did not always produce the same analysis as Propp himself. While trained annotators performed better than untrained annotators, a certain vagueness in Propp's descriptions of *Dramatis Personae* and functions leaves room for interpretation [19].

Furthermore, for the specific case of the comparison of the *Obaraedo* tale as described above, the annotators' comments on why a function was chosen were added to the function instance as an rdfs:comment. This way, users can comprehend the annotators' reasoning on why a particular function was chosen.

ProppOntology can facilitate a comparison between existing analyses like those by Azuonye [15] and Okodo [18], and those of tales from other regions. Existing annotations can be accessed easily through SPARQL queries or by accessing the triple search of the lightweight query system. Therefore, the system allows the study of the Proppian morphology interculturally and language-independently which might lead to new findings in folktale research.

3.3. Motif Indices

The Aarne-Thompson-Uther index (ATU) [20] is used to classify a tale into exactly one class, the *tale type*. For instance, the tale of the *Frog Prince or Iron Henry* [1] falls into the ATU class *440 - The Frog King*, which belongs into the broader category *Magic Tales*. Type classes are relatively wide, describing the main story line of the tale. Therefore, each tale can only have one ATU type. Tale types also indicate the relation of tales that belong to the same class. The Aarne-Thompson-Index was first published in 1910, revised by Hansjörg Uther and republished as ATU-Index in 2004.

In contrast, the Thompson-Motif-Index [21] is more fine-grained, describing single motifs, i.e. "recognizable object[s], character[s], or event[s]" [17, p.127], such as characters, actions, or numerical patterns. A tale can contain more than one TMI motif, e.g. the *Frog Prince* tale includes motifs such as *B211.7.1 Speaking Frog*, *P40 Princesses*, *P23 Children and Parents*, *P320 Hospitality*, or *D935 Transformation: Frog to Person*.

4. Modelling of the Domain as an Ontology

The choice to model Propp's theory as an ontology was influenced by the flexibility an ontology can provide, e.g. in contrast to a more rigid database approach. Classes and instances can be considered "static facts" of a tale, whereas relations describe the dynamics of the interaction between characters. Furthermore, the ontology-based approach allows us to infer knowledge where annotations are incomplete. The use of an OWL ontology allows us to consider the class hierarchy of Proppian functions, and to describe the instance level of the tales, their instances of Dramatis Personae, and the specific relationships between them that make up the instances of the Proppian annotations.

Since the purpose of ProppOntology is narrative annotation and queries on those annotations, and not for folktale generation like other ontologies, e.g. [6], it focusses on representations of Proppian functions that allow multiple ways of querying for narratives, as described in Section 6. To allow both character-focussed and function-focussed queries, functions have representations as ontology classes and as relations.

Therefore, each Proppian function in a tale has at least two representations within the ontology: as instance of a class, and as one or more relations between

instances of character classes, and between the function instance and the characters.

As an example, the Indian tale *Kathanar and the Yakshi* in which the hero Kathanar kills the Yakshi, as depicted in Fig. 2 is investigated. The object property *defeats* is used to connect the two instances of *Hero* (Kathanar) and *Villain* (Yakshi). Additionally, the appearance of the function *Victory I* is annotated as an instance of the *Victory* class that holds the verbalisations in both English and Malayalam. In addition, the instance of *Victory* is connected to the characters by :hasDefeated and :hasDefeater properties with the corresponding ranges (Villain and Hero). Table 1 in Section 6.3 gives some examples of these restrictions.

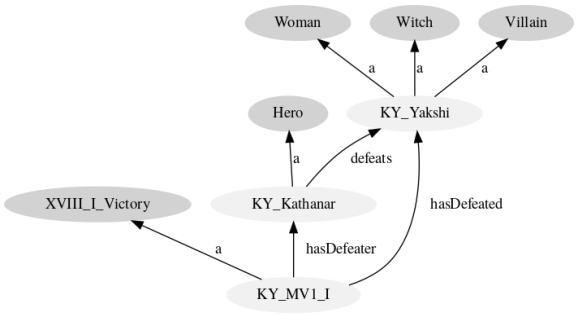


Figure 2. Example illustration on the representation of Proppian functions as class and object properties

ProppOntology includes bibliographical information about original publications of the tales. Anthology individuals provide metadata such as title or date of publication. Editors and authors of folktale collections are represented as individuals, since many of them are influential in their fields, e.g. Grimm and Grimm, Afanasyev, or Harold Scheub for tales of the Zulu people. In the future, ProppOntology can be extended with respect to the relationships between a real person and a specific tale beyond the publication, e.g. how collectors influenced the story telling.

In the system, and for visualization, URIs such as *O_ 2012_F* are used for the appearance of the evil spirit in the 2012 version of the *Obaraedo* tale. The specific individual can then hold further information on how it is verbalised in the given tale.

To extend the available classes of characters, ProppOntology was extended by including generational hierarchies from [13], see Section 2.

ProppOntology provides multilingual information for classes, supplied by either ourselves[7], Koleva [13] for the generational classes or Declerck et al. [8].

For this purpose, common vocabularies, such as SKOS [22] or the Dublin Core metadata schema [23] are used. Additionally, common owl:AnnotationProperties were included, such as rdfs:label, rdfs:comment[8], e.g. to explain why a specific Proppian function was chosen in a particular analysis.

Basically, the ontology contains information analogous to a traditional database schema: class hierarchy, domains and ranges of properties, required properties. Therefore, its core is in $\mathcal{ALC}$. It also uses inverse roles, subroles and functionality requirements, which extends it to $\mathcal{SHIF}$/OWL-Lite. Further extensions that use counting, nominals and datatypes (especially dates), which still is in $\mathcal{SHOIQ}(\mathrm{D})$, as supported by the common tools (here, mainly Protégé) should not be excluded.

5. Data Sources

The African tales were taken from a number of anthologies. To achieve a broad representation, a "healthy" mix between scholarly collections of tales, and typical children's stories was selected. This includes Harold Scheub's collection *African Tales* [24] (2005), Nick Greaves' children's book *When Hippo was hairy and other tales from Africa* (1990) [25], Children of Wax (1989) [4] by Alexander McCall Smith, and Phillis Savory's *Bantu Folk Tales From Southern Africa* (1974) [26].

Secondly, a small corpus of Indian tales, from the state of Kerala, published in Malayalam with their English translation, was collected. The tales in Malayalam have predominantly been taken from *Aithihyamaala* [27], a corpus of all the prevalent legends in Kerala written in the 20th century. All the stories, history, mythology, and romance of the Keralite community of the time, are presented in 126 articles. It represents the social and cultural life in the state at that time, and popularised characters, such as *Kayamkulam Kochunni*, *Naranathu Bhrandan* and *Kadamattathu Kathanar*.

The book is still an indispensable reference for historians of the Keralite society, which lacks in historical

[7]Special thanks to Siya Sikobi and Nokubonga Mkhize for the translation into isiZulu.

[8]https://www.w3.org/TR/owl-ref/#Header

record keeping. The English versions of the tales have been extracted from a translation of the book *Aithihyamaala, The Great Legends of Kerala* [28].

To encompass poetical literature in the scope of the study, some stories have been taken from the famous *Vadakkan Pattukal*, a collection of Ballads in Malayalam. These have survived by oral passage from generation to generation, and are believed to have been written down in the 17th or 18th century. There may have been some additions or reductions over time, but they still remain largely intact. The epic poem *Poothapattu* has also been included in the corpus [29, 30].

6. Ontology Design

The choice to model Propp's theory by using an ontology has two main motivations. Firstly, the functions are highly hierarchical as they are divided in categories, functions, and subfunctions. Secondly, the use of an OWL ontology allows us to represent the Proppian functions not only as classes within the ontology, but additionally to model the connection between the instances of the subclasses of *Dramatis Personae*, and the relationships between character and Proppian function, as shown in Fig. 2.

This approach allows us to query not only instances of functions, but also the relationships they represent between characters in a tale. After all, the functions are defined as "Functions of the Dramatis Personae" [3] and should therefore not be separated from the characters in a tale. To our knowledge, the representation of functions as classes and separate object properties as followed in this project is a novel approach.

To demonstrate how a thoroughly modelled ontology in combination with natural language processing approaches can be employed to semi-automatically populate the ontology, an information extraction component for folktale characters and Proppian functions has been added. This module, as described in Section 8, should be seen as a proof-of-concept study rather than a perfect tool for extracting information from folktale texts. The implementation of the ontology-guided information extraction is currently not accessible on the project website.

Instead of using the information extraction tool, manual annotation of folktale texts is also possible to populate the ontology with additional folktales.

6.1. Competency Questions

For the design of the ontology, following Noy and McGuinness' recommendations, a set of competency questions was formulated [31]. If these questions can be answered by the final ontology, it has fulfilled its expressive purpose. They should be seen as a minimal requirement to the expressitivity of the system.

1. Which folktales fall into a given motif class, e.g. ATU 70-99 Other Wild Animals?
2. Which Dramatis Personae appear in a given tale?
3. Which Proppian functions appear in African folktales?
4. How are Dramatis Personae interacting in the African folktales, e.g., which figures use the "interdiction" function?
5. Which sequences of Proppian functions appear in a given tale? Which sequences appear in tales in general?
6. Which Proppian functions follow a given function predominantly, i.e., are there patterns within the Proppian sequences?
7. Who is the editor of an anthology of folktales from a given origin?
8. How are Proppian functions verbalised, i.e., which words are used to describe events that fall into a given function class?
9. Is there a dominating interaction between certain classes of Dramatis Personae?

6.2. Axioms

Following Noy and MacGuiness' design pipeline [31] further, a set of axioms was defined before the implementation of the ontology. Some of these axioms refer to the publication of the tale and its metadata, e.g.:

- Each tale is published in an anthology, or as part of a journal article.
- Each anthology has at least one editor, a title, a publisher, and a date of publication.
- Each tale has a title.
- A tale can have an author and an origin if known.
- Each tale falls into one of the ATU type classes.
- Each ATU class has an ATU number and a description.

Furthermore, content-related axioms include:

- Each tale has a set of Dramatis Personae.

- Each fictional character belongs to one or more character classes and is represented by one or more verbalisations.[9]
- If a Proppian function applies to a tale, there is some verbalisation in the text.
- In a tale, Proppian functions always follow a specific order (see below), which is represented by a sequence.
- Each Proppian function is represented by a symbol.

In addition to these axioms, following Propp's approach, axioms for the description of the narrative were derived. These restrictions mainly model the scope of Proppian functions, e.g., the *Wedding* function can only be applied if it describes a relation between the *Hero* and the *Princess*. These restrictions were modelled using rdfs:range restrictions, e.g.

```
:hasBride  rdfs:domain  :Wedding;
           rdfs:range   :Princess,
:hasGroom  rdfs:domain  :Wedding;
           rdfs:range   :Hero.
```

If a function applies to a tale, the axiom holds. Not all of the functions need to occur in every tale, but all axioms, regarding which *Dramatis Personae* fulfill them, need to be fulfilled. Additionally, their order needs to remain the same. An exception to the sequential order can be made under special circumstances when a function is inverted [3, p.107].

6.3. Modelling Folktales in Description Logic

Modelling folktale narrative in Description Logic was particularly challenging, since certain real-life restrictions do not necessarily hold for the folktale domain. For instance, while in real life the classes of humans and animals (in the sense of non-human biological animals) would certainly be distinct, these classes might mix in folktales, e.g. transfiguration of humans into animals or a human mother giving birth to animals are recurrent pattern especially in African tales.

Especially with regard to future extensions of the ontology, it is crucial that the logical foundations are not preventing the annotation of unforeseen patterns in folktales. Therefore, only general description logic statements, such as those that are indicated by Propp's theory, have been defined in awareness that this approach might lead to a limited application of ontology reasoning in the future.

First, a set of description logic statements that model the class hierarchy was defined. They are divided between statements that are content-related, such as *Princess* $\sqsubseteq$ *DramatisPersonae*, and those that are metadata related, such as *Anthology* $\sqsubseteq$ *Publication*. Secondly, since ProppOntology is designed to model Propp's functions not only as classes but also as relations between folktale characters, a set of restrictions regarding the range and domain of Proppian functions were defined, e.g., Donor $\sqsubseteq \forall$ requestsService.Hero. Furthermore, Proppian functions are connected to the corresponding characters, e.g. *Departure* $\sqsubseteq \forall$ *hasDepartee.Hero*, as shown in Fig. 2.

Table 1 shows exemplarily how Proppian functions are modelled as relations between characters. In addition, functions and their subfunctions are represented as ontology classes. Since the function hierarchy follows directly from Propp's theory [3], the authors refrain from listing description logic statements on the class hierarchy for Proppian functions.

6.4. Implementation of Ontology Classes

Fig. 3a shows how the 31 function classes are implemented. They are divided into the five main categories *Preparation*, *Complication*, *Functions of the Donor*, *Struggle*, and *Dénouement*. Fig. 3b shows the main character classes, in particular the Proppian characters and the classes imported from [13]. The subclasses of *Animal* are far from complete, and can be extended where needed.

In contrast to the ontology by Declerck et al. [8], the classes modelling the Proppian functions and their subfunctions have been named after their original description as published in [3].

Furthermore, in ProppOntology the types of Dramatis Personae are modelled as subclasses and not as individuals of the Dramatis Personae class. This way, characters can be assigned appearing in a specific tale as individuals of character classes, such as *O_Obaraedo* as *Victim*.

Following Propp's naming conventions, the subfunctions are named following the same pattern as the parent function, e.g., δ1*_Interdiction_violated*. The ontology was designed in OWL using the Protégé desktop application [32].

Alternative labels consist of translations of the skos:prefLabels in different languages, such as German, Russian, and Bulgarian, that were either im-

[9] e.g., in the tale Snow White 'the stepmother' and 'the evil queen' describe the same individual.

Table 1

Selection of Important Concepts
∃ consistsOf.Move ⊑ Tale
∃ FollowedBy.ProppianFunction ⊑ ProppianFunction
∃ PrecededBy.ProppianFunction ⊑ ProppianFunction
∃ correspondsTo.ProppianFunction ⊑ ProppianFunction
ProppianFunction ⊑ ∃ applies.Tale
Tale ⊑ ∃ publishedIn.Publication
FictionalCharacter ⊑ ∃ appearsIn.Tale
FictionalCharacter ⊑ ¬ RealPerson
Hero ⊑ ∀ acquires.MagicalAgent
Hero ⊑ ∀ marries.Princess
Hero ⊑ ∀ combats.(Villain ⊔ Donor)
Hero ⊑ ∀ recognizedByMeansOf.Object
Donor ⊑ ∀ attemptsHarm.Hero
Donor ⊑ ∀ begsForFreedom.Hero
Donor ⊑ ∀ requests.(Task ⊔ Object)
Villain ⊑ ∀ causesHarm.FamilyMember
Villain ⊑ ¬ (Victim ⊔ Hero)
Villain ⊑ ∀ threatens.Victim
Reward ⊑ ∃ isRewardedTo.Hero
Reward ⊑ ¬ MagicalAgent
Task ⊑ ∃ proposedBy.FictionalCharacter
Task ⊑ ∃ proposedTo.Hero
Wedding (W) ⊑ ∀ hasGroom.Hero
Wedding (W) ⊑ ∀ hasBride.Princess
Departure ⊑ ∀ hasDepartee.Hero
Victory ⊑ ∀ hasDefeater.Hero
Victory ⊑ ∀ hasDefeated.Villain
Villain ⊑ ∀ taskesPosessionFrom.Victim
Donor tests the Hero (D1) ⊑ ∀ hasTester.Donor
Donor tests the Hero (D1) ⊑ ∀ hasTested.Hero
Donor tests the Hero (D1) ⊑ ∀ hasTask.Task

ported from the Family Ontology [13] or [8], provided by native speakers of isiZulu for the possible application of the system for African tales in their native languages, or created by ourselves. Some English skos:altLabels have been derived from WordNet synsets via the NLTK WordNet interface[10], in order to increase the number of matches between the folktale text and the skos:prefLabels for the information extraction.

Example specifications for function classes and character classes are given in Listings 1 and 2, an illustration of a Proppian function instance is shown in Fig. 4.

Some classes appear in pairs, such as the *A Lack* function and *K Liquidation of Lack*. They can be combined using the *correspondsTo* relation.

6.5. Object and Data Properties and their Constraints

As mentioned before, Proppian functions are modelled as classes, capturing their appearances in tales as individuals. To be able to examine the interaction between folktale characters, they are also represented by object properties.

Propp [3] defines strictly which character has to perform a certain action in order for a function to apply. For instance, the *Hero* can be only interrogated by the donor, which implies the function *D2 Donor greets and interrogates the Hero*. If another person e.g., the villain interrogates the hero, in order to find out more about him or her, the function *Reconnaissance* ϵ1 applies.

Data properties mainly provide metadata information, such as the tale title or the key used for distinguishing the individuals. A few data properties come with the Family Ontology, such as *hasGender*.

6.6. Folktale Annotation

Note that the use of the term "annotation" in this section follows the linguistic definition, i.e. the analysis of tales, not the sense of owl:AnnotationProperties.

Five different student annotators were asked to provide Proppian analyses for different folktales. Annotators were first introduced to the theory, before they annotated Dramatis Personae and their respective Proppian roles, as well as Proppian functions as they appear in the tales. Each character and function instance was annotated as an individual in the ontology. They are identifiable by a key that indicates to which folktale they belong, e.g., individuals starting with *COW* belong to the tale *Children of Wax*. Despite a function instance always being connected to a tale by an *applies* relation, and respectively a character by a *appearsIn* relation, this naming convention can be used for filtering query results later on and helps keeping the list of individuals comprehensible.

Each function or character of a tale comes with a verbalisation, i.e. their representation in the text ("the Yakshi", "the witch", "she", etc.) In the case of annotations of the same tale in more than one language, ver-

[10] http://www.nltk.org/howto/wordnet.html

Listing 1: Example specification of a Proppian function

```
###  https://teaching.gcdh.de/ProppOntology/1.0.1#XVIII_I_Victory
ProppOntology:XVIII_I_Victory rdf:type owl:Class ;
                              rdfs:subClassOf ProppOntology:IV_Struggle ;
                                    [ rdf:type owl:Restriction ;
                                          owl:onProperty :hasDefeated ;
                                          owl:someValuesFrom :Villain
                                          ] ,
                                          [ rdf:type owl:Restriction ;
                                          owl:onProperty :hasDefeater ;
                                          owl:someValuesFrom :Hero
                                          ] ;
                              rdfs:comment "Der Gegenspieler wird besiegt."@de ,
                                           "The villain is defeated."@en ,
                              skos:altLabel "Sieg"@de ,
                                           "ukunqoba"@zu ,
                              skos:prefLabel "Victory"@en ;
                              skos-xl:literalForm "I" .
```

Listing 2: Example specification of a Character class

```
###  https://teaching.gcdh.de/ProppOntology/1.0.1#Villain
ProppOntology:Villain rdf:type owl:Class ;
                      rdfs:subClassOf ProppOntology:Dramatis_Personae ;
                      dc:source "Vladimir Propp: Morphology of the Folktale. Austin, Texas 1968"@en ;
                      skos:altLabel "Schurke, Boesewicht"@de ;
                      skos:prefLabel "Villain"@en .
```

balisations in both languages are provided, e.g., in English and Malayalam. This feature allows interesting insights into the cultural transfer that folktales undergo during the translation process.

Furthermore, annotators were asked to provide metadata of the tale, such as the title or the publication it was published in.

Each tale was annotated by one annotator. As illustrated in the discussion about the Obaraedo tale in Section 3.2, and in experiments by Bod et al. [19], it is very unlikely that two annotators produce the same analysis. Moreover, the ontology is accessible through the institutional Webprotégé server, which creates an environment that allows users to discuss different Proppian analyses, e.g. using the Webprotégé comment functionality, and foster scholarly communication within the discipline of Folkloristics. However, ProppOntology does not aim to provide a ground truth in the sense of indisputable Proppian annotations.

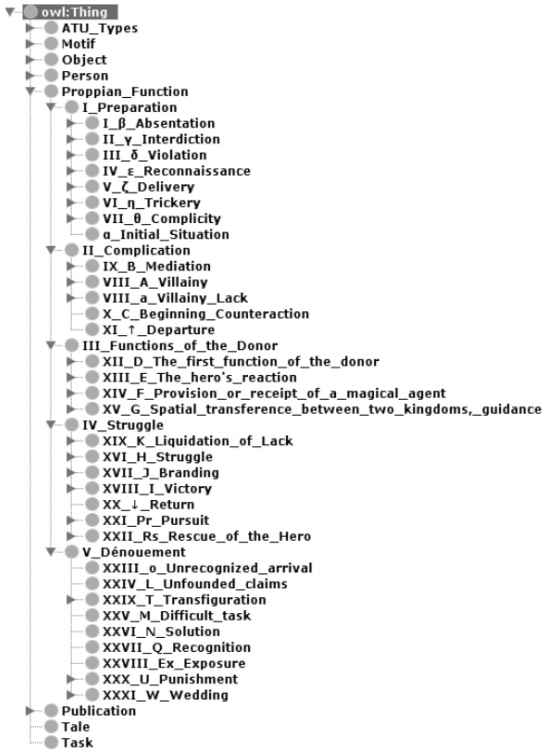
Note that every annotation of a tale is represented by an own RDF subgraph that is only connected to the tale, and via rdf:type edges to the classes of ProppOntology, and by the verbalisations to the tale text, but not to any other annotation. Especially, every *Dramatis Personae* p of a tale t detected in an annotation a becomes a separate RDF node $uri(t, a, p)$ – thus, different annotators could make different assignments of the same individuals named in a tale, as described in Section 3.2 for different annotations of the *Obaraedo* tale. This is not only useful for the current cases of manual annotation, but can also help much to generate automated annotations, or in an interactive process.

6.7. Usage of Ontology Reasoning

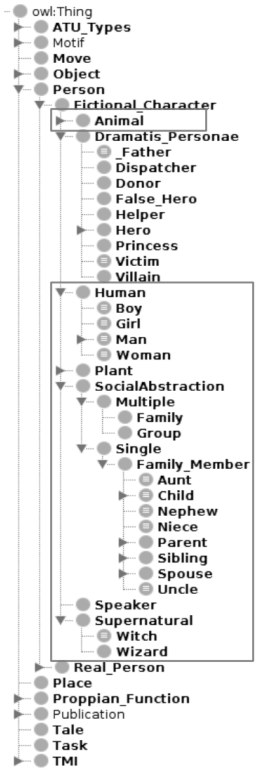
As described in Section 4, the ontology is in SHIF and might be extended with $\mathcal{SHOIQ}(\mathrm{D})$, which is supported by e.g. Protégé.

The ontology reasoning is usually not actually used to *derive* new knowledge which would be interesting to the users (the only derived information could come from assertions like rdfs:domain/range, e.g., that a defeated *Dramatis Personae* belongs to class *Villain* – a fact that a annotator is (or should be) aware of). Instead, the DL framework is merely used as a formal framework that allows for a logical axiomatization with correctness guarantees, and for consistency checks.

Furthermore, as described in Section 3.2 for existing annotations of the *Obaraedo* tale, sometimes these annotations are actually inconsistent usage of Propp's approach. One could conclude that these annotators



(a) Class Hierarchy of Proppian Functions



(b) Class Hierarchy of the Dramatis Personae, boxes indicate classes imported from the Family Ontology [13]

Figure 3. Subclasses Function and Dramatis Personae

should have used such a logic-based validation system (giving evidence that such a system is useful). In reality, these annotations exist, and might also contribute to research. So, keeping the reasoner turned off they are stored in the underlying database, seen as a pure RDF graph.

On the other hand, considering any (new or existing) annotation of a tale, this subgraph, can be considered (i.e., extracted from the RDF database) separately, to validate it together with the ProppOntology specification using the reasoner.

7. Ontology Implementation

This section presents the general layout and implementation aspects of the lightweight query system.

The core of the system is a Flask[11]-based web application which provides three major functionalities: queries, annotation, and ontology browsing. While most modern web applications are developed using programming languages like PHP or Ruby, Python was used in the context of this project because of the extensive availability of libraries and toolkits especially for the information extraction. This way, the system was developed in one language, avoiding the need to exchange data back and forth between different applications written in different programming languages.

The Flask application builds the Web pages from HTML templates, and communicates with a Fuseki Web server[12] via a RESTful API. The Fuseki server processes the SPARQL queries and sends the results back to the Flask application.

For the ontology processing, an Apache Jena Fuseki server application is used. It provides a comfortable handling of SPARQL updates and queries via a RESTful API. For development purposes, the employment of Fuseki came with the advantage that its interface could be used to check whether the ontology-driven information system that was developed behaves as desired, especially for the verification of the queries.

For the production system, the Fuseki server is hosted on a port that is only accessible from the server on which the Flask application is deployed. This ensures that no requests, especially no SPARQL updates, are sent to the RESTful API except those that come from the Flask application. This way, the risk of harmful injections into the ontology is reduced.

[11] http://flask.pocoo.org/
[12] https://jena.apache.org/documentation/fuseki2/

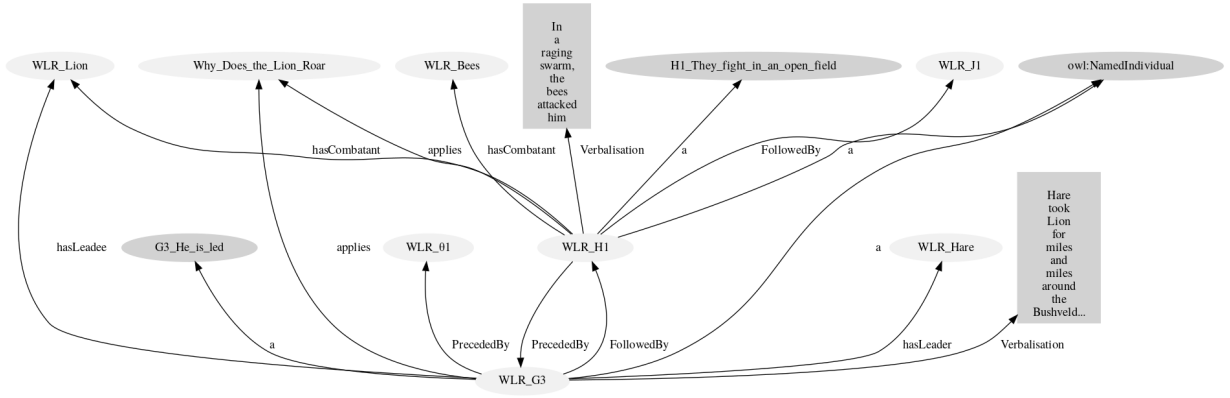


Figure 4. Graph representation of a Proppian function (dark grey ellipses indicate classes, light grey represents individuals, data property values are indicated by boxes)

Webprotégé is used for the ontology browsing and annotation part of the system [33]. While querying in itself already provides a lot of insight, especially a good overview of individuals that were added, users might want to see how classes and subclasses are defined. The Webprotégé instance is not directly connected to the Flask application. A MongoDB database is used for managing Webprotégé user accounts.

7.1. Connectivity of the ProppOntology to other Knowledge Bases

The system itself is a full-fledged computer-supported cooperative work-style tool for annotating tales according to Propp's functions. Its users in general log in with personalized accounts and use the Webprotégé user interface where they can add data and also edit/extend the ontology. The additions to the ontology are exported regularly, inspected for consistency and made available via the Fuseki-Server.

Since the data itself is stored as RDF data, it is also possible to use an RDF level API, e.g., to add bulk bibliographical data permanently to the system's knowledge.

The system's knowledge can also be exported as RDF data, either as a file, or as Linked Open Data, and it provides a SPARQL interface where the ontology can be queried using the corresponding SPARQL endpoint[13]. Both variants make it interoperable with remote data (similar annotations) that use the same ontology, or with different information that uses the same URIs for the tales (or, connected by owl:sameAs statements). Listing 3 illustrates how the ontology can be connected to additional knowledge bases, such as the Mondial ontology [34] for geographical information.

Listing 3: Example query connecting ProppOntology to other knowledge bases

```
PREFIX mondial:<http://www.semwebtech.org/mondial/10/meta#>
PREFIX ProppOntology:<https://teaching.gcdh.de/ProppOntology/
1.0.1#>

SELECT ?t ?p ?d
WHERE {
    ?t a ProppOntology:Tale .
    ?t ProppOntology:Origin ?p .

    SERVICE <http://www.semwebtech.org/mondial/10/> {
    FILTER (?o = mondial:Country || ?o = mondial:Province ) .
    ?s a ?o .
    ?s mondial:name ?p .
    ?s mondial:population ?d
  }
}
```

7.2. Queries

The users of the light-weight query system can query the ontology in three ways. Firstly, a basic text field can be used for advanced queries, triple queries can be used to investigate relations between rdf triples, and single queries provide means to investigate single classes. Additionally, access to the institutional Fuseki server is provided.

The user can provide a syntactically correct SPARQL query, including prefixes, interpunctation, query limits or regex restrictions in the text field. However, for this purpose users are advised to query directly via the institutional Fuseki server as it provides a more robust and comfortable query environment including syntax highlighting.[14]

[13] https://teaching.gcdh.de/sparql/ds/sparql

[14] https://teaching.gcdh.de/sparql

The second way of querying the ontology is provided by a simple user interface. Users can fill a triple query pattern and enter either one or two classes, leaving the ones empty that would be represented by the variables in a SPARQL query. The first and third field are dedicated to classes, while the second field is assigned to the relation. When the query page is loaded, relations, ranges and domains are queried from the ontology to create a dropdown menu for each of the fields. A star at the end of a class name is used as a flag to query not only the class itself but also its subclasses. If the checkbox next to the first or the last of the fields is ticked, the query yields individuals of the respective classes. Thirdly, single classes and instances of classes can be accessed through a single text field.

All query results from either of the three ways to query the ontology, i.e. triple query, single query or input field based query, can be exported as a CSV file.

8. Excursus: Information Extraction from Tale Texts

This project attempted to extract some of the information encoded in the text semi-automatically. Specifically, nominal phrases that describe characters or animals, and instances of Proppian functions were of interest. On the other hand, nominal phrases of non-living objects that are repeated through the text can indicate a motif, such as the tree that *Cinderella* repeatedly visits which supplies her with the ball gown [1] corresponds to the TMI motif *D950 Magic Tree*. As of yet, the project focussed on the extraction of characters and instances of Proppian functions and leave the motif extraction efforts for a future project.

While Wimalasuriya and Dou argued that linguistic extraction rules should be part of the ontology [35], the natural language processing elements were implemented entirely on the Flask side of the application. With a rule based approach, e.g., using regular expressions or gazetteer lists, it would make sense to include it within the ontology. However, this project followed a machine learning approach that used the Python module NeuralCoref [15].

8.1. Entity Recognition for Folktale Characters

Initially, a set of syntactic rules were defined to extract potential candidates of *Dramatis Personae* from the text. However, this approach did not yield satisfying results. The main reason might be that the rules for the appearence of characters in tales must naturally be relatively broad.

A rule like: $NP : < DT >? < JJ > * < NN >$, would deliver correct nominal phrases, such as *the girl*, but also yield many false positives, since not every grammatically correct nominal phrase indicates a folktale character. Stricter, more sophisticated rules would likely not find entities that are verbalised in a simple manner, like *the man*.

The NLTK toolkit for Python provides a named entity chunker *ne_chunk*[16]. Expectedly, fairy tale texts do seldomly supply *named* entities, with exception of some popular tales like *Hans in Luck* or the *Obaraedo* tale discussed before. Usually, characters are introduced in a more general way, e.g., *the girl*. Therefore, the pure named entity recognition task was abandoned.

Since verbalisation of characters is one of the interesting features the ontology is supposed to supply, the focus shifted to the resolution of coreferences instead. The main idea behind using coreferences was that entities or other important features will likely be repeated throughout the text. One hypothesis is that instances of *Dramatis Personae* yield particularly long coreference chains since they are key elements in folktale plots.

A satisfyingly working coreference resolution tool would not only provide characters that occur in the text, it would also provide reoccuring motifs, e.g., a tale revolving around an apple tree would yield many coreferences for *apple tree* or *tree*. Using a coreference approach yields results for named entities as well as unnamed entities, which is the most significant advantage and the main reason this approach was chosen.

From the available coreference resolution approaches, the NeuralCoref[17] approach was found to be the most promising. Although NeuralCoref was initially designed for coreference resolution in chatbot systems[18], this approach seems to work reasonably well on English folktale texts.

The text is first preprocessed using Spacy's *nlp* method[19]. Subsequently, coreferences are resolved using NeuralCoref. Candidate entities of characters are identified from the text using Spacy's named entity recognition method *ents*, finding token sequences from

[15] https://huggingface.co/coref/

[16] https://www.nltk.org/api/nltk.chunk.html

[17] https://huggingface.co/coref/

[18] https://medium.com/huggingface/state-of-the-art-neural-coreference-resolution-for-chatbots-3302365dcf30

[19] https://spacy.io/api/doc

the entire text that are labelled as *Person*. Since this list alone yields very noisy results, the candidate tokens are then compared to the antecedents in the coreference clusters. Candidates that do not appear in the coreference clusters are abandoned.

8.2. Extraction of Instances of Proppian Functions

For extracting occurrences of Proppian Functions, the extensive SKOS labels provided by the ontology were employed. For the time being, function instances are extracted from English tale texts, therefore only skos:prefLabel fields are used. However, skos:altLabels could be used to identify instances for classes in different languages in the future.

For the information extraction, the text is preprocessed as described above. A SPARQL query yielding the values of all pref:Labels and their corresponding classes is sent to the Fuseki server at the beginning of the text processing.

After the coreferences are identified, a list of first mentions in all the coreference chains is created. Each mention is tokenized and stripped of punctuation. A list of tokenized prefLabels is created. Both lists are then lemmatized using the NLTK WordNetLemmatizer and compared. If one antecedent matches a token in a prefLabel, it is added to the list of potential candidates for that particular class.

The results of both approaches are then handed back to the Flask application, which creates an input form. If a potential person is found, a dropdown list allows the user to select the correct ontology class. If a candidate class is found by the second approach, the class name is shown next to the input field. Users can then change the data and create their own annotation.

9. Results and Evaluation

This section reports the quantitative results of the application of the Proppian annotations of African and Indian folktales that can be gathered by querying the ontology. To date, the corpus of annotated tales includes 20 (mostly sub-Saharan) African tales and 15 tales from the region of Kerala in southern India.

This evaluation investigates the annotations with respect to the structure of the tales (Section 9.1), patterns of Proppian functions (Section 9.2), and see how characters are represented in culturally different tales (Section 9.3). It should also be borne in mind that the corpora investigated in this paper are relatively small and results can therefore only be indicative of potential tendencies that would have to be verified on a larger corpus.

The results presented here are potentially interesting for folklorists who want to compare Proppian analyses of African and Indian tales. Furthermore, existing theories about those tales, e.g. [14], [36], can be investigated and supported.

To evaluate the ontology, natural language questions are phrased as SPARQL queries to the lightweight query system.

Extensions to the functionality of the front end are planned for the future, e.g., by adding visualizations for the data that is currently only displayed in a list.

9.1. Narrative Structure of Tales

The following section investigates how the structure of tales differs throughout the small corpus. Propp divided the 31 functions into five categories, *Preparation*, *Complication*, *Functions of the Donor*, *Struggle*, and *Dénouement*. The annotated tales were analysed to determine how prevalent these five categories are. Fig. 5 shows the mean percentage of each of the categories among function sequences from African and Indian tales.

The data shows that African tales focus more strongly on the preparatory functions, e.g., the description of the initial situation. Indian tales, on the other hand, stress the complicating functions more, e.g., the acts of villainy or the beginning counteraction. While 40 % of the mean function sequence length in African tales consists of preparative functions ($\alpha-\theta$), and 27 % consist of functions from the *Struggle* category, only 5 % of the sequence length is made up of function from the *Dénouement* category. In Table 2, only four of the eleven functions fall into that category. The rest of the final functions belong to the *Struggle* category, with the exception of *Provision of Magical Agent F* (*Functions of the Donor*).

The data shows a different behaviour in Indian tales, which focus more on the *Complication* aspects of story telling, i. e. the acts of *Villainy* or the *Departure* of the Hero. The complication functions make up 31 % of the mean function sequence length. Neither story telling cultures seem to make extensive use the *Functions of the Donor*.

Tables 2 and 3 illustrate the different tale beginnings and endings for African and Indian tales. The African tales that were investigated showed a clear preference for the Proppian functions *Interdiction* γ, *Absentation*

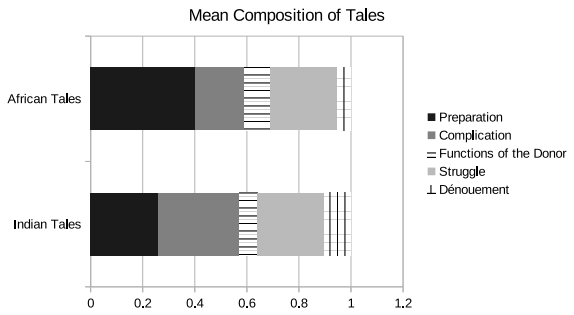


Figure 5. Composition of African and Indian tales by function classes

Table 2

Distribution of introductory and concluding functions in African tales (bold functions belong to the *Dénouement* category)

Beginning	End
Interdiction γ: 8	*Return* $\downarrow$: 4
Absentation β: 8	*Liquidation of Lack K*: 3
Trickery η: 4	***Transfiguration T***: 3
	Wedding W : 3
	Pursuit P: 1
	Provision of Magical Agent F: 1
	Unrecognized Arrival o: 1
	Rescue Rs: 1
	Punishment U: 1
	Victory I: 1
	Branding J: 1

Table 3

Distribution of introductory and concluding functions in Indian tales (bold functions belong to the *Dénouement* category)

Beginning	End
Absentation β: 9	*Return* $\downarrow$: 5
Villainy/Lack A: 4	***Transfiguration T***: 4
Reconnaissance ϵ: 1	***Solution N***: 2
Complicity θ: 1	***Punishment U***: 2
	Recognition Q: 1
	Liquidation of Lack K: 1

β and *Trickery* η at the beginning of the tales. This indicates, that Propp's preparatory functions are well suited for representing African tales beginnings. The only exception is Azuonye's analysis of the *Obaraedo* tale [15], where the initial function is *initial situation* α[20] and *Absentation* β is the second function. Also note that in Okodo's analysis [18] of this tale the initial function is indeed *Absentation* β as discussed in Section 3.2.

Interestingly, the distribution of ending functions reported in Tables 2 and 3 might allow some new interpretations. While the functions belonging to the *Dénouement* class symbolise some sort of reward for the hero's struggles, only four out of eleven different ending functions in African tales belong into that category, corresponding to seven out of 20 tales.

This could indicate that the reward for heroes in African tales is not to gain something, e.g. a throne, the princess, monetary reward, or fame, as described in Propp's *Dénouement* functions. Instead, the "reward" seems to be to restore the status from the beginning of a tale, e.g. returning home, liquidation of lack brought onto the hero by the villain, or victory over some form of evil. These end functions indicate a lack of individual reward, e.g. monetary, in African tales which is in line with previous analyses for African tales as discussed in Section 2. [14]

This particularity should further be investigated as the population of the ontology grows.

It might be worth studying the function sequence endings in greater detail. Folklorists might come to the conclusion that an alternative to *Dénouement* with an new set of functions might be worth defining for African tales.

The Indian tales show a slightly different division of initial and concluding functions, as shown in Table 3. The non-preparatory function *Villainy/Lack A* appears four times as an initial function, if the *Initial Situation* α is ignored, which appears 14 times in total. The other start functions fall into the *Preparation* category.

The tale endings *Return* $\downarrow$ and *Liquidation of Lack K* fall into the *Struggle* category. The remaining eight ending functions belong into the category of *Dénouement*. Indian tales use less diverse tale endings than the African tales, and Propp's *Dénouement* category seems to be better suited. Nine out of 15 tales end in *Dénouement* functions.

9.2. Patterns of Functions

Spatial distance seems to play a certain role in all tales. The functions *Departure* $\uparrow$ and *Return* $\downarrow$ appear alone or together in eight of 20 African tales. In Indian tales, they appear 13 times, counting occurrences in multi-move tales separately. In four African tales and nine Indian tales, both *Departure* $\uparrow$ and *Return* $\downarrow$ can be found as a pair.

[20] Propp himself states that the initial situation function is not technically a function. [3] Therefore, in Tables 2 and 3 the beginning of the tale starts with the first plot-driving function.

The *Departure* ↑ function appears without a corresponding *Return* ↓ once in African tales and four times in Indian tales, while *Return* ↓ appears on its own three times in African tales.

The appearance of the functions *Return* ↓ and *Departure* ↑ on their own could be an indicator towards the prevalence of transformation patterns, in this case spatial transformation, which Harold Scheub found "reveal[ing] the way people of the region survived the onslaught of colonialism." [36, p.20] He argues that oral story telling serves as a form of resistance in which metaphors help listeners to identify with characters. This might also explain why African tales give more room for preparatory functions, as shown in Fig. 5, e.g. to create a setting that recipients can recognize.

Another prominent pattern is the pair *Villainy A/Villainy Lack a* and the corresponding function *Liquidation of Lack K* and their subfunctions. The pair appears together in nine African tales and ten Indian tales resp. moves. The distance between *Villainy A/Villainy Lack a* and *Liquidation of Lack K* ranges between one function and seven functions in African tales, and three to six functions in the Indian corpus.

Additionally, *Villainy A/Villainy Lack a* appears alone in five sequences of African tales and seven times in the Indian corpus. This indicates that in 25 % of the analysed African tales and 47 % of the Indian tales, some form of harm is done to the hero or his/her family members without being resolved later.

In line with Propp's theory, there is no occurence of *Liquidation of Lack* without a preceding *Villainy A/Villainy Lack a* in African or Indian tales.

9.3. Representation of Characters

Characters in the annotated African tales mainly belong to three upper classes, *Animal* (23), *Family Member* (30) and *Dramatis Personae* (48). Of course, one character can belong to multiple of those upper classes. Fig. 6 shows the distribution of Proppian characters in the corpus. For the tales from India, the most prevalent character classes are *Human* (44) and *Dramatis Personae* (57). The Dramatis Personae fall into seven categories as defined by [3]: *Hero/Victim, Villain, Helper, Donor, Dispatcher, Princess and her father, and False Hero*.

The classes *Hero* and *Villain* appear 16 and 17 times in the corpus of twenty African tales. Five instances of *Victim*, three instances of *Donor*, and two instances

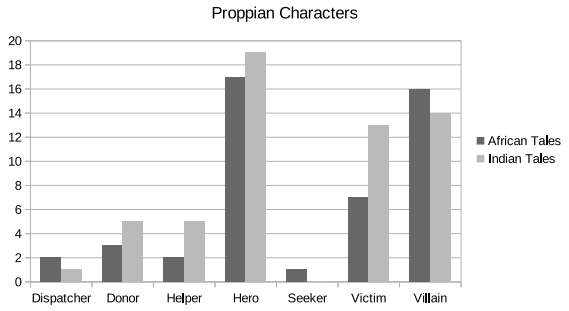


Figure 6. Distribution of Dramatis Personae class instances

of *Helper* occur; *Seeker*[21] and *Dispatcher* both appear exactly once. There was no instance of either *Princess* or *Princess' Father*.

In the collection of fifteen Indian tales, the most common Proppian characters are *Hero* (19), *Villain* (14), and *Victim* (13). In addition, the classes *Donor* and *Helper* appear five times each, and there is one occurence of the *Dispatcher* class.

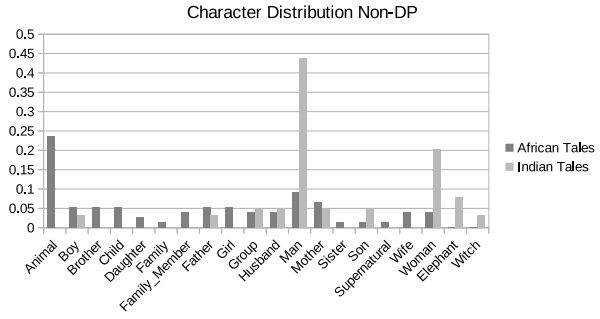


Figure 7. Distribution of non-Dramatis Personae class instances

Since the Family Ontology [13] was imported to gain more insights into how Proppian characters are represented in tales, characters can belong to more than one character class. For instance, if the victim in a tale is the father of the hero, his character might fall into the classes victim, man, and father, and husband if the hero's mother appears in the tale as well.

Fig. 7 shows character classes in both Indian and African tales that do not belong into the group of Proppian Dramatis Personae. The African tale data shows a preference for animal characters, and agents are more diverse than in Indian tales. Especially family relations

[21] The role of the seeker is a specification of Hero and could therefore also be counted into the Hero class.

seem to play a more significant role. On the other hand, Indian tales show a strong preference towards male characters.

Since one character can belong to multiple classes, users can investigate the distribution of Proppian roles among other classes. Fig. 8 shows the distribution of the hero class among other character classes. The data shows similar preferences towards animal resp. male character classes as in Fig. 7.

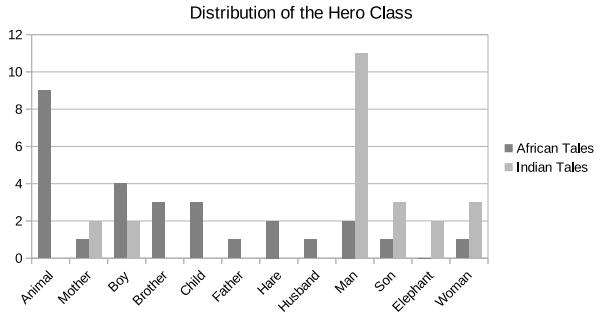


Figure 8. Distribution of the Hero class among other character classes (multiple occurrences possible)

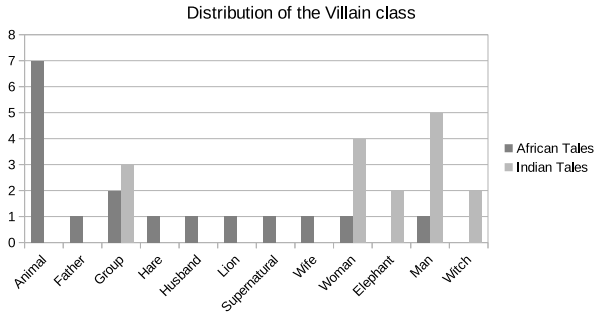


Figure 9. Distribution of the Villain class among other character classes (multiple occurrences possible)

Fig. 9 shows the distribution of the villain class among other character classes. Interestingly, while the African tales follow the same pattern as before, i.e., the *Villain* mainly belonging to animal classes, the Indian tales show almost the same number of female and male villains.

Regarding the representation of agents in the corpus of tales that were annotated, it is apparent that the Dramatis Personae mainly consist of the Proppian roles *Hero* and *Villain* in African tales, and *Hero*, *Villain* and *Victim* in Indian tales.

Especially the lack of *Donor* figures in the annotations seems to indicate that this role is a specific feature in Russian magic tales, for which Propp's theory was initially developed.

As expected, animal characters play a dominant role in African tales. Noticably, they mainly seem to fulfill roles of *Villain*, see Fig. 9 and *Hero*, see Fig. 8. This could be an indicator that a clear separation of characters into good and evil is characteristic for animal tales. In Indian tales, heroes are predominantely male figures, see Fig. 8. While the corpus studied for this project is very limited and by no means representative, the representation of characters might be a relict of patriarchical structures in early Indian society where stories originated. [37]

9.4. Literary and Social Interpretations

In general, comparative ontologies like the Propp-Ontology have the potential to reveal the universal nature of powerful ideologies and traditional stereotypes. The data indicated a strong bias towards male characters in Indian tales especially in the *Hero* class. Ortner quoted in Tuğlu [38, p.18] indicates that "universality of female subordination, the fact that it exists within every type of social and economic arrangement" is "something very profound, very stubborn". In this case, glimpses of patriarchy may be seen in the character distribution of the male and female figures. In Indian tales, the male figures have the highest distribution. Patriarchy creates hierarchical binaries across genders which manifest in the narrative in particular ways [39, p.161]. The male character is most often coded as the rational, prime mover acting in a range of capacities and roles, while the female is confined to the stereotype governed by perceived biological imperative, usually represented in the role of the mother [40, p.6]. Moutsou further argues that the female in narrative structure is cast within the 'Madonna –Whore axis'. The 'Madonna' (or mother figure) is passive and subordinate, and hence not a plot-driving character. The 'Whore' (or witch figure) is active, independent and uncontrollable [41, p.184]. The 'mother' figure reinforces female connection to biology as the key marker of identity, and is usually self-effacing, keen to obey [41, p.185]. She is present yet either not heard or serves only as a frame for action, e.g. by giving an interdiction at the beginning of a tale, never or rarely the pole position of prime mover [38, 15]. This basic analysis of gender indicates the range of work possible by modelling ontologies to represent ideological and

traditional stereotypes in folktales. Future modelling could extend Proppian functions to include voice and further delineations of gender.

10. Limitations & Future Work

10.1. Fulltexts

By design, no fulltexts are stored in the context of the ontology. This should not be interpreted as a limitation of the usability of the system, as the verbalisations that are provided are sufficient proof for the Proppian analyses. However, first time users might expect to be able to access the entire tale and not only the verbalisations stored when annotating functions and characters. This could be potentially achieved by storing the fulltexts as annotated XML-TEI[22] files and referencing the verbalisations by using pointers to the specific parts of the document. For this, the Web Annotation Data Model [42] (which follows the concept of XLink third-party links [43]), can be employed by using annotation instances that point to the fulltext fragment (and/or also video/audio) on their target side, and to the URI of the Propp function instance on their body side. An alternative would be to incorporate textual representations of characters and Proppian functions by using TEI as Linked Open Vocabulary, as proposed by Ciotti and Tomasi [44].

However, copyright aspects need to be taken into consideration when following this fulltext approach.

10.2. Natural Language Questions

Efforts have been made to generate the SPARQL queries answering the competency questions automatically. However, a natural-language-to-SPARQL-system would either have to rely on an extensive rule system or needs to be trained on a large set of questions and corresponding queries if a machine learning approach is used. Unfortunately, the implementation of this feature exceeds the scope of this project. However, for the system at hand such a feature would certainly be useful, especially since it would allow users with lower levels of IT-profiency to use it in a more intuitive manner. Attempts in this direction have been made by the ORAKEL project [45], or [46].

[22] Text Encoding Initiative https://tei-c.org/

10.3. Future Work

As the ontology grows, potentially also linking additional media types such as video and voice recordings, one might consider taking into account additional features, such as features like facial expressions, reactions of the audience, interaction between narrator and audience, degree of attention, and composition of the audience "from the standpoint of age, sex, class or other social division" [17] should be added as datatype properties.

Furthermore, it is planned to add the possibility to visualize findings, e.g., by showing origins of tales on a map.

Measuring occurrence of function pairs and their distance, as discussed in Section 5, could be automated with relatively low effort. This feature would certainly become more interesting as the ontology grows.

As the ontology can be extended by folklorists with different cultural foci, we hope to create a larger foundation for the intercultural comparison of folktales. The more folktales are provided, the more possible applications the tool could yield. If ProppOntology could host a substantial number of instances per function with their respective verbalisations, this data could be used to train a machine learning system for automatic function suggestion, extending the information extraction functionalities discussed in Section 8. The more tales of different origins are added by specialists, the more thorough investigations can be made through queries. The authors aim to make this system available to folklorists around the world, in order to build a community-driven knowledge base on Proppian analyses.

11. Conclusion

This project aimed to show how ontologies can help formalise traditional theories from the Humanities.

In contrast to many successful ontology-related Digital Humanities projects, ProppOntology was not modelled on a vast amount of data. Instead, it was created in a bottom-up approach from a theory-oriented point of view, with a specific purpose – the comparison of Proppian analyses.

Vladimir Propp's theory on the *Morphology of the Folktale* [3] was modelled and used to demonstrate how data about folktales from different cultural backgrounds can be easily accessed and compared by translating traditional folkloristic questions about the struc-

ture of tales or the representation of characters into queries. A carefully modelled ontology cannot only serve as means to access data and put it into context, but it can also assist traditional Humanities researchers approaching research questions that are commonly solved by manual analysis and comparisons even today.

The system allows users to compare different analyses of the same tale, and therefore holds potential to spark scientific discourse, providing a platform for different interpretations of Proppian functions, e.g., in the case of the *Obaraedo* tale as discussed in the beginning.

Proppian analyses are used both for teaching and research. Unfortunately, many of these analyses could previously not be contextualised and compared, because a digital tool to collect annotations was still missing.

This paper presents an ontology that is accessible and invites folklorists to share their annotations on our Webprotégé instance. This way, ProppOntology serves as a tool for folklorists who are interested in contextualising their analyses in an intercultural environment. Furthermore, folklorists and linguists are invited to expand the set of translations for Proppian functions, Dramatis Personae, motifs and other concepts.

ProppOntology may be of interest for intercultural research on folktales, but also for translation studies because verbalisations of the same character or function can be provided together. While still work in progress, the lightweight query system allows users to access the data and to draw own conclusions about Proppian morphology and character representations in tales of different origins.

12. Acknowledgements

We would like to thank Thierry Declerck, Jean Vincent Fonou Dombeu, and Yasar Abbas for their help with the conceptualization of the project. We are grateful to Yuvika Singh and Danielle Russel for help with the annotation of the African folktales.

References

[1] J. Grimm and W. Grimm (eds), *Kinder- und Hausmärchen*, 7. edn, Verlag der Dieterichschen Buchhandlung, Göttingen, 1857.

[2] A.N. Afanasyev, *Russian Folk-Tales*, E. P. Dutton & Company, New York, 1916.

[3] V. Propp, *Morphology of the Folktale*, Vol. 10, University of Texas Press, 1968.

[4] A. McCall Smith (ed.), *Children of Wax: African Folk Tales*, Canongate, 1989.

[5] F. Pannach, An Ontology-Driven Information System based on Vladimir Propp's Morphology of the Folktale for Southern African Folktales, Master's thesis, University of Göttingen, Germany, 2019. http://www.dbis.informatik.uni-goettingen.de/Teaching/Theses/PDF/MSc-Pannach-Folktales-mrz-2019.pdf.

[6] F. Peinado, P. Gervás and B. Díaz-Agudo, A description logic ontology for fairy tale generation, in: *Procs. of the Workshop on Language Resources for Linguistic Creativity, LREC*, Vol. 4, 2004, pp. 56–61.

[7] B. Díaz-Agudo and P.A. González-Calero, CBROnto: a task/method ontology for CBR, *Proceedings of the 15th International Florida Artificial Intelligence Research Society Conference* **2** (2002), 101–106.

[8] T. Declerck, A. Aman, M. Banzer, D. Machácek, L. Schäfer and N. Skachkova, Multilingual Ontologies for the Representation and Processing of Folktales, in: *Proceedings of the First Workshop on Language technology for Digital Humanities in Central and (South-)Eastern Europe*, A. Dinu, P. Osenova and C. Vertan, eds, INCOMA Ltd, 2017, pp. 20–24.

[9] T. Declerck, A. Kostová and L. Schäfer, Towards a Linked Data Access to Folktales classified by Thompson's Motifs and Aarne-Thompson-Uther's Types, in: *Proceedings of Digital Humanities 2017*, ADHO, 2017.

[10] T. Declerck and L. Schäfer, Porting past Classification Schemes for Narratives to a Linked Data Framework, in: *Proceedings of the 2nd International Conference on Digital Access to Textual Cultural Heritage*, ACM, 2017, pp. 123–127.

[11] F. Karsdorp, M. van der Meulen, T. Meder and A. van den Bosch, MOMFER: A Search Engine of Thompson's Motif-Index of Folk Literature, *Folklore* **126**(1) (2015), 37–52. doi:10.1080/0015587X.2015.1006954.

[12] F. Karsdorp, Retelling Stories. A Computational-Evolutionary Perspective, PhD thesis, Radboud Universiteit Nijmegen, 2016.

[13] N. Koleva, *Ontology-based iterative detection of characters and their recognition in folktales*, Bachelor Thesis, Saarland University, 2011.

[14] U. Reuster-Jahn, Gute und schlechte Ausgänge in europäischen Märchen und in den Volkserzählungen der Mwera in Tansania, *Märchenspiegel* **13**(2) (2002), 17–18.

[15] C. Azuonye, Morphology of the Igbo Folktale: Ethnographic, Historiographic and Aesthetic Implications, *Folklore* **101**(1) (1990), 36–46.

[16] A. Scheidel, *An Augmented Annotation Scheme for Fairy Tales*, Bachelor Thesis, Saarland University, 2010.

[17] D.J. Crowley, The Uses of African Verbal Art, *Journal of the Folklore Institute* **6**(2/3) (1969), 118–132.

[18] I. Okodo, Obaraedo: Conformity to Proppian Morphology, *AFRREV IJAH: An International Journal of Arts and Humanities* **1**(2) (2012), 100–111.

[19] R. Bod, B. Fisseni, A. Kurji and B. Löwe, Objectivity and Reproducibility of Proppian Narrative Annotations, in: *Proceedings of the Third Workshop on Computational Models of Narrative (CMN'12)*, M.A. Finlayson, ed., 2012, pp. 15–19.

[20] H.-J. Uther, *The types of international folktales: a classification and bibliography, based on the system of Antti Aarne and Stith Thompson*, Suomalainen Tiedeakatemia, Academia Scientiarum Fennica, 2004.
[21] S. Thompson, *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folk Tales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jest-Books, and Local Legends*, Vol. 1, Indiana University Press, 2001.
[22] A. Miles and S. Bechhofer, SKOS simple knowledge organization system reference, *W3C recommendation* **18** (2009), W3C.
[23] S. Weibel, The Dublin Core: a simple content description model for electronic resources, *Bulletin of the American Society for Information Science and Technology* **24**(1) (1997), 9–11.
[24] H. Scheub (ed.), *African Tales*, Univ. of Wisconsin Press, 2005.
[25] N. Greaves (ed.), *When the Hippo Was Hairy*, Lutterworth Press, 1990.
[26] P. Savory, *Bantu folk tales from southern Africa*, Howard Timmins, 1974.
[27] K. Sankunni, *Aithihyamaala*, Kottayam Nasanal Bukk Sttal, 1974.
[28] S. Ramachandran, *Aithihyamaala*, Mathrubhumi Books, 2014.
[29] E.G. Nair, *Poothappattu*, Poorna Publications, 1955.
[30] Story of the Poem “Poothappattu” (Ode to a "Pootham"). https://www.edasseri.org/English/poothapattu_story.htm.
[31] N. Noy and D.L. McGuinness, Ontology development 101: A Guide to Creating Your First Ontology, *Knowledge Systems Laboratory, Stanford University* **KSL-01-05** (2001).
[32] M.A. Musen, The protégé project: a look back and a look forward, *AI Matters* **1**(4) (2015), 4–12. doi:10.1145/2757001.2757003.
[33] T. Tudorache, J. Vendetti and N.F. Noy, Web-Protege: A Lightweight OWL Ontology Editor for the Web., in: *OWLED*, 2008.
[34] W. May, Information Extraction and Integration with FLORID: The MONDIAL Case Study, Technical Report, 131, Universität Freiburg, Institut für Informatik, 1999, Available from http://dbis.informatik.uni-goettingen.de/Mondial.
[35] D.C. Wimalasuriya and D. Dou, Ontology-based information extraction: An introduction and a survey of current approaches, *Journal of Information Science* **36**(3) (2010), 306–323. doi:10.1177/0165551509360123.
[36] H. Scheub, *The uncoiling python: South African storytellers and resistance*, Ohio University Press, 2010.
[37] U. Chakravarti, Conceptualising brahmanical patriarchy in early India: Gender, caste, class and state, *Economic and Political Weekly* (1993), 579–585.
[38] B. Tuğlu, From Culturazing Nature to Naturalizing Culture: The Differing Function of Animal Imagery in Defining Bodies from Homer’s Odysseus to Margaret Atwood’s The Penelopiad, *European Journal of Social Science Education and Research,* **3**(2) (2016), 15–20.
[39] A. Kertzer, This quiet lady: Maternal voices and the picture book, *Children’s Literature Association Quarterly* **18**(4) (1993), 159–164.
[40] M.A. Ferguson, *Images of Women in Literature*, University of Massachusetts, 1986.
[41] C. Moutsou, The maternal and the erotic: An exploration of the links between maternal and erotic subjectivity, in: *The Mother in Psychoanalysis and Beyond*, R. Mayo, ed., Routledge, 2016, pp. 199–215.
[42] W3C, Web Annotation Data Model, 2017. https://www.w3.org/TR/annotation-model/.
[43] XML Linking Language (XLink) Version 1.1, 2006.
[44] F. Ciotti and F. Tomasi, Formal ontologies, linked data, and TEI semantics, *Journal of the Text Encoding Initiative* (2016).
[45] P. Cimiano, P. Haase, J. Heizmann, M. Mantel and R. Studer, Towards portable natural language interfaces to knowledge bases–The case of the ORAKEL system, *Data & Knowledge Engineering* **65**(2) (2008), 325–354.
[46] J.-D. Kim and K.B. Cohen, Natural language query processing for SPARQL generation: A prototype system for SNOMED CT, in: *Proceedings of BioLINK 2013*, 2013, pp. 32–38.