

Ontology of Rhetorical Figures for Serbian

Miljana Mladenović¹ and Jelena Mitrović²

¹ Faculty of Mathematics, University of Belgrade, Serbia

{ml.miljana, jmitrovic}@gmail.com

www.matf.bg.ac.rs

² Faculty of Philology, University of Belgrade, Serbia

www.fil.bg.ac.rs

Abstract. The paper presents *RetFig*, a formal domain ontology of rhetorical figures for Serbian. This ontology is one of the necessary steps in developing tools for Natural Language Processing in the Serbian language, especially for tools pertinent to discourse analysis, sentiment analysis and opinion mining. The *RetFig* ontology was developed taking into account a plethora of rhetorical figures in the morphologically rich Serbian language, as well as in regard to various classifications of rhetorical figures that exist. We propose a system of linguistic classes and properties that are best suited for this ontology, as well as some of the possible usages for this particular ontology of rhetorical figures.

Keywords: domain ontology, rhetorical figures, Semantic web.

1 Introduction

Natural language texts are not always "flat" with unique, ordinary, untwisted literal meaning. On the contrary, texts written in a natural language almost always have more than one meaning, due to the usage of various linguistic operations over words, phrases, sentences, et cetera. Without taking these facts into consideration, we can get incomplete and imprecise results in some NLP tasks. This especially holds true in areas of opinion mining, sentiment analysis and discourse analysis. For example, if we say "He is as fast as light", this statement will be marked as a positive opinion statement. On the other hand, if we say "He is as fast as a turtle", opinion mining techniques will not show the correct result unless we include the process of detection of rhetorical figures. Our first task, in this direction, is to create the very first formal and comprehensive domain ontology of rhetorical figures in Serbian that will lead us, primarily, towards an ontology based semantic tool for annotation of rhetorical figures and implementations in other NLP tasks.

2 Related Work

Rhetoric is a means of spoken or written communication that we use in order to influence our listeners or our readers in a special way. Rhetorical figures (rhetorical devices, stylistic figures or figures of speech) have been a subject of research since ancient times in Aristotles major work, *Rhetoric*, it was pointed out that: "Rhetorician is someone

who is always able to see what is persuasive". Furthermore, the first rhetorical classification originated from Latin. It is known as "quadripartita ratio" and it describes four fundamental rhetorical operations on linguistics elements: addition, omission, permutation and transposition. Classical rhetoricians claimed that for any text taken as a literal model, all figures could be obtained with a combination of the four fundamental rhetorical operations by application on different linguistic levels: word forms, phrases, sentences, paragraphs, texts, etc. This kind of classification had been applied by Peacham [1] from 16th century on schemates rhetorical. He used rhetorical operations: repetition, omission, separation and conjunction. Also, it was the basis for later research. In that regard, Morris [2] created a (semio-) syntactic twodimensional classification table made of: linguistic operations and linguistic levels. Similarly, Durand introduced linguistic elements relationships like: identity, similarity, difference, opposition and false homologies [3]. Nowadays, we meet different classification systems. Harris¹ classifies rhetorical figures into three groups: "those involving emphasis; those involving physical organization, transition, and disposition; and those involving decoration." Sutcliffe² gives us a classification into six categories: figures of grammar, figures of meaning, figures of comparison, figures of parenthesis, figures of repetition and figures of rhetoric. Another classification is made by Schwartz³: figures of speech, sounds and other rhetorical devices. One of the most comprehensive researches on rhetorical figures can be found in the Inkpot⁴ Rhetfig project. According to Kelly et al. [4], the members of Inkpot group, there are four kinds of conceptual classifications of rhetorical figures. The first kind of classification represents the basic classification based on rhetorical characteristics, the classification into: *tropes*, *schemes* and *chroma*. The second classification is based on linguistic characteristics of the figures. Harris and DiMarco [5] marked it as a "linguistic domain" classification where domains are branches of linguistics. The third kind of conceptual classifications of rhetorical figures refers to different linguistic techniques used in rhetorical figures creation: *repetition*, *omission*, *series*, *identity*, *similarity*, *symmetry* and *opposition* applied over letters, words, clauses, phrases, sentences, etc. The fourth kind of conceptual classification is based on the generalization-specialization relationship of certain rhetorical figures and groups of figures. Evidently, many different methods for classification of rhetorical figures exist. Some of them are made from the perspective of rhetoric, some are made from the perspective of linguistics, and the others took both of those approaches into account. In order to create an ontology, we must consider all aspects of rhetorical figures research.

3 Building the Ontology of Rhetorical Figures for Serbian

Keeping in mind the complex and modular approach to building an ontology, which by Devedzić [6] includes: gathering and organizing of domain knowledge, defining usage, the range of validity and granularity in ontology, building the taxonomy, defining

¹ <http://www.virtualsalt.com/rhetoric.htm>

² <http://opundo.com/figures.php>

³ <http://cla.calpoly.edu/~dschwart/teaching.html>

⁴ http://create.uwaterloo.ca/matt/inkpot/projects/rhetorical_about.html

relations, restrictions and rules over ontology entities, we divided our work into two phases. First, we collected rhetorical figures and their examples in order to create domain knowledge of rhetorical figures and a solid basis for further procedures in the process of building of our ontology. Second, we developed a formal domain ontology of rhetorical figures for Serbian and prepared it for further usage.

3.1 Creating a Collection of Rhetorical Figures in Serbian

In the process of gathering and organizing domain knowledge, the first step was to create a database structure for collecting rhetorical figures [7]. It contains information about: rhetorical figure name in Serbian, name that is referred to corresponding rhetorical figure in English⁵, a definition or description, etymology of the name and additional notice. Also, three types of classifications according to rhetorical types, linguistic types and linguistic operations were introduced. We have developed and installed a web application⁶ for maintaining and serialization of the database (*RetFig*). In the process of acquiring data about rhetorical figures, we searched novels, poems and journal texts in order to find examples of all of the relevant figures. As the Corpus of contemporary Serbian language mostly consists of daily newspaper articles, we needed to find relevant texts elsewhere. We marked 98 distinct rhetorical figures and manually classified them into 4 rhetorical types: *figures of pronunciation* (figure naglašavanja), *figures of meaning tropes* (figure zamene značenja tropi), *figures of construction* (figure konstrukcije) and *figures of thoughts* (figure širenja i sužavanja misli). Typical representatives of the group of figures of construction are: aphaeresis, apocope, diatesis, ellipsis, etc; representatives of the group of figures of pronunciation are: alliteration, anaphora, paromoiosis, epistrophe, etc; of the group of tropes are metaphor, metonymy, oxymoron, simile, etc.; of the group of figures of thoughts: antitheton, auxesis, climax, paradox, etc. All figures are also divided into five linguistic categories. If the linguistic elements participating in the creation of a rhetorical figure are letters or groups of letters or syllables, we are talking about a group of *phonological rhetorical figures*. If a rhetorical figure is created using Inflectional forms of a word, or a word formation, that figure belongs to the *morphological* group. If a rhetorical figure changes ordinary linguistic order of words in a sentence or if it changes lexical categories of some words, if it adds or omits parts of a sentence, that figure belongs to the *syntax* group. In the case when figures are used to change the literal meaning of a sentence, they belong to the *semantic* group. When a change of literal meanings spreads over the context of more sentences, we are talking about a *pragmatic* group. At last, every rhetorical figure must also be defined by linguistic operations over linguistic elements. We use linguistic operations of *addition, omission, repetition, transposition, joining, separation* and *symmetry*. The RetFig XML file can be downloaded and used locally from the web application address.

3.2 Creation of the RetFig Ontology

Ontology of rhetorical figures in Serbian (The RetFig ontology) is meant to have the following roles: to represent a formal domain ontology that unambiguously describes

⁵ <http://rhetfig.appspot.com/>

⁶ <http://resursi.mmiljana.com/MemberZone/RetFig.aspx/>

and defines rhetorical figures in Serbian; to be shared and merged with other linguistic resources and ontologies, such as Serbian WordNet (SWN) [8], Princeton WordNet [9] and Suggested Upper Merged Ontology (SUMO) [10]; to represent the basis upon which a task ontology will be built and used in processes of ontological annotation of rhetorical figures in Serbian. We have decided to use the *top-down* modelling technique. The RetFig ontology was created in Protege 4.2., the free, open source ontology editor and knowledge-based framework, using of OWL 2 Web Ontology Language. RetFig ontology is a domain ontology filled manually. Its growth is not intensive and will depend on the instantiation of new types of figures.

3.3 Building a Taxonomy

As the term "rhetorical figure" is used equally in the fields of rhetoric and linguistics, we have primarily defined two top-concepts: the *RhetoricalEntity* ("RetorickiEntitet") and the *LinguisticEntity* ("LingvistickiEntitet"). The concept *RhetoricalFigure* ("RetorickaFigura") is defined as both a rhetorical and a linguistic concept. On a lower level, the rhetorical concept is represented by concepts: *RhetoricalGroup* ("RetorickaGrupa") and *RhetoricalFigure* ("RetorickaFigura"), while the linguistic concept is represented by concepts: *LinguisticObject* ("LingvistickiObjekat"), *LinguisticRange* ("LingvistickiOpseg"), *LinguisticGroup* ("LingvistickaGrupa"), *LinguisticPosition* ("LingvistickaPozicija"), *LinguisticElement* ("LingvistickiElement") and *RhetoricalFigure* ("RetorickaFigura") (Figure 1).

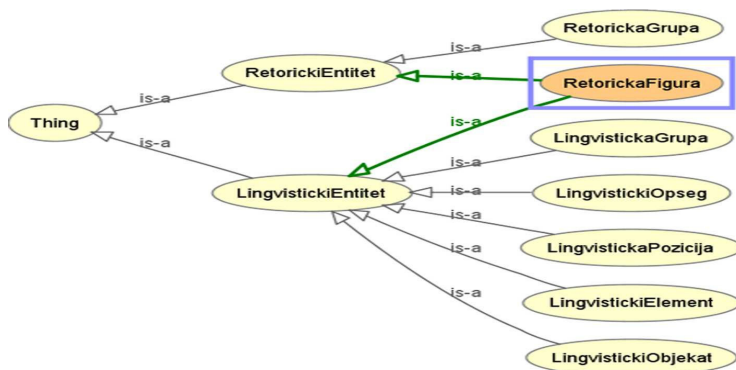


Fig. 1. Taxonomy of linguistic and rhetorical concepts

Each rhetorical figure in a text is characterized by the scope or the range of the text (the context) in which it appears. Looking from the inside out, the scope can be: a word, a phrase, a sentence, a verse, a strophe, a paragraph. Inside of such *linguistic scope*, we defined a *linguistic object* whose transformation via *linguistic operations* leads us to the structure that can be recognized as a certain *rhetorical figure*. *Linguistic object* can be a word, a phrase, a verse or a sentence. Transformation processes are either done over the

entire *linguistic object* or over a part of that object. In that regard, we defined *linguistic element* as part of *linguistic object* that is being transformed. If the *linguistic operation* is being performed over the entire object of transformation the *linguistic object* and the *linguistic element* are identical. Otherwise *linguistic element* is smaller than the *linguistic object*. In Figure 2 an example of rhetorical figure Aphaeresis (afereza) detection by *RetFig* ontology is shown. The given example is taken from Shakespeares "King Lear" — "The King hath cause to plain."

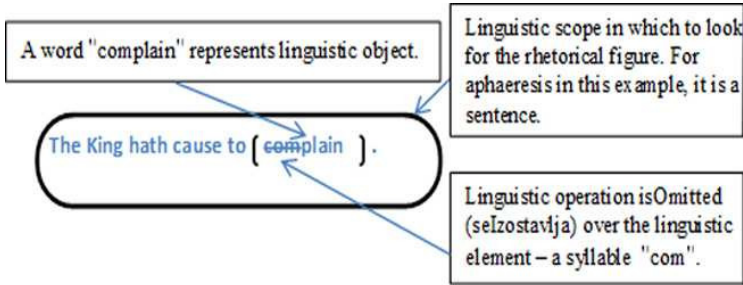


Fig. 2. Mutual relationship of linguistic scope, object and element

Analysis of processes of rhetorical figures creation has shown that mutual relationship between *linguistic objects* and *linguistic elements* differentiates the figures themselves. For example, if a *linguistic object* is a word, *linguistic element* is a letter, and a *linguistic operation* is "letter omission", then we could have: rhetorical figure aphaeresis (afereza), if an omitted letter is the first letter in the word; rhetorical figure apocope (apokopa), if an omitted letter is the last letter in the word; rhetorical figure syncope (sinkopa), if an omitted letter is not in the first or in the last position in the word to which it belongs. Because of that, positional relation between the *linguistic element* and the *linguistic object* is important, therefore, we introduce the concept: *Linguistic position* ("LingvistickaPozicija"), in order to define the position in which a *linguistic element* appears inside of a *linguistic object*.

Linguistic operations are defined in the ontology as relations that connect instances of *RhetoricalFigure* ("RetorickaFigura") class (Domain) and instances of *LinguisticElements* ("LingvistickiElement") class (Range). The division of relations in the *RetFig* ontology is to the following relations: *addition*, *omission*, *repetition*, *trans-position*, *joining*, *separation* and *symmetry* at the ObjectProperty level. Also, SubObjectProperty levels are defined.

3.4 Inserting Attributes and Individuals

The most important set of members of the *RetFig* ontology is the *RhetoricalFigure* ("RetorickaFigura") set of Individuals. They represent rhetorical figures themselves and these Individuals are formally defined to be uniquely identified. For each Individual, the rhetorical and linguistic groups it belongs to are defined, the linguistic scopes,

objects, elements and linguistic operations used for the creation of the said rhetorical figure. Conclusively, each rhetorical figure was appointed with its appropriate annotation: *comment* a short definition of the rhetorical figure, *seeAlso* information about the name of the rhetorical figure in English (keeping in mind the goal of mapping to linguistic ontologies [11] in English) and the alternative name of the same rhetorical figure in Serbian. This naming principle has been chosen to allow easier usage of this ontology for Serbian annotators but also to keep the possibility of alignment. Each *Rhetorical-Figure* ("RetorickaFigura") class member is declared as it is shown in the declaration of rhetorical figure Dysphemismus (Disfemizam):

```
<owl:NamedIndividual rdf:about="&ont;DISFEMIZAM">
<rdf:type rdf:resource="&ont;RetorickaFigura"/>
  <ont:naziv rdf:datatype="&xsd:string">DISFEMIZAM
  </ont:naziv>
  <rdfs:comment>Namerno koriscenje ruznijeg, ostrijeg
    izraza umesto normalnog.</rdfs:comment>
  <rdfs:seeAlso xml:lang="en">dysphemismus</rdfs:seeAlso>
  <rdfs:seeAlso xml:lang="sr">KAKOFEMIZAM</rdfs:seeAlso>
  <ont:jeNaPoziciji rdf:resource="&ont;CELINA"/>
  <ont:jeRetorickaGrupa
    rdf:resource="&ont;FIGURE_ZAMENE_ZNACENJA-TROPI"/>
  <ont:seZamenjujeDrugimElementomJacegZnacjenja
    rdf:resource="&ont;FRAZA-LELEMENT"/>
  <ont:jeNadObjektom rdf:resource="&ont;FRAZA-LOBJEKAT"/>
  <ont:seZamenjujeDrugimElementomJacegZnacjenja
    rdf:resource="&ont;REC-LELEMENT"/>
  <ont:jeNadObjektom rdf:resource="&ont;REC-LOBJEKAT"/>
  <ont:jeNadOpsegom rdf:resource="&ont;RECENICA"/>
  <ont:jeLingvistickaGrupa rdf:resource="&ont;SEMANTIKA"/>
  <ont:jeNadOpsegom rdf:resource="&ont;STIH"/>
</owl:NamedIndividual>
```

From the example given above, we can see that this figure represents usage of an intentionally harsh word or expression instead of an expected, or a polite one. We also find that the name of this figure in English is *Dysphemismus*, and that there is also an alternative name for this figure in Serbian *Kakofemizam* (*Kakophemismus*). *Disfemizam* is a rhetorical figure from the group named tropi (tropes), it is a subject of research in the area of linguistics called Semantika (Semantics). It can be found inside a sentence or a verse (*linguistic scope*) and it is formed by replacing the existing phrase or a word (*linguistic object/element*) in its entirety (*linguistic position*) by a different phrase or a phrase or a word of a stronger meaning (*linguistic operation* – "seZamenjujeDrugimElementomJacegZnacjenja").

4 RetFig Ontology Testing

RetFig ontology is meant to give a couple of significant answers. First, it is prepared for ontological annotation of rhetorical figures in Serbian. In this regard, for pre-selected individuals for *linguistic scope* and/or *linguistic object* of observation, this ontology

has to give a candidate or candidates for a certain rhetorical figure. For example, if the pre-selected individual is "REC" ("word") for the *linguistic scope*, there is no sense in expecting for the rhetorical figure Dysphemismus (Disfemizam) to be annotated, but it is expected that figures like aphesis (afereza), diarexis (dijareza), prothesis (proteza) et cetera will appear. Second, it gives us an insight into the rhetorical figures used in a particular text. For example, if we determine that the analysis of a certain text shows a frequent loss of letters in words, mapping onto the ontological relation "*selzostavlja*" (isOmitted) gives us a set of rhetorical figures that are formed that way, by omission of letters. Those figures are: aphaeresis (afereza), syncope (sinkopa), apocope (apokopa) and ecthipsis (elizija). The SPARQL (recursive acronym for SPARQL Protocol and RDF Query Language) queries that give the answers to both of the mentioned tasks are represented in Figure 3. Moreover, the RetFig ontology will also be able to give answers about the statistical data regarding the annotated rhetorical figures. Each rhetorical figure in the RetFig ontology is defined by a finite set of RDF triples that uniquely describe that figure. With SPARQL queries, figures can be selected individually or in groups, which is the purpose of the ontology. The RetFig.owl ontology can be downloaded from the address of the web application.

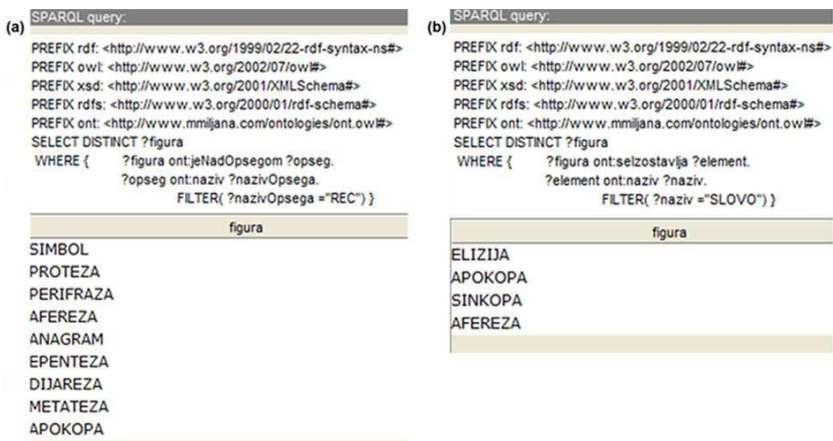


Fig. 3. (a) SPARQL query used to find the rhetorical figures generated over words (b) Query for finding rhetorical figures formed by omission of letters in words

5 Conclusion and Future Work

In this paper we have proposed the *RetFig* ontology that describes and defines linguistic entities and relations used in generation of rhetorical figures in Serbian. We have collected linguistic knowledge about rhetorical figures and their examples in the Serbian language. Furthermore, we have shown that the *RetFig* ontology can be our starting point in the future process of annotation of rhetorical figures in Serbian. Our future work will include connecting RetFig ontology with SUMO, Adimen-SUMO [12] and with the SWN (especially after recent enhancements of SWN which include transformation

to RDF) as a part of the process of development of the RetFig application ontology in order to get a usable, ontology based semantic annotation tool for the rhetorical figures in Serbian. Metaphors are the focus of our research in the scope of developing the *SimNet* resource for automatic annotation of metaphors, which will further improve the results of semantic analysis processes aided by the tools we are developing. Implementation of annotation methods for figurative types will also be leaning on extensive research done in the field of MWEs exploration [13] so far.

References

1. Peacham, H.: *The Garden of Eloquence*. Perseus Digital Library (1593), <http://rhetoric.byu.edu/primary%20texts/Peacham.htm>
2. Sloane, T.O.: *Encyclopedia of Rhetoric*. Oxford University Press, New York (2001)
3. Durand, J.: *Rhetorical Figures in the Advertising Image*. In: *Marketing and Semiotics: New Directions in the Study of Signs for Sale*, pp. 295–318. Walter de Gruyter, New York (1987)
4. Kelly, R.A., Abbott, A.N., Harris, A.R., DiMarco, C., Cheriton, D.: *Toward an ontology of rhetorical figures*. In: *SIGDOC 2010 Proceedings of the 28th ACM International Conference on Design of Communication*, pp. 123–130 (2010)
5. Harris, R.A., DiMarco, C.: *Constructing a rhetorical figuration ontology*. Symposium on Persuasive Technology and Digital Behavior Intervention, Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB), Edinburgh (2009)
6. Devedžić, V.: *Understanding Ontological Engineering*. *Communications of the ACM* 45(4), 136–144 (2002)
7. Tartalja, I.: *Teorija književnosti*. Zavod za udžbenike i nastavna sredstva, Beograd (2003)
8. Krstev, C., Lažetić, G.P., Vitas, D., Obradović, I.: *Using Textual and Lexical Resources in Developing Serbian Wordnet*. *Romanian Journal of Information Science and Technology* 7(1-2), 147–161 (2004)
9. Fellbaum, C.: *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge (1998)
10. Niles, I., Pease, A.: *Linking lexicons and ontologies: Mapping wordnet to the suggested upper merged ontology*. In: *Proceedings of the IEEE International Conference on Information and Knowledge Engineering*, pp. 412–416. IEEE (2003)
11. Bateman, J.A.: *Ontologies of Language and Language Processing*. In: *Theory and Applications of Ontology: Computer Applications*, pp. 393–409. Springer, Netherlands (2010)
12. Àlviz, J., Lucio, P., Rigau, G.: *Adimen-SUMO: Reengineering an Ontology for First-Order Reasoning*. *International Journal on Semantic Web and Information Systems (IJSWIS)* 8(4) (2012)
13. Sag, I.A., Baldwin, T., Bond, F., Copestake, A., Flickinger, D.: *Multiword Expressions: A Pain in the Neck for NLP*. In: Gelbukh, A. (ed.) *CICLing 2002*. LNCS, vol. 2276, pp. 1–15. Springer, Heidelberg (2002)