LDK 2023

Modeling and Comparison of Narrative Domain Knowledge with Shallow Ontologies

Franziska Pannach
Göttingen Centre for Digital Humanities

University of Göttingen
franziska.pannach@
uni-goettingen.de

Theresa Blaschke
Marburg Center for
Digital Culture and Infrastructure
Philipps-Universität Marburg
blaschkt@
students.uni-marburg.de

Abstract

Ancient myths have fascinated scholars and laymen for centuries. In comparatistic efforts, classical scholars try to detect and interpret variations between versions of the same myth. We present a way to structure the underlying background information in myth variants. The background knowledge of twelve different versions of the popular myth *Orpheus and Eurydice* has been modeled in individual shallow ontologies that allow inter- and intra-myth comparison.

1 Introduction

The story of *Orpheus and Eurydice* is one of the most popular Greek myths with a long tradition of re-use and re-adaptation. Each of the variants of the myth uses certain elements of the narrative while leaving others out. One aspect of comparing those variants is to investigate not only the plot, but also which assumptions we can make about the circumstances in which the myth takes place, i.e. the background knowledge about the world it takes place in. Projects like Wikidata or Mythoskop[1] combine information from different sources and give a good overview of how characters and concepts are connected. However, investigating the difference between narratives, especially contradictory information, is an interesting research objective in itself.

Consider the following example: In most variants, *Orpheus* manages to reach the netherworld and is allowed to take *Eurydice* with him. But once he turns around to look at her, he loses her forever. However, why do we still consider it a variant of the same myth if *Orpheus* reaches the surface without turning around and is hence successful in bringing *Eurydice* back from the dead [6, L.1-14], [4]? The reason is, that we know that both variants concern the same characters and the circumstances are overall the same (e.g. Eurydice is in the netherworld. Orpheus has nothing but his musical talent to convince the inhabitants of the netherworld to release her.) In other words, the background is the same or at the very least similar. Additionally, the similarity of ancient mythical plots has already been studied thoroughly, e.g. by Bowra [1] or Marlow [8].

Hence, we focus this work on the question "Who is who and what is what?" and not "What happens?"

Comparatistic efforts of mythological narratives are still conducted mainly manually. In this paper, we demonstrate how we approach the comparison of the background information in mythological (and other) narrative domains in a manner that results in re-usable, machine-readable domain ontologies, and how we can use them to compare variants of the same myth.

2 Related Work

Nakasone and Ishizukua [10] use Rhetorical Structure Theory (RST) as a basis for a generic ontology model that focusses on storytelling paradigms.

In his work on the narrative formalism of Vladimir Propp [14], Peinado et al. [13] uses ontologies for automatic fairy tale generation. Ciotti [2] uses character-centric domain ontologies and highlights their importance in the Digital Humanities and the field of digital narratology.

Most digital analyses of narration focus on texts. Xu et al. [16] propose a model that uses ontologies and human annotation to capture narration on digitized artifacts, such as vase paintings, and other cultural heritage objects.

Re-tellings of folktales, similar to myth variants in this work, have been studied by [7]. Their story networks represent ancestral relationships between folktale variants, such as "Little Red Riding Hood". However, they do not focus on the content of the tales.

For the mythical domain, the Mythoskop

[1] https://mythoskop.de/

project[2] presents a knowledge graph focussed on the relationship and geneaology of characters of the Greek mythology. The VAST (Values across space & time) project presents a semantic knowledge graph[3] of annotatations on "past of values", including *Peace* or *Justice*. Their sources include Greek tragedies, among others.

3 Data

For this project, we use twelve myth variants of the myth of *Orpheus and Eurydice* from various antique sources. A complete list of sources and their abbreviations can be found in the project repository. The variants span a considerable time, with the earliest source approx. 400 BCE (Plato, Symposium) and the latest 875–1075 CE (Mythographus Vaticanus). The data consist of a number of statements per myth variant that form one narrative sequence describing the plot. They have been derived by domain experts of classical studies according to the hylistic approach [18][17].

This approach was developed specifically to extract and analyze narrative structures from mythological sources. It has been applied to different temporal and geographical backgrounds, such as ancient Mesopotamia, ancient Greece, or Egypt [3].

The individual statements in each sequence are derived from the original text of the source, e.g. a Greek poem, but they are not re-tellings of the story nor direct quotes from a translation, as the examples in Section 4 illustrate. Each sequence of statements was extracted by one or more domain experts, and reviewed, discussed, and agreed-upon within the research group.

The sequences of statements that describe a myth variant include two coarse types of elements: 1. statements concerning the background or circumstantial knowledge (durative) and 2. narrative statements that form the plot (single-point). We can distinguish these types of sequence elements by their truth values over the sequence.

For instance, the statement "Eurydice is the wife of Orpheus" is *true* at all times during the narrative sequence, while "Orpheus turns around" is *true* only once, at one point in the sequence. Consequentially, "Eurydice is dead" is *true* after she was killed by a snake, so only over a part of the sequence.

To compare narrative domains, i.e. all circumstantial and background knowledge available from a source about a myth, which are the basis of a myth variants, we only consider statements that are *true* over the entire sequence. In hylistic terms [18], those are considered *durative-constant*. Statements that are only *true* before or after a certain event, e.g. "Eurydice is dead.", depend on the context of the narrative sequence (*durative-initial* or *-resultative*). Therefore, we do not consider them as parts of the overall background knowledge.

Each of the twelve sequences corresponding to one myth variant contains one or more of those statements, i.e. statements that hold true over the course of the entire variant (*durative-constant* statements). Those statements are *assertions* we can make about the domain knowledge, i.e. the world in which a plot takes place.

The statements describing the background knowledge were originally in German, but translated for this paper.

4 Domain modeling

We demonstrate the domain modeling approach using two variants of the myth of *Orpheus and Eurydice*. The English translation of the source text is shown below in Examples 1 and 2. The sequences of statements describing the plot and the narrative background knowledge are derived by experts in diverse mythological studies according to the hylistic approach [18].[4] From both texts, we can derive background information that holds true in the respective variant. Table 1 shows which assertions can be made from the information in the sequences. Those assertions form the ground truth, the *a priori* knowledge for the ontology modeling process. According to the hylistic approach, we only consider statements that are relevant to the *Orpheus and Eurydice* myth. Statements like "Linus is Orpheus' brother" are not considered, since they pertain to a different myth.

> (1) "But Orpheus, son of Oeagrus, they sent back with failure from Hades, showing him only a wraith of the woman for whom he came; her real self they would not bestow, for he was accounted to have gone upon a coward's quest, too like the minstrel that he was, and to have lacked the spirit to die as Alcestis did for

[2] https://mythoskop.de/
[3] https://ontology.vast-project.eu/
[4] https://www.uni-goettingen.de/en/556429.html

> the sake of love, when he contrived the means of entering Hades alive. Wherefore they laid upon him the penalty he deserved, and caused him to meet his death."[5]
>
> (2) "Now Calliope bore to Oeagrus or, nominally, to Apollo, a son Linus, whom Hercules slew; and another son, Orpheus, who practised minstrelsy and by his songs moved stones and trees. And when his wife Eurydice died, bitten by a snake, he went down to Hades, being fain to bring her up, and he persuaded Pluto to send her up. The god promised to do so, if on the way Orpheus would not turn round until he should be come to his own house. But he disobeyed and turning round beheld his wife;"[6]

Background information was collected for all twelve variants of the myth of *Orpheus and Eurydice*. Subsequently, a small controlled vocabulary[7] specifically for the myth was created that allows matching of concepts, such as consort/wife/female spouse → wife. The concepts are given in German (*skos:prefLabel*) and English (*skos:altLabel*). The vocabulary also includes definitions (*skos:definition*) for the interpretation of the concepts, e.g. the definition of the concept *son* would be "direct male descendant of a person".

While matching synonyms for the target languages, German and English, is a fairly straightforward task to automate, e.g. using WordNet [9] and GermaNet [5], the controlled vocabulary allowed us to create the ontologies more uniformly. Using controlled vocabulary for classes and relationships also helps to compare ontologies visually or by manual inspection. Additionally, the controlled vocabulary can be extended and re-used for other myths that contain similar concepts. Using those concepts, a set of twelve shallow ontologies were constructed following the guidelines outlined by Noy and McGuinness [11]. Important concepts are represented as ontology classes, such as terms for spouses or descendants, geographic concepts, or concepts of arts and music. Narrative characters like *Orpheus*, geographic locations and specialised concepts such as *Kitharodie* are individuals of the ontology. The resulting class hierarchy is shallow in the sense that only important higher-class concepts are modeled (e.g. wife → spouse → person).

In this regard, the ontologies are to our knowledge the only machine-readable and re-usable source of *source-specific* background knowledge for the individual sources. Figure 1 shows an example of a shallow domain ontology for the *Orpheus* myth in Apollodorus' library.

The information in each ontology corresponds to one myth variant and one source. The information is not combined into a single ontology for two reasons. Firstly, background or circumstantial information between ontologies may be contradictory, e.g. with regard to a characters ancestry. Secondly, if one statement is missing in one source, but is present in another we cannot make assumptions about the truth value of the information in the first source. For example, "Orpheus is the beloved son of Oiagros." implies *loves*(Oiagros,Orpheus) to be $true$. If another source does not contain that information, we cannot assume it to be $true$ or $false$.

Object properties in the domain ontologies contain all relations that are not *isA*-relations derived from the background statements. These contain information such as spousal relationships or locations, e.g. *isIn*(Person, Location). Object properties have role restrictions for domain and range, depending on the classes they apply to.

Each ontology has translations of class concepts, object properties and data properties in German and English (*skos:altLabel*).

Public semantic sources such as Wikidata contain information on narrative characters, but they do not distinguish between source-specific and general information. For instance, the Wikidata entry on *Orpheus*[8] states that his occupation is poet and writer, and that he was killed by Maenad. In the myth variants studied for this project, we can only derive that his profession was that of a musician, more specifically that he was a minstrel who practised Kitharodie (κιθαρῳδία). The manner of his death is discussed in multiple variants, where it is stated as 'being killed and torn to pieces by the

[5] Plato Symp. 179d http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0174:text=Sym.:section=179d&highlight=Orpheus

[6] Apollod. Lib. 1.3 http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0022:text=Library:book=1:chapter=3&highlight=Orpheus

[7] Controlled Vocabulary and ontologies in TTL-format are available under: https://gitlab.gwdg.de/franziska.pannach/hylva under a creative commons license.

[8] https://www.wikidata.org/wiki/Q174353

Table 1: Background information from two myth variants

Plato	Apollodorus
Orpheus is the son of Oeagrus. Eurydice is in Hades.	Orpheus is the son of Calliope and Oeagrus. Eurydice is the wife of Orpheus. Eurydice is in Hades. Orpheus practises minstrelsy. (κιθαρῳδία)

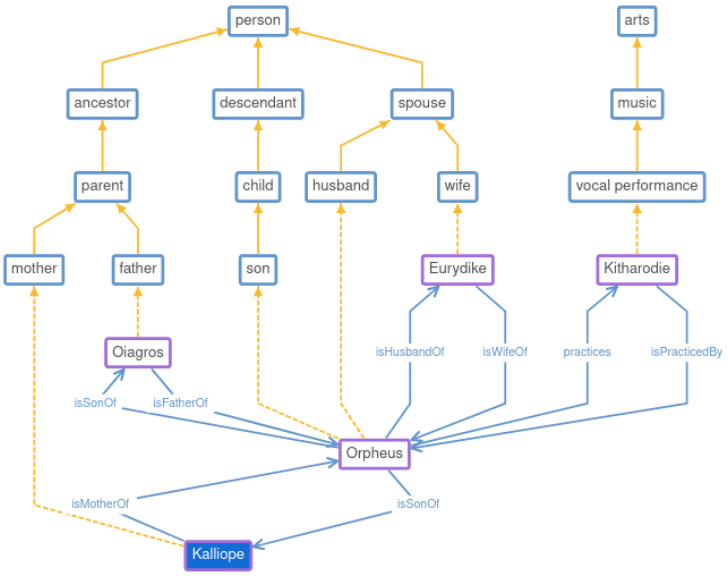


Figure 1: Ontology of 'Orpheus and Eurydice' concepts in Apollodorus' library

Thrakian women'.

Therefore, we cannot investigate or compare different views on the character *Orpheus* using resources like Wikidata. However, we link the information in the domain ontologies with the corresponding concepts in Wikidata via Wikidata ID, and Pleiades[9] ID in case of geographic locations.

5 Domain Comparison

The resulting domain ontologies can be used to compare the domains, i.e. the background information we have about the characters and the setting within the narrative variant. We can do so by applying two measures: Firstly, we can compare classes of the ontology. This answers the question 'Which general concepts are present in this narrative variant?' This way, we can interpret the background information in Apollodorus' library, as shown in Figure 1, as 'some people who are related to each other either by marriage or ancestry', and 'some music presented in the form of song'. Since the controlled vocabulary was created during the ontology modeling process, we can match classes easily.

[9] https://pleiades.stoa.org/

We define class overlap as:

$$CO = \frac{|C_1 \cap C_2|}{|C_1 \cup C_2|}, \tag{1}$$

where C_i is the set of classes of ontology i. Secondly, we can map the individuals of the ontologies to answer questions like "Who appears in this story?" and "Who is this story about?". Figure 1 shows the characters *Orpheus*, *Eurydice*, *Calliope*, and *Oiagros* as individuals. We match individuals iteratively by: name, alias, and WikidataID or PleiadesID if the individual is a geographic entity. The node for *Orpheus* in the example ontology in Figure 1 has the most in- and out-going relations, represented as arrows. Graphically, he is the most 'connected' character is the domain ontology, we can derive that he is most likely the main character. We define the individual overlap as:

$$IO = \frac{|I_1 \cap I_2|}{|I_1 \cup I_2|}, \tag{2}$$

where I_i is the set of individuals of ontology i. Furthermore, the characters in different mythological sources can be compared using relations (object properties) between two or more characters or con-

cepts. This way, a degree of similarity between two different characters can be estimated, e.g. if both characters are sons of a father who is a king, or if both characters love a woman who is located in the netherworld. For the twelve variants of *Orpheus and Eurydice* such a measure was implemented but yielded few interesting results because the characters were either already matched due to name, alias or Wikidata ID, or too dissimilar to be compared in a meaningful way (e.g. *Hades* and *Kitharodie*).

6 Results

The twelve ontologies are freely available in TTL-format for download.[10] Figure 2 shows the class and individual overlaps across variants. We see that for both class and individual overlap, the variant P_6 (Pausanias 9) is very dissimilar to the other variants.

We can match our results against information that is available have about the sources. e.g. Pausanias only briefly mentions the story of Orpheus and Eurydice in his travel report [12].[11]

On the level of ontology individuals, IO, the pair Plato-Hermesianax has the highest overlap score of 0.6. The highest score for the class overlap is 0.65 between the domain ontologies derived from statements based on the Mythographus Vaticanus and Apollodorus' Library.

In Figure 2c and 2d, we highlight only the closest matches between variants (without the self-matches on the matrix diagonal). Neither time of creation of the sources nor their geographic origin, e.g. Roman or Greek, seem to correspond to the similarity of the domain descriptions.

7 Discussion

The statements about the background information are based on the source texts of the original versions of a myth variant. To extract these is not a matter of simple NLP technique. Especially the decision on the truth value (single-point or durative) of a statement needs to be based on the source texts and made by informed experts on the material. This means that the extraction of these statements happens manually which is time-consuming. The construction of the ontology based on the background information, on the other hand, can be assisted semi-automatically using simple rules, e.g. for *isA*-relationships. Common concepts, such as geographical concepts and entities, are available in common thesauri and semantic web resources such as Wikidata. Their freely available data could be re-used for our purposes. However, to link them in shallow ontologies instead of creating them as classes might not always be the best option. For instance, we suggest modeling locations with a distinction between mythological (e.g. Hades) and real – past or present – locations (e.g. Macedonia). In this sense, the class distance (in our case the depth to the lowest common ancestor (LCA)) in the shallow ontology between Hades and Macedonia would have a value of two. If we applied Wikidata classes, those two concepts would not share a meaningful common ancestor beyond *Wikidata metaclass*.

As discussed at the end of Section 5, we do not report similarity measures for relationships (object properties) for the myth variants studied in this paper. However, this measure is interesting for inter-myth comparison, where different characters with similar features appear. It can also serve useful to compare re-use of mythological storytelling in modern fiction, e.g. comparing the myth of Persephone to Ginny Weasleys story in Harry Potter and the Chamber of secrets [15]. Furthermore, the hylistic analysis and the comparison of narrative domain knowledge using shallow ontologies can applied to other fictional genres as well, e.g. the study of folktales or comparison of different character representations in fanfiction, among others. We leave these efforts for future studies.

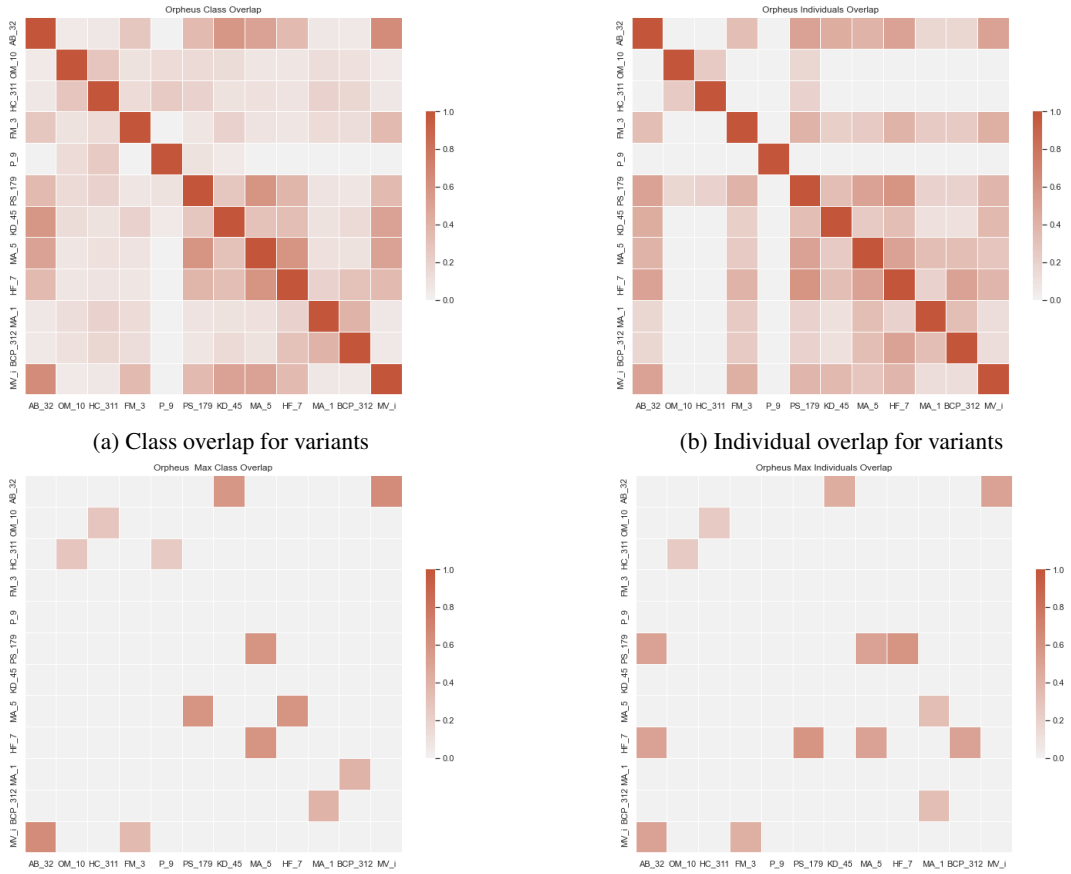
When studying modern texts in well-resourced languages, such as German or English, the extraction of sequences and subsequent ontology modelling could be assisted by automation through NLP methods, such as named entity recognition and semantic role labelling. With a larger number of texts and corresponding sequences, it would also be possible to automatically identify candidate statements from text. However, the creation of final sequences and knowledge bases, like the ones presented here, will most likely continue to include some form of manual work.

8 Acknowledgements

This work is funded by the DFG as part of the Myth-Research Group 2064 STRATA. The authors would like to thank Dr. Balbina Bäbler for her

[10] https://gitlab.gwdg.de/franziska.pannach/hylva Creative Commons license (CC-BY 4.0)

[11] http://www.perseus.tufts.edu/hopper/text?doc=Paus.+9.30.6&fromdoc=Perseus%3Atext%3A1999.01.0160



(a) Class overlap for variants

(b) Individual overlap for variants

(c) Maximum class overlap

(d) Maximum individual overlap

Figure 2: Overlap for variants of *Orpheus* journey to the netherworld

helpful insights and comments.

References

[1] C. M. Bowra. 1952. Orpheus and Eurydice. *The Classical Quarterly*, 2(3-4):113–126.

[2] Fabio Ciotti. 2016. Toward a formal ontology for narrative. *MATLIT: Materialidades da Literatura*, 4(1):29–44.

[3] Gösta Ingvar Gabriel, Brit Kärger, Annette Zgoll, and Christian Zgoll, editors. 2021. *Was vom Himmel kommt:Stoffanalytische Zugänge zu antiken Mythen aus Mesopotamien, Ägypten, Griechenland und Rom*. De Gruyter, Berlin, Boston.

[4] Fritz Graf and Ernst Badian. Eurydice.

[5] Birgit Hamp and Helmut Feldweg. 1997. Germanet-a lexical-semantic net for german. In *Automatic information extraction and building of lexical semantic resources for NLP applications*.

[6] Hermesianax. *Fragment 7*.

[7] Folgert Karsdorp and Antal Van den Bosch. 2016. The structure and evolution of story networks. *Royal Society open science*, 3(6):160071.

[8] A. N. Marlow. 1954. Orpheus in Ancient Literature. *Music and Letters*, XXXV(4):361–369.

[9] George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

[10] Arturo Nakasone and Mitsuru Ishizuka. 2006. Storytelling ontology model using RST. In *2006 IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, pages 163–169. IEEE.

[11] Natalya F Noy, Deborah L McGuinness, et al. 2001. Ontology development 101: A guide to creating your first ontology.

[12] Henry Arderne Ormerod, Richard Ernest Wycherley, William Henry Samuel Jones, et al. 1918. *Pausanias description of Greece*, volume 1. W. Heinemann.

[13] Federico Peinado, Pablo Gervás, and Belén Díaz-Agudo. 2004. A description logic ontology for fairy tale generation. In *Procs. of the Workshop on Language Resources for Linguistic Creativity, LREC*, volume 4, pages 56–61.

[14] Vladimir Propp. 1968. *Morphology of the Folktale*, volume 10. University of Texas Press.

[15] Richard A Spencer. 2015. *Harry Potter and the Classical World: Greek and Roman Allusions in JK Rowling's Modern Epic*. McFarland, Jefferson.

[16] Lei Xu, Albert Merono-Penuela, Zhisheng Huang, and Frank Van Harmelen. 2017. An ontology model for narrative image annotation in the field of cultural heritage. In *Proceedings of the 2nd Workshop on Humanities in the Semantic web (WHiSe)*, pages 15–26.

[17] Annette Zgoll and Christian Zgoll, editors. 2020. *Mythische Sphärenwechsel: Methodisch neue Zugänge zu antiken Mythen in Orient und Okzident*. De Gruyter, Berlin, Boston.

[18] Christian Zgoll. 2019. *Tractatus mythologicus: Theorie und Methodik zur Erforschung von Mythen als Grundlegung einer allgemeinen, transmedialen und komparatistischen Stoffwissenschaft*. De Gruyter, Berlin, Boston.