Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 4001–4010
Marseille, 20-25 June 2022

GRhOOT: Ontology of Rhetorical Figures in German

Ramona Kühn, Jelena Mitrović, Michael Granitzer
University of Passau
Innstrasse 43, 94032 Passau, Germany
{ramona.kuehn, jelena.mitrovic, michael.granitzer}@uni-passau.de

Abstract

GRhOOT, the **G**erman **Rh**et**O**rical **O**n**T**ology, is a domain ontology of 110 rhetorical figures in the German language. The overall goal of building an ontology of rhetorical figures in German is not only the formal representation of different rhetorical figures, but also allowing for their easier detection, thus improving sentiment analysis, argument mining, detection of hate speech and fake news, machine translation, and many other tasks in which recognition of non-literal language plays an important role. The challenge of building such ontologies lies in classifying the figures and assigning adequate characteristics to group them, while considering their distinctive features. The ontology of rhetorical figures in the Serbian language was used as a basis for our work. Besides transferring and extending the concepts of the Serbian ontology, we ensured completeness and consistency by using description logic and SPARQL queries. Furthermore, we show a decision tree to identify figures and suggest a usage scenario on how the ontology can be utilized to collect and annotate data.

Keywords: Ontologies, Language Modelling, Information Extraction, Information Retrieval, Semantics, Semantic Web, Knowledge Discovery/Representation

1. Introduction

Rhetorical figures are omnipresent: in advertising, argumentation, fake news, hate speech, music, and in everyday communication. But how can one recognize a rhetorical figure? In her book "Rhetorical Figures in Science", Fahnestock (2002) has analyzed different definitions of rhetorical figures: she states that most definitions have in common that figures are considered as a "departure from the normal usage". Quintilian (1996) describes a rhetorical figure as a departure from ordinary meaning, syntax, or form. This raises the question, what "ordinary" means, as figures are not exclusively used on certain occasions or by specific communities, but "across situations and registers" (Fahnestock, 2002). Fontanier (1968) also describes figures as principles of departure from a simpler expression: for example, the sentence "he is not the brightest bulb in the box" could be replaced by a simpler expression, e.g., "he is stupid".

Rhetorical figures are used to serve a certain communicative function, to increase memorability, or to attract the attention of the reader or listener. Depending on the figure, a certain effect is caused. This "form-function correlation" as Givón (1991) describes it, was first mentioned by Aristotle in *De Partibus Animalium* (Aristoteles et al., 1972).

Given the large amount of unstructured data generated every day on the Internet, automated text processing including natural language processing (NLP) techniques and machine learning algorithms has become essential to analyze and understand the human language: For example, using opinion mining or sentiment analysis gives insights into opinions about products, companies, or political parties (Solangi et al., 2018). Another field of application is the detection of fake news or hate speech to ensure compliance with policies of online platforms or to prevent criminal acts. The usefulness of rhetorical figures for argumentation mining purposes was described by Mitrović et al. (2017). The authors highlight that a formal ontological representation of rhetorical figures could positively influence investigations related to the persuasiveness of arguments. A similar idea is also considered in the context of propaganda detection by Hamilton (2021). The author highlights the relevance of ontologies of rhetorical figures for this research direction.

However, this task is not easy: figures are often implicit or hidden comparisons between two different domains that have common aspects, as in the case of metaphor[1] (Lakoff and Johnson, 2008). Although it is easy for a computer to detect e.g., repetitions, it is difficult to find the types of repetition that result in the creation of a rhetorical figure because of "accidental or irrelevant repetitions" (Dubremetz and Nivre, 2018). Another problem is that data on rhetorical figures are scarce, especially for figures that are not commonly used (Dubremetz and Nivre, 2017). In addition, as figures require knowledge about context and are highly dependent on cultural background, it is even hard for humans to agree on the presence of a figure (Hall Maudslay et al., 2020). Another problem is that only few formal, often contradictory descriptions and classifications of rhetorical figures exist. Automatically detecting rhetorical figures requires a formal description of each figure that respects the different cate-

[1]The reader can find a glossary at the end with the translated German explanation of the rhetorical figures mentioned throughout the paper.

gorizations, but is specific enough to uniquely describe and group the figures. Kelly et al. (2010) took a crucial step in this direction by classifying rhetorical figures into either tropes (figurative usage and conceptual similarities e.g., metaphor, irony), schemes (formal figures based on a certain syntax or structure e.g., antimetabole: "All for one, one for all"), or chroma (a figure used intentionally, e.g., rhetorical questions: "Are you serious?"), and assigning a linguistic domain (phonological, morphological, etc.). This work was the inspiration for "RetFig", an ontology of rhetorical figures in the Serbian language (Mladenović and Mitrović, 2013), that contains formal descriptions of 98 distinctive figures.

Current approaches, however, focus only on the modeling of two or three figures, and almost always for the English language (e.g., chiasmus, epanaphora, epiphora (Dubremetz and Nivre, 2018), or gradatio, incrementum and climax (O'Reilly et al., 2018)). RetFig is so far the only known approach that tries to model all relevant rhetorical figures in one language, namely Serbian. It is therefore necessary to extend this ontology for other languages (Hamilton, 2021).

In this work, we aligned with the RetFig approach to develop an ontology of rhetorical figures in German – called GRhOOT – that contains 110 different figures at the moment. It is available online[2], as well as the documentation generated with the Live OWL Documentation Environment (LODE)[3]. The ontology and further information is also available at the author's github page[4]. As the ontology contains rhetorical figures in a machine-readable format, it will lead to a better understanding on how and why different texts have different effects because they are using certain figures. This approach has a high potential for offering increased explainability of NLP systems.

This paper describes the following contributions:

- Development of an ontology in Protégé[5] (open-source ontology framework) for the most common rhetorical figures in German (110 in total) in line with the RetFig ontology.
- Design of a modular ontology that respects different categorization approaches and allows for an easy extension or translation into other languages (we provide the English and the Serbian name for each figure).
- Formulation of four competency questions in accordance with the guidelines of Grüninger and Fox (1995) for the design and validation of ontologies.
- Establishing the foundation to create annotated datasets for rhetorical figures.

We also ensured the completeness of the ontology with description logic (DL) and SPARQL queries. We show how the ontology can be used in the form of a decision tree or an interactive web application to identify and annotate rhetorical figures.

The remainder of the paper is structured as follows: Section 2 describes similar ontologies of rhetorical figures. Next, the modeling process of the ontology and its details are described in Section 3. We show how a rhetorical figure in the ontology looks like and investigate the commonalities with and differences to the RetFig ontology. The completeness of the newly built ontology is checked in Section 4, whereas Section 5 presents two possible applications for the ontology. Finally, the paper ends with the conclusion and ideas for future work in Section 6.

2. Related Work

The problem with automatic detection of rhetorical figures is the lack of formal descriptions or categorization of figures that could be used for automated text processing. In her book "Rhetorical Figures in Science", Fahnestock (2002) provides a clear overview of different attempts to categorize rhetorical figures over the centuries. It starts with the *Rhetorica Ad Herennium* differentiating between figures of diction and figures of thought. According to Fahnestock, Quintilian (1996) reuses this classification in his *Institutio oratoria*, but differentiates between tropes and figures. Operations of figures such as repetition, omission, separation, and conjunction are mentioned by Peacham (1971). In the following years, more attempts were made to categorize different figures. However, this led to an inconsistent and sometimes contradicting view on rhetorical figures. For the German language, Lausberg (1990) wrote the de-facto standard reference book on this topic. Still, the author is aware that it is not free of contradictions. Harris and DiMarco (2009) made one of the first attempts to construct a formal machine-readable description of rhetorical figures. They focused on linguistic operations (addition, deletion, etc.) and cognitive affinities (comparison, symmetry, etc.). A formal description of rhetorical figures with addition or omission operations is provided. However, the authors notice that this description finds its limitations when a structure with same phrases with different meaning, or the same word in different forms appear. Kelly et al. (2010) try to overcome those limitations by defining an advanced set of relations.

This work then inspired Mladenović and Mitrović (2013) to model an ontology of rhetorical figures customized for the Serbian language in a machine-readable format, the Web Ontology Language (OWL). O'Reilly and Paurobally (2010) built an ontology of several rhetorical figures in Protégé that should be able to locate rhetorical figures of speech in text. However,

[2] https://ramonakuehn.de/grhoot
[3] https://essepuntato.it/lode/
[4] https://github.com/kuehnram/GRhOOT-Ontology
[5] https://protege.stanford.edu/

the authors criticize their own approach by mentioning that they cannot describe all rhetorical figures with "simple logic rules". A further attempt to model ontologies was made by Harris et al. (2017): The authors present a top-down, middle-out, and a bottom-up approach in combination with cognitive aspects to create their ontology. They also focus on the relation of different figures and their interplay. An *isA* relationship is modeled that expresses a hierarchical dependency of figures.

O'Reilly et al. (2018) described an ontology for "the figural structure and aspects of argumentation". The ontology focuses on three rhetorical figures (gradatio, incrementum, and climax) and is written in OWL. The authors consider the connectedness of figures, namely the fact that some figures are the intersection of others. However, they remark that they "only model the surface features of an argument" by modelling only three figures.

A survey of ontological modeling of rhetorical concepts is given by Mitrović et al. (2017). They compare the Rhetorical Structure Theory (RST), the Serbian RetFig ontology, and the Lassoing Rhetoric project.

Hamilton (2021) models an ontology for propaganda detection while pointing out that rhetorical figures are highly relevant in this field. A timeline of the different ontologies developed in the recent years is shown in Fig. 1.

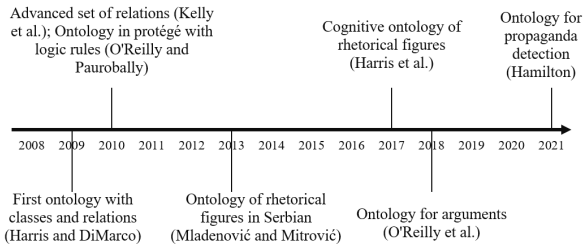

Figure 1: Timeline of ontologies for rhetorical figures.

The ontology presented in this paper combines several approaches from the research mentioned above. It is the first ontology with a focus on rhetorical figures in the German language. In addition, we show possible use cases of how the ontology can be used to identify rhetorical figures and to collect annotated data.

3. Modeling an Ontology of Rhetorical Figures

The process of modeling rhetorical figures is a challenging task: multiple definitions, classifications, and names for the same rhetorical figure exist that have to be respected. Below, we describe our approach to modeling the ontology (Section 3.1). GRhOOT is presented in detail in Section 3.2. Section 3.4 compares the RetFig ontology with our ontology.

3.1. Modeling Approach

Gómez-Pérez (1999) specifies several properties of an ontology. The ontology shall be consistent and not having contradictory output for valid input. In addition, the ontology should be complete. It is difficult to prove completeness, but it should be possible to infer missing elements. Furthermore, the ontology has to be concise by not having unnecessary definitions or redundancies. The effort of adding new definitions or more knowledge without changing existing properties should be low (expandability) as well as the sensitiveness that describes how small changes affect the defined properties. Devedzić (2002) extends this list by mentioning that terms should be precisely defined and definitions should be structured. The users of the ontology should be able to express what they want to say (high expressiveness). The ontology shall be coherent and interoperable as well as providing stability and scalability.

As shown in Section 2, attempts at modeling rhetorical figures can never be complete or without contradiction as it lies in the nature of those figures to have several intersections and similarities. It was a challenge to achieve consistency for our ontology as it must be as abstract and general as possible while still providing enough information to identify the different figures. Conciseness is guaranteed by not modeling synonyms (e.g., different notations of a figure) separately, but by linking them to the original figure by the relation *isSynonym*. New classes, definitions, or properties can be added easily, guaranteeing expandability and sensitiveness. Due to the modular structure, the ontology can be translated to other languages. The used formal language OWL offers interoperability with different popular programming languages, such as Python.

A detailed description of the different steps in an ontology development process from scratch is provided by Fernández-López et al. (1997). They extract three key points that should be considered during each development process:

- The purpose of the ontology, use cases, possible end-users.
- Level of formality.
- Scope and terms of the ontology.

We followed this development process and defined the following: The goal of the GRhOOT ontology is to formally describe rhetorical figures in the German language to have a machine-readable representation. It should be able to specify several rhetorical figures with different properties. The use cases are manifold: It can be used to develop a rule-based approach for rhetorical figure detection. It can also be used to guide people to identify rhetorical figures. Therefore, it can support human annotators, too. The scope of the ontology are the most important rhetorical figures of the German language. The rhetorical figures were chosen on the basis of representative literature: Lausberg's Handbook

of Literacy Rhetoric (Lausberg, 1990) is the standard reference book for rhetorical figures in the German language. Furthermore, literature from Berner (2011), Mayer (2007), Göttert (2009), Plett (1991), and the lexicon of language "Metzler Lexikon Sprache" (Glück, 2016) were utilized to find the most common rhetorical figures with definitions and examples in the German language.

Our methodology to create the ontology was to first define for each figure in the Serbian RetFig (Mladenović and Mitrović, 2013) if a German counterpart exists. If yes, we evaluated together with a German linguistic expert and the literature mentioned above if the German figure has similar properties or if they need to be remodeled. Furthermore, we identified common German rhetorical figures that are not present in the RetFig ontology and modeled them accordingly.

The challenge was to reach consensus as the definitions of rhetorical figures can be very different, even within one language. Some of the literature mentioned above consider some figures to be completely equal whereas others see a slight difference. For example, Berner (2011) considers isocolon just a different name for parallelism, while other say that parallelism and isocolon have both a parallel structure but isocolon is also of equal length. It is not possible to reflect all views in the ontology, so we tried to focus on the opinion presented in the majority of our resources.

To assure that the ontology fulfils the tasks for which it was developed, we use informal competency questions as described by Grüninger and Fox (1995). The number of questions that reflect all possible combinations is quite high (e.g., show all figures where a word/letter/syllable is affected; where a word is affected in a phrase/sentence etc.). Therefore, we will show only five exemplary informal competency questions that have been formulated for GRhOOT:

Q1: Which figures have their defining element in the beginning?

Q2: Which figures with their defining element in the beginning are a figure of speech or a figure of thought?

Q3: Which figures belong neither to the rhetorical group of tropes nor to the group of figures of speech?

Q4: Which figures occur in a word or affect a word?

Q5: In which figures is a letter omitted?

Questions Q4 and Q5 were also used for validation in RetFig (Mladenović and Mitrović, 2013). This allows for a comparison between the two ontologies.

Those questions will be used to ensure that the ontology behaves as expected (Section 4).

3.2. GRhOOT Details

The following section describes the structure, classes, properties, and individuals of GRhOOT. They are highlighted in *italic* except for the rhetorical figures to maintain clarity. A glossary of the mentioned rhetorical figures can be found at the end of this paper.

3.2.1. Structure

The GRhOOT ontology uses the same basic structure as RetFig (Mladenović and Mitrović, 2013). It consists of classes, object and data properties, and individuals. The hierarchical tree-like structure of the classes is shown in Fig. 2. At the top level is the *owl:Thing*. The "top-concepts", as Mladenović and Mitrović (2013) name them, are the class of *rhetorical entities* and the class of *linguistic entities*. A *rhetorical figure* (highlighted in bold in the figure) belongs to both classes as it has both linguistic and rhetorical features. A *linguistic entity* has six subclasses, whereas a *rhetorical entity* has only two, namely *rhetorical group* and *rhetorical figure*. The ontology has 14 object properties, where three of them have 10, 15, and 4 subproperties. Overall, 167 individuals are defined, where 110 of them are *rhetorical figures* and the others are individuals like *sentence*, *verse*, etc. The class and property names have been translated from German here for better readability.

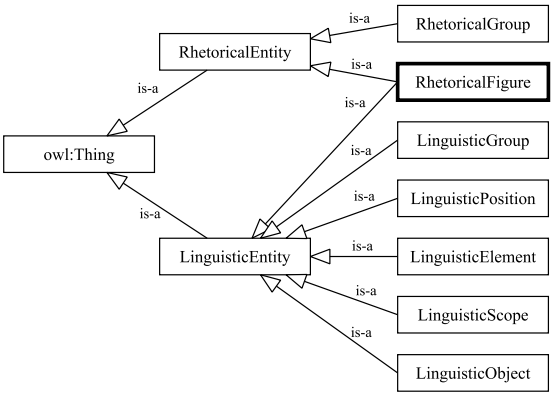

Figure 2: Class structure of GRhOOT and RetFig. Adapted from Mladenović and Mitrović (2013).

In the future, we think that the structure of the GRhOOT ontology can be adapted easily to other languages, especially for languages similar to German or Serbian. The names of classes and individuals can be translated at once in Protégé. The time-consuming part is to compare the properties for each figure, to agree on properties that differ, and to find the definitions and exemplary sentences containing the rhetorical figure (the examples are actually not required for the formal description, but rather a nice-to-have feature).

3.2.2. Linguistic Group

The *linguistic group* describes how a figure is constructed. A figure belongs to one or more of five linguistic groups: *morphological*, *phonetic*, *pragmatic*, *semantic*, and *syntactic*.
If the figure is *morphological*, it is formed by the use of different inflectional forms of a word (e.g., a play on words can be *morphological*). If the figure is created by letters, group of letters or syllables, it belongs to the group of *phonetic* figures (e.g., alliteration: "Two terrible tigers"). Changing the literal meaning over more sentences results in the figure belonging to the *pragmatic* figures (e.g., rhetorical question). If the use of the figure can change the literal meaning, it belongs to the *semantic* figures (e.g., metaphor, rhetorical question). It is *syntactic* if the linguistic order is changed, therefore adding, repeating, or omitting elements (e.g., anaphora: "I have [...]. I have [...].").

3.2.3. Linguistic Position

Depending on the figure, a *linguistic position* is assigned to express where the defining element (e.g., repetition of a word, a letter, etc.) of the figure appears: e.g., in the *beginning*, *middle*, *end*, *beginning and end*, *whole*, etc. In case of an anaphora, the position is *beginning*. The relationship *isInPosition* indicates the position.

3.2.4. Linguistic Scope

The *LinguisticScope* describes whether a rhetorical figure occurs in a *sentence*, a *verse*, a *paragraph*, a *phrase*, *stanza*, or a *word*. It is expressed by the relationship *isInAreaOf*. For example, an alliteration occurs over a complete *sentence* or a *verse*.

3.2.5. Linguistic Element

The *linguistic element* encompasses the element of a figure that is affected by an operation. The element can be either a *letter*, *vowel*, *consonant*, *sentence*, *phrase*, *verse*, *word*, etc. An operation can be e.g., *adding*, *omitting*, *repeating* this element (cf. 3.2.6). For example, in the figure aphaeresis, the element *letter* or *syllable* is omitted ("The King hath cause to 'plain." (Shakespear's King Lear) where 'plain is used instead of complain).

3.2.6. Linguistic Operation

Operations are *addition*, *conversion*, *reduction/omission* or *repetition*. For each operation, the affected linguistic element is specified. The details of the operation provide additional information, e.g., if it is a *repetition* or *addition* of a *word* with the *same meaning*, *different meaning*, *different form*, *stronger* or *weaker meaning*, etc. In case of the figure aphaeresis (see 3.2.5), it is an *omission*.

3.2.7. Linguistic Object

The individuals of the class *LinguisticObject* are used to describe in which object the rhetorical figure appears by the relation *isInObject*. Possible individuals are *phrase object*, *sentence object*, *verse object*, and *word object*.

3.2.8. Rhetorical Group

The *rhetorical group* specifies to which category a figure belongs. Exactly one *rhetorical group* is assigned to a figure. Harris and DiMarco (2009) state that there are two "traditional categories", the tropes and schemes. Tropes are figurative or abstract (e.g., metaphor: "brightest candle"), whereas schemes are "formal" (e.g., antimetabole: "all for one one for all"). Through this formality, the authors claim that those figures are easier to be detected automatically. Kelly et al. (2010) differentiate between tropes, schemes, and chroma. Tropes are again described as conceptual figures, and schemes are figures that are also formal figures, for example rhymes that are based on the sound of the figure. Figures of the group chroma are those that are used intentionally to gain the attention of the audience (e.g., rhetorical questions: "Are you serious?").
In our ontology, the scheme is divided into three further groups, as we are modeling the groups in alignment with RetFig (Mladenović and Mitrović, 2013):

- *Figures of speech*
- *Figures of construction*
- *Figures of thought*
- *Tropes*

Chroma is not considered as it is difficult to decide when a figure is used intentionally. The group of *figures of speech* contains all figures that are based on sounds, e.g., alliteration. A simple explanation would be that even if one does not understand the language, one would notice the use of the figure because of prosody. *Figures of thought* are defined by a certain syntax that helps to support the claim of thought (Lausberg, 1990), e.g., rhetorical questions or parallelism. *Figures of construction* are also based on mostly syntactic features because they are actually constructed by adding, omitting, or repeating elements. For example, syllables or letters are omitted in the figure aphaeresis ("plain" instead of "complain").

3.2.9. Rhetorical Figure

This class contains all individuals of the type *Rhetorical Figure*, so the actual rhetorical figures. As described in 3.1, the standard reference books and representative literature of rhetorical figures in German were used to define the most common rhetorical figures.
The figures modeled in our GRhOOT ontology have a unique internationalized resource identifier (IRI), a textual definition of the figure, and the English name. GRhOOT also provides the Serbian name of a figure (if it exists). The German naming is more convenient to be used by German speaking users, but the names in other languages allow a better alignment with other ontologies or the translation of GRhOOT.

There is at least one exemplary sentence for each figure. All figures contain a description of their relations and properties, e.g., a figure → *isInAreaOf* (*LinguisticScope*) → *word* means that the figure appears within a word.

3.3. Exemplary Model of a Rhetorical Figure

Through the example of alliteration ("*t*wo *t*errible *t*igers"), we want to show the structure of our ontology in Turtle syntax. The prefix before each # was omitted and properties/classes translated for better readability.

```
### #Alliteration
<#Alliteration> rdf:type owl:NamedIndividual ,
<#RhetoricalFigure> ,
  [ rdf:type owl:Restriction ;
  owl:onProperty <#isLinguisticGroup> ;
  owl:someValuesFrom <#LinguisticGroup>] ,
  [ rdf:type owl:Restriction ;
  owl:onProperty <#isRhetoricalGroup> ;
  owl:someValuesFrom <#RhetoricalGroup>] ;
<#isInPosition> <#Whole> ;
<#isLinguisticGroup> <#Phonetic> ;
<#isRhetoricalGroup> <#FigureOfPronunciation> ;
<#isInObject> <#Sentenceobject> , <#Verseobject> ;
<#isInAreaOf> <#Sentence> , <#Verse> ;
<#isRepeated> <#Consonant> ;
<#Name> "Alliteration"^^xsd:string ;
<#isExample>"\"Pickled Peppers." "^^xsd:string ,
    "\"Two terrible tigers\" "^^xsd:string ;
rdfs:comment "homophone initial sound of stressed syllable
      inside a group of words (Metzler)."^^xsd:string ;
rdfs:label "Alliteration"^^xsd:string ;
rdfs:seeAlso "Aliteracija"@sr ,
"Alliteration"@en .
```

The formal RDF description is suited for automated processing, while the graphical representation shown in Fig. 3 is easier for humans to grasp. The graph indicates the relations (e.g., *isInAreaOf*) and the respective values (e.g., *sentence*, *verse*, etc.). Those relationships are assigned to the classes by dotted arrows. To be completely precise, *owl:thing* is the superclass of *LinguisticEntity* and *RhetoricEntity* at the top, but it is not shown in the graph. The reader may notice that the graph is a more detailed version of the general structure that was presented in Fig. 2.

3.4. Retfig vs. GRhOOT

GRhOOT, the ontology of rhetorical figures in German that we describe in this paper, is based on RetFig (Mladenović and Mitrović, 2013), the ontology of rhetorical figures in Serbian. In the following, we describe the properties and structure of RetFig and compare it to our ontology. Mladenović and Mitrović (2013) mention the following properties of their ontology:

- It is a formal domain ontology.
- It describes and defines rhetorical figures in the Serbian language.

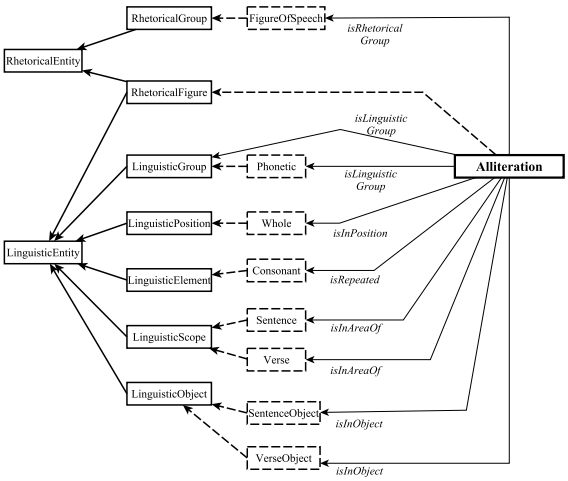

Figure 3: Tree-like structure of the figure alliteration.

- It can be shared and combined with other linguistic projects and ontologies.
- It serves as the basis for other specific task ontologies that are used for annotating rhetorical figures in Serbian.

RetFig contains twelve figures that do not exist in the German language or are not considered as rhetorical figures and are therefore not represented in GRhOOT. Vice versa, we identified and modeled 36 additional figures in the German language that are not modeled in RetFig.
Most of those figures that exist in both languages (approximately 60 figures) can be modeled identically, but some figures have different properties: For example, in the German figure syllepsis, a word element is omitted, whereas a word element is added in the Serbian equivalent silepsa. GRhOOT also contains at least one exemplary sentence for each figure to illustrate the use of a rhetorical figure in context.

4. Completeness Check of GRhOOT

We test if the ontology is complete, consistent, and delivers the expected output. We formulate the competency questions from Section 3.1 as queries and check if the result contains all expected rhetorical figures. The used ontology development environment Protégé offers different possibilities to execute queries on ontologies: The less formal way is to group individuals by their usage according to specific properties. For example, individuals can be grouped by the property *trope* to show all figures of the rhetorical group tropes. However, this only allows simple searches as multiple properties cannot be combined. For validation purposes, Protégé offers description logic (DL). In contrast to the less formal approach, more complex queries can be formulated specifying multiple properties and their values. It is also possible to execute SPARQL queries directly in

Protégé. SPARQL is a query language for the Resource Description Framework (RDF) that is used to query graphs like our OWL-ontology. In addition, Mladenović and Mitrović (2013) and O'Reilly et al. (2018) used SPARQL for the validation of their ontologies.
For Q1 and Q2, we will use DL queries. Q3-Q5 will be answered by SPARQL queries. All queries can be found online[6]. The result of the DL query for Q1 shows us all figures with the defining element in the beginning. With DL queries, multiple statements can be logically combined with `and/or` operators. So we can answer Q2, showing all figures with the position in the beginning that are a figure of speech or a figure of thought.
When looking into the details of the query results, it can be seen that all the specified properties are fulfilled and no figure matching the criteria has been left out. An interesting aspect is however, that the basic DL queries used by Protégé cannot handle negations. For example, the figures cannot be filtered for instances that do not fulfill a certain property like figures `not` belonging to a class (e.g., not a figure of speech). This is due to the open world assumption of DL queries (Calvanese et al., 2007). If a figure does not have a certain property, it does not mean that it could not belong to the class nevertheless. This means that DL queries can only handle explicit negations, therefore specifying all properties that it does not fulfill. As this is too much effort for the modeling process, a more powerful query language like SPARQL is necessary for those queries, as it is not based on the open world assumption.
To show negation in SPARQL, we use the competency question Q3 to show all figures with their defining element in the beginning, but belonging neither to the rhetorical group of tropes nor figures of construction (see Listing 1). It is actually Q2 inverted, so we expect the same output.

```
SELECT distinct ?figure
WHERE {
  ?figur ont:isRhetoricalGroup ?group .
  ?group ont:Name ?groupName .
  ?figur ont:isInPosition ?position .
  ?position ont:Name ?posName .
Filter (?groupName != "Trope" &&
   ?groupName != "figureOfConstruction"
   && ?posName = "Beginning") }
```

Listing 1: Q3 Query.

In the following, we want to intentionally execute the same queries used by Mladenović and Mitrović (2013) in RetFig so that we can compare the results. We first execute the SPARQL query to retrieve all figures that occur in or affect a word (Q4).
The comparison with the result of the same query in RetFig shows that the result of the German ontology contains 12 figures, whereas the Serbian contains only 9. On the one side, the Serbian figures diaeresis and symbol are missing as they are not considered figures in German. On the other side, the German ontology contains the figures malapropism, chiffre, archaism, and pejoration that are not modeled in Serbian.
Using a SPARQL query to show all figures where a letter is left out (Q5), we obtain the same result as Mladenović and Mitrović (2013).
Both the SPARQL and DL queries have proven that the ontology delivers the expected results. SPARQL queries are more complex, but also more powerful. As easy integration in Python is possible, it seems as a good choice for further applications of the ontology.[7]
In the following section, we will draft different applications of our ontology.

5. Possible Applications

The ontology can support its users in the process of identifying rhetorical figures and their names. It can serve as a codebook and guideline for people in the process of annotating rhetorical figures in text. A decision tree can facilitate the detection of figures, e.g., for figures containing repetitions as they follow a syntactical pattern. We constructed such a decision tree for figures of repetition in our ontology. It is shown in Fig. 4. Some leaf nodes contain multiple figures as figures can be related: For example, anticlimax is a more specific form of gradatio ("Anticlimax: Gradatio downwards" (Berner, 2011)). In those cases, the textual definition and the example sentences of the ontology can help to uniquely identify a figure. The disjunctive normal form for a path through this tree could look like the following:
`(Repetition = yes) ∧ (DifferentForm = Word) → Gradatio`
or
`(Repetition = yes) ∧ (SameForm = Word) ∧ (isInPosition = Beginning) → Anaphora`

However, the decision tree does not guide users through the process and is not interactive. Another possibility to make the ontology accessible to a broader audience is the development of a web application. It can collect a great amount of annotated data and actively support users in finding rhetorical figures. We developed a preliminary web application that is not yet publicly available. A clearly structured interface guides the users through the process of selecting matching properties.
Users can either submit an own text/sentence or select an example from the pool containing unlabeled data (e.g., news articles, social media, etc.). For example, if users found a repetition of the same word, the interface asks about the location (e.g., *Beginning*) or which

[6] https://github.com/kuehnram/GRhOOT-Ontology/blob/main/competency-questions.txt

[7] The SPARQL queries in Python can be found here: https://github.com/kuehnram/GRhOOT-Ontology/blob/main/Python_SPARQL.py

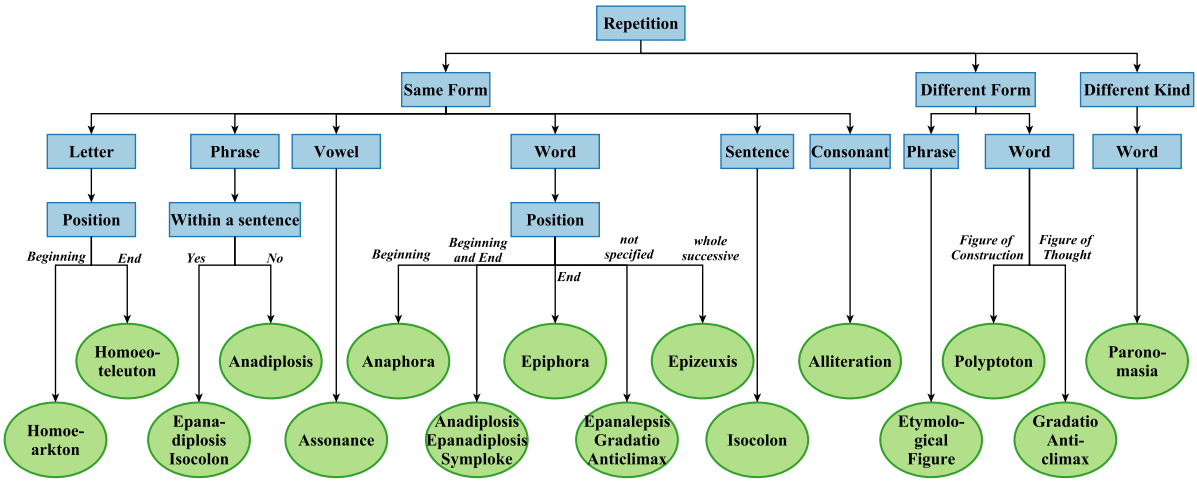

Figure 4: Decision tree for figures of repetition.

element is affected. The possible solutions are shown with definitions and exemplary sentences. In the background, the selected properties are transformed into SPARQL queries that are executed on the ontology.
In addition, an active learning algorithm (a semi-supervised machine learning algorithm) will be used in the future. The advantage is that this algorithm can be trained on a small set of labelled data. An extensive overview and literature survey was written by Settles (2009). They also state that pool-based sampling is more common in application scenarios and that there are no limitations regarding memory or processing power. Therefore, we will use this sampling technique. Also, a human-in-the-loop approach can be used: Budd et al. (2021) involve a human for medical image analysis, and Hauptmann et al. (2006) let a human help in efficient video search.

6. Conclusion and Future Work

The paper presents GRhOOT, an ontology of rhetorical figures in the German language. The RetFig ontological modeling of Mladenović and Mitrović (2013) was used as a basis for our work. We took special care in modeling the figures according to the rules and tradition of the German language, and modeling additional figures that do not exist in RetFig. Furthermore, we added examples for each figure to facilitate both human and automatic annotation aided by our ontology. The formal structure helps to automatically detect rhetorical figures in texts, therefore contributing to a better understanding and a higher explainability of NLP systems.
We showed that the ontology is complete, presented possible applications, and how the ontology can facilitate the process of identifying rhetorical figures and collecting annotated data. Future work includes the following tasks:

- Extending the ontology by adding hierarchical relations between several figures (O'Reilly et al., 2018).
- Using the ontology to detect certain figures that are present in implicit hate speech, fake news, and complex arguments, as suggested by Mitrović et al. (2017).
- Extending the ontology by functions/cognitive aspects of figures, e.g., argumentative power of antimetabole, memorability of alliteration, or persuasiveness of litotes (Mitrovic et al., 2020).
- Building an ontology of rhetorical figures in English and other languages.
- Development of a web application with an active learning approach to collect data.
- Assisting in building a multilingual Wikipedia (Vrandečić, 2021) by using the ontology for articles about rhetorical figures.

For now, the focus lies on the extension of the ontology and the collection of data via the web application that will be developed. In the near future, the focus will be on special figures that occur in hate speech or fake news to detect them more precisely.

Acknowledgments

The project on which this report is based was funded by the German Federal Ministry of Education and Research (BMBF) under the funding code 01—S20049. The author is responsible for the content of this publication.
We further thank Ms Yvonne Zörner for providing her linguistic expertise. We also thank the anonymous reviewers for their valuable comments and suggestions.

7. Bibliographical References

Aristoteles, Balme, D. M., and Gotthelf, A. (1972). *Aristotle's "De Partibus Animalium I": And "De Generatione Animalium I"(with Passages from II. 1-3)*. Clarendon Press.

Berner, G. (2011). *Vollständiges Kompendium der rhetorischen Mittel, Stilfiguren und Tropen für Oberstufenschüler und Studienanfänger -*. GRIN Verlag, München.

Budd, S., Robinson, E. C., and Kainz, B. (2021). A survey on active learning and human-in-the-loop deep learning for medical image analysis. *Medical Image Analysis*, 71:102062.

Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., and Rosati, R. (2007). Eql-lite: Effective first-order query processing in description logics. In *IJCAI*, volume 7, pages 274–279.

Devedzić, V. (2002). Understanding ontological engineering. *Communications of the ACM*, 45(4):136–144.

Dubremetz, M. and Nivre, J. (2017). Machine learning for rhetorical figure detection: More chiasmus with less annotation. In *Proceedings of the 21st Nordic Conference on Computational Linguistics*, pages 37–45, Gothenburg, Sweden, May. Association for Computational Linguistics.

Dubremetz, M. and Nivre, J. (2018). Rhetorical figure detection: chiasmus, epanaphora, epiphora. *Frontiers in Digital Humanities*, 5:10.

Dudenredaktion. (a). Antimetabole. `https://www.duden.de/rechtschreibung/Antimetabole`. Accessed: 2022-01-16.

Dudenredaktion. (b). Malapropismus. `https://www.duden.de/rechtschreibung/Malapropismus`. Accessed: 2022-01-16.

Fahnestock, J. (2002). *Rhetorical figures in science*. Oxford University Press on Demand.

Fernández-López, M., Gómez-Pérez, A., and Juristo, N. (1997). Methontology: from ontological art towards ontological engineering.

Fontanier, P. (1968). *Les figures du discours*, volume 15. Paris, Flammarion.

Givón, T. (1991). Isomorphism in the grammatical code: cognitive and biological considerations. *Studies in language*, 15(1):85–114.

Glück, H. (2016). *Metzler Lexikon Sprache*. Springer.

Gómez-Pérez, A. (1999). Evaluation of taxonomic knowledge in ontologies and knowledge bases.

Göttert, K.-H. (2009). *Einführung in die Rhetorik*. utb GmbH.

Grüninger, M. and Fox, M. S. (1995). Methodology for the design and evaluation of ontologies.

Hall Maudslay, R., Pimentel, T., Cotterell, R., and Teufel, S. (2020). Metaphor detection using context and concreteness. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 221–226, Online, July. Association for Computational Linguistics.

Hamilton, K. (2021). Towards an ontology for propaganda detection in news articles. In Ruben Verborgh, et al., editors, *The Semantic Web: ESWC 2021 Satellite Events*, pages 230–241, Cham. Springer International Publishing.

Harris, R. and DiMarco, C. (2009). Constructing a rhetorical figuration ontology. In *Persuasive Technology and Digital Behaviour Intervention Symposium*, pages 47–52. Citeseer.

Harris, R. A., Di Marco, C., Mehlenbacher, A. R., Clapperton, R., Choi, I., Li, I., Ruan, S., and O'Reilly, C. (2017). A cognitive ontology of rhetorical figures. *Cognition and Ontologies*, pages 18–21.

Hauptmann, A. G., Lin, W.-H., Yan, R., Yang, J., and Chen, M.-Y. (2006). Extreme video retrieval: Joint maximization of human and computer performance. In *Proceedings of the 14th ACM International Conference on Multimedia*, MM '06, page 385–394, New York, NY, USA. Association for Computing Machinery.

Kelly, A. R., Abbott, N. A., Harris, R. A., DiMarco, C., and Cheriton, D. R. (2010). Toward an ontology of rhetorical figures. In *Proceedings of the 28th ACM International Conference on Design of Communication*, pages 123–130.

Lakoff, G. and Johnson, M. (2008). *Metaphors we live by*. University of Chicago press.

Lausberg, H. (1990). *Handbuch der literarischen Rhetorik: eine Grundlegung der Literaturwissenschaft*. Franz Steiner Verlag.

Mayer, H. (2007). *Rhetorische Kompetenz*. utb GmbH.

Mitrović, J., O'Reilly, C., Mladenović, M., and Handschuh, S. (2017). Ontological representations of rhetorical figures for argument mining. *Argument & Computation*, 8(3):267–287.

Mitrovic, J., O'Reilly, C., Harris, R. A., and Granitzer, M. (2020). Cognitive modeling in computational rhetoric: Litotes, containment and the unexcluded middle. In *ICAART (2)*, pages 806–813.

Mladenović, M. and Mitrović, J. (2013). Ontology of rhetorical figures for serbian. In Matoušek Habernal, editor, *Text, Speech, and Dialogue*, pages 386–393, Berlin, Heidelberg. Springer.

O'Reilly, C. and Paurobally, S. (2010). Lassoing rhetoric with owl and swrl. *Unpublished MSc dissertation. Available: http://computationalrhetoricworkshop. uwaterloo. ca/wpcontent/uploads/2016/06/LassoingRhetoricWithOWLAndSWRL. pdf*.

O'Reilly, C., Wang, Y., Tu, K., Bott, S., Pacheco, P., Black, T. W., and Harris, R. A. (2018). Arguments in gradatio, incrementum and climax; a climax ontology. In *Proceedings of the18th work-*

shop on Computational Models of Natural Argument. Academic Press.

Peacham, H. (1971). *The Garden of Eloquence...* H. Jackson.

Plett, H. F. (1991). *Einführung in die rhetorische Textanalyse*. Buske.

Quintilian, M. F. (1996). Institutio oratoria (translated by H.E. Butler).

Settles, B. (2009). Active learning literature survey.

Solangi, Y. A., Solangi, Z. A., Aarain, S., Abro, A., Mallah, G. A., and Shah, A. (2018). Review on natural language processing (nlp) and its toolkits for opinion mining and sentiment analysis. In *2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, pages 1–4.

Vrandečić, D. (2021). Building a multilingual wikipedia. *Commun. ACM*, 64(4):38–41, March.

Glossary

alliteration The same sound of the stressed syllables within a word group (Glück, 2016). 5, 6, 8

anaphora Repetition of the initial word in successive sentences or lines (Berner, 2011). 5, 7, 10

anticlimax Gradatio downwards (sequence of steps) downwards (anti: against, climax: ladder, stairs) (Berner, 2011). 7

antimetabole Repetition of words in two sentences of the same structure in reverse order (e.g., We do not live to eat, but we eat to live) (Dudenredaktion, a). 2, 5, 8

aphaeresis Removal, disappearance of a vowel in the initial sound of a word (Berner, 2011). 5

archaism Antiquity; use of obsolete words or phrases, also called: antiquitas (Berner, 2011). 7

chiasmus Crossing over of syntactically or semantically corresponding parts of sentences ("art is long, and short is our life" (Berner, 2011). 2

chiffre Image or word compounds take on a meaning set by the author, regardless of their original meaning; [...] "city" as a cipher for hopelessness in some expressionist poets; occasionally called an absolute metaphor (Berner, 2011). 7

climax Ladder, staircase; an increase of at least three parts from the weaker to the stronger expression, sometimes also called gradation [...] (Berner, 2011). 2

epanaphora Another word for anaphora. 2

epiphora Reversal of anaphora, repetition of one or more words at the end of a sentence or verse (Berner, 2011). 2

gradatio Gradual increase; sequence of stages, generic term for climax and anticlimax (Berner, 2011). 2, 7, 10

incrementum A series of words in the same domain where subsequent words mark an increase on a semantic scale (O'Reilly et al., 2018). 2

irony Pretense, mockery; feigned ignorance, pretense; the opposite of what is said is meant, irony is context-dependent or arises from the contradiction to the person and situation of the person using it; it has an alibi function, if one thereby hides one's true opinion, then the contradiction to the context is only recognizable to insiders; however, it can also have a mocking and aggressive character, if one presents the value judgment corresponding to the listener's expectation in the form of ironic praise (Berner, 2011). 2

isocolon Identical words of almost the same number of syllables are placed in equivalent sentences/-parts of sentences; today, the term parallelism is consistently used instead. (Berner, 2011). 4

malapropism [deliberately] incorrect choice of words, in which a word is replaced by one that is phonologically similar but semantically (very) different (Dudenredaktion, b). 7

metaphor To transfer to something; a figure of figurative speech, the use of a word not in its proper (lexical) but (non-proper) transferred sense; the transfer of a word into a contextual image while retaining a similarity. A metaphor is a word in a context by which it is determined to mean something other than it means (Berner, 2011). 1, 2, 5

parallelism Repetition of the same syntactic construction in at least two or three successive clauses, often associated with an intensification or growing members (Berner, 2011). 4, 5

pejoration In linguistics, a word whose meaning has deteriorated in the course of development (Berner, 2011). 7

rhetorical question Rhetorical figure (of thought); question to which the speaker does not expect an answer and in which he or she dresses a request or statement. When the speaker wants to be particularly forceful and emotional (Glück, 2016). 2, 5

syllepsis Form of syntactic shortening in which a predicate verb is related to two or more grammatically identical but semantically different parts of a sentence, also called oblique parenthesis, since disparate things are brought together in a formal but stylistically oblique parenthesis (Berner, 2011). 6