# Ontology-based Recognition of Rhetorical Figures

UDC 811.163.41'322.2
DOI 10.18485/infotheca.2016.16.1_2.2

Miljana Mladenović
ml.miljana@gmail.com
*University of Belgrade, Faculty of Mathematics*

**ABSTRACT:** Automatic recognition of rhetorical figures (similes, irony, sarcasm, humor, metaphors, etc.) is increasingly used in natural language processing tasks, primarily to improve sentiment classification, machine translation, but also for analysis of linguistic structures on different levels. In this paper, it is proposed a method of automatic recognition and classification of rhetorical figures from the group of tropes that uses ontological inference rules in an ontology based on Serbian WordNet (SWN). A binary classification method was carried out on the rhetorical figure simile and evaluated by ROC curve ($AUC = 0.696$) which indicates that it can be successfully used in solving these types of tasks. It is also proposed a semi-automatic ontology learning method, for further learning of SWN ontology, by increasing the number and the type of relationships that can assist in the detection of figurative language in the texts in Serbian.
**KEYWORDS:** rhetorical figures, simile, ontology-based classification, WordNet

**PAPER SUBMITTED:** 25 February 2016
**PAPER ACCEPTED:** 09 May 2016

## 1 Introduction

Automatic recognition of rhetorical figures (figurative language) in text and its annotation are not new fields of natural language processing. One of the first pieces of research in the field of automatic identification and interpretation of rhetorical figures, which deals with the identification of *anaphora* (anaphora resolution), was given in the 1964 paper "A question-answering system for high school algebra word problems" (Bobrow, 1964), whereas in the paper "met*: A method for discriminating metaphor and metonymy by computer" (Fass, 1991) subjects of automatic recognition were *metaphor* and *metonymy*. The literature describes methods of automatic recognition of *metaphors* (Mason, 2004; Hardie et al., 2007; Koller et al., 2008; Shutova et al., 2013), *anaphora* (Mitkov, 2002; Poesio and Artstein, 2010), *metonymy* (Farkas et al., 2007; Leveling, 2007; Nicolae et al., 2007), *irony* (Carvalho et al., 2009; Veale, 2012), *sarcasm* (Tsur et al., 2010; González-Ibáñez et al., 2011) etc. However, basic sentiment analysis (SA) methods mainly explore literal meaning of the text not taking into account figurative language.

In (Hao and Veale, 2010) authors analyzed figurative language and showed that certain rhetorical figures (e.g. *irony*) in text act as sentiment polarity *modifiers*, whose role in changing the sentiment polarity of words or phrases appears in the scope of the role of *valence shifters* presented in the paper "Sentiment Classification of Movie Reviews Using Contextual Valence Shifters" (Kennedy and Inkpen, 2006). In general, the *valence shifters* are words and phrases that reduce, increase or completely change the polarity of emotion words or phrases that they appear with (e.g. sentiment polarity of the word *happy* is changed by the phrase *not at all*). A certain group of rhetorical figures, which can affect the sentiment polarity of words or phrases in text, act in a similar way. By definition, figures *irony* and *sarcasm* change polarity, *dysphemism* and *hyperbole* enhance the existing level of sentiment expression, while *litotes* and *euphemism* reduce that level. *Metaphor*, *metonymy*, *oxymoron*, and *simile* have more complex mechanisms of action in both directions of changing the sentiment strength and polarity.

The process of automatic or semi-automatic identification of rhetorical figures in text can improve the sentiment analysis, machine translation and other natural language processing tasks. For example, the traditional sentiment polarity classification system (Pang et al., 2002) would interpret sentences "*He is quick as a rabbit*" and "*He is quick as a snail*" in the same way. The inclusion of figurative language recognition would show that the second sentence can be noted as ironical, which would invert its sentiment polarity compared to the literal meaning. Research published in (Reyes and Rosso, 2012b) has shown that the accuracy of sentiment classification can be significantly improved (from 54% to a maximum of 89.05%) when predictors of identifying figurative language were included, compared to a set of predictors that treated the text in the literal sense. In (Rentoumi et al., 2010) a sentiment analysis task has been improved by integration of a machine learning (ML) method with the method based on the rules which recognize the use of figurative language. The paper (Williams et al., 2015) discussed the role of idioms in sentiment analysis. In two experiments, the authors showed improvement of classification, measured by F-measure (from 45% to 64%, and from 46% to 61%), when idioms that carry figurative meaning were used as predictors of classification. In the paper (Carvalho et al., 2011) results of the sentiment analysis on comments published on the web portals of newspaper show that 11% of the observed set of comments would have been incorrectly labeled as positive, had the analysis and identification of rhetorical figure *irony* not been used. Therefore, sentiment analysis of texts using figurative language is a new challenge in the field of natural language processing. In 2015, for the first time on a global scale, a task of this kind was set — sentiment analysis of figurative language in Twitter (International Workshop on Semantic Evaluation — SemEval-2015).[1]

Research in the field of processing of figurative language moves in two directions:

- to improve sentiment classification methods (Reyes and Rosso, 2012b; Rentoumi et al., 2010);
- to better understand the structure of a language — in (Veale and Hao, 2009) comparative analysis of ironic similes in English and Chinese languages showed that *irony* is a linguistic and also a cultural phenomenon, because a set of ironic simile expressions used in the paper as similes that are used in English can be applied in the Chinese language only in amount of 3–4%.

In the study of figurative language semantic networks such as WordNet (Mason, 2004; Barbieri et al., 2015), ontology (Harris and Di Marco, 2009; Kelly et al., 2010), lexical resources such as corpora (Mason, 2004; Hao and Veale, 2010; Reyes and Rosso, 2012a), specialized dictionaries of emoticons and punctuation (Carvalho et al., 2009; González-Ibáñez et al., 2011; Barbieri et al., 2015) and lexicons like SentiWordNet (Rentoumi et al., 2010; González-Ibáñez et al., 2011; Barbieri et al., 2015) play an important role.

In this paper we propose a method of automatic recognition of rhetorical figures belonging to the group of tropes, using the rules defined in an ontology SWN which is based on the Serbian semantic network WordNet. Afterwards, we propose an ontology learning method in terms of increasing the number and type of relations in SWN ontology that can help in the detection of such figures in texts in Serbian.

## 2 Methods for automatic recognition of rhetorical figures

Automatic recognition of rhetorical figures in text depends on the nature of a figure and its language structure. In order to build an effective system of automatic recognition of figurative language, it is necessary, first of all, to formally define and describe the rhetorical figures. The process of building the first formal domain ontology of rhetorical figures for Serbian (*RetFig*) is shown in the paper "Ontology of Rhetorical Figures for Serbian" (Mladenović and Mitrović, 2013). The *RetFig* ontology describes 98 rhetorical figures. For each of them, rhetorical and linguistic group

[1] http://alt.qcri.org/semeval2015/index.php?id=tasks (Retrieved on March, 23th 2016.)

they belong to were defined, the linguistic scope, objects and elements, mutual relationship of objects and elements, as well as linguistic operations involved in the process of their creation. All figures are divided into four groups: figures of pronunciation, figures of construction, figures of twisted meaning — tropes and figures of thought. Figures of pronunciation are based on the influence of certain letters in text (phonemes in speech). The repetition of a certain letter or group of letters in text (phonemes or groups of phonemes in speech), their omission or insertion in unexpected places, imitating certain sounds and noises from nature, affects the increase or decreases the importance of linguistic structures over which they run. This type of figures has no effect on the meaning of text and does not change it, but only emphasizes its basic meaning. Figures of construction are formed by changing the usual arrangement of words in a sentence or in another larger text portion (verse, for example). This kind of figures also does not change the basic meaning of the language structures that are being built. In contrast to these groups, figures of twisted meaning — tropes and figures of thought change the basic meaning of a word or larger text portion in their scope.

Automatic recognition of rhetorical figures of pronunciation and construction in text can be carried out using regular expressions as applied to the works (Gawryjolek et al., 2009; Hromada, 2011). For example, the figure *antimetabole*, belonging to the group of figures of construction, is defined by the repetition of a word or phrase which stands in the first part of the sentence and also in the second, but in reverse order.[2] The paper (Gawryjolek et al., 2009) describes a model of automatic recognition of such kind of figures, as well as the appropriate tool for automatic annotation (Java Annotation Tool of Rhetoric — Jantor). It gives the syntax pattern for automatic recognition of *antimetabole* in the form $[W_a] \ldots [W_b] \ldots [W_b] \ldots [W_a]$ in which it finds the occurrence of this figure in a sentence, where $W_a$ and $W_b$ are words or phrases in the sentence which in the second part of the sentence appear in reverse order compared to the first. Results of Jantor show that the rhetorical figures of pronunciation and construction can be successfully recognized, because it is possible to define regular expressions that define them unambiguously. Both classes of figures are based on syntactic and morphological operations within the grammatical rules of a natural language. However, the structure of rhetorical figures in classes of figures of twisted meaning — tropes and figures of thought cannot be defined and recognized by syntactic or morphological patterns.

In processes of automatic recognition and annotation of rhetorical figures that can change meaning of the text ML methods are successfully used. In the works that dealt with the *metonymy* recognition, supervised ML methods, such as making lists (Markert and Nissim, 2002), the maximum entropy (Farkas et al., 2007), k-

[2] For example: "*We eat to live, not live to eat.*", "*Write as you speak, speak as you write.*"

nearest neighbors (Leveling, 2007)[3], logistic regression (Nicolae et al., 2007) are used. In the paper (Tsur et al., 2010) a method of semi-supervised learning to identify and classify sarcastic tweets and comments from the site e-commerce Amazon is proposed. Support vector machines and logistic regression were used in the process of classifying sarcastic tweets and both positively and negatively sentiment polarized tweets without *sarcasm* (González-Ibáñez et al., 2011).

Besides by ML methods, *irony* and *sarcasm* can be successfully detected and by other techniques. In (Carvalho et al., 2009) authors created eight forms (e.g. $P_{laugh} = (LOL|AH|EMO+)$, $P_{quote} = (ADJ_{pos}|N_{pos})\{1,2\})$ which indicate the existence of structural *irony* in comments published on news portals in Portuguese language. The crowdsourcing method proposed by (Filatova, 2012) is used in the detection of *irony* and *sarcasm* in Amazon product reviews, while author (Veale, 2012) defined automatic extraction method of semantic knowledge in examples of using the *simile* figure in order to discover examples of ironic comparisons. In (Mitrović et al., 2015; Mladenović et al., 2016a) authors used a crowdsourcing method for detecting *similes* comparing it with the algorithm proposed for automatic extraction of simile candidates based on frequency of occurrence in an annotated corpus.

In this paper, in the process of detection of rhetorical figures belonging to the group of tropes we propose usage of ontological reasoning of the ontology derived from the semantic network Serbian WordNet (SWN).[4]

## 3 Ontology-based recognition of rhetorical figures tropes

Ontology is one of the forms of knowledge representation. If an ontology is described by a formal language and stored in a computer-readable format, it is a formal ontology. Understood as a "specification of shared conceptualization" (Gruber, 1993), ontology indicates the kind of knowledge that can be transferred, exchanged and used. According to (Devedžić, 2010), the main purpose of ontology is to be shared and reused by various intelligent agents and applications. Depending on which part of reality it describes, an ontology can be:

- top level ontology — when it describes the general concepts and the knowledge that it represents is comprehensive, systematized and applicable in a wide range of applications;
- domain ontology — when it represents knowledge of a certain domain;
- task ontology or application ontology — contains only the knowledge necessary to carry out the given class tasks.

[3] A software package TiMBL, based on the implementation of the algorithm Memory-Based Learning which is derived from the method "k-nearest neighbor", is used.

[4] http://sm.jerteh.rs/MemberZone/eW3.aspx?id=miljanam

In this paper, we propose a method for recognition of rhetorical figures that belong to the class of tropes, which, by definition, are those figures whose role is based on changing the basic meaning of words or phrases within a scope represented by surrounding words, phrase, verse or a sentence. Tropes base their role on the following characteristics:

– replace the meanings of a word, substituting that word with another word or phrase (e.g. in the case of an utterance representing *irony* "*You're my best friend*", the word *worst* is substituted by the word *best*);
– assign multiple meanings to a word or a phrase (e.g. in the case of an utterance representing figure of *synecdoche* "*He got a roof over their heads*" a phrase *roof over their heads* can get the meaning of words: *home*, *house*, *safety*, *shelter* or their synonyms);
– generation of a new meaning of a word or a phrase (e.g. figure *oxymoron* combines the concepts of opposite meaning in the new term, for example, *virtual reality*).

In the case of tropes it is not possible to define the morphological or syntactic patterns for recognitions, but it is possible to determine what kind of semantic relations exists between participants who build a certain figure and on that basis to define rules of inference in an ontology obtained from the semantic network WordNet (Fellbaum, 1998). Using a SPARQL query against a domain ontology *RetFig* (Mladenović and Mitrović, 2013) we can get all tropes used in Serbian.

```
select distinct ?figura
where { ?figura ont:jeNadObjektom ?objekt.
?figura ont:jeRetorickaGrupa ?retGrupa.
?objekt ont:naziv ?nazivObjekta.
?retGrupa ont:naziv ?nazivRetGrupe.
FILTER (?nazivObjekta ="REC" &&
?nazivRetGrupe="FIGURE_TROPI")}
order by ?figura
```

The SPARQL query produced a collection of 26 rhetorical figures (Figure 1), which belong to the group of figures tropes (?nazivRetGrupe="FIGURE_TROPI") and are characterized by some form of change in the usual meaning of a word (?nazivObjekta="REC"). At the top of this part of the *RetFig* ontology is a class named Tropi (tropes). It contains three subclasses (VišestrukoZnačenje, IzmenjenoZnačenje, NovoZnačenje) which define the way in which is changing the usual meaning of the word on which a figure is applied (multiple meaning, changed meaning and a new meaning). The idea is, if it is possible, for each of these figures to be defined by rules (expressed using the Semantic Web Rule Language — SWRL)[5] over ontology SWN.

[5] Semantic Web Rule Language (SWRL) is the language used in Semantic Web for presenting formal logical expression, obtained by combining the features of OWL DL language

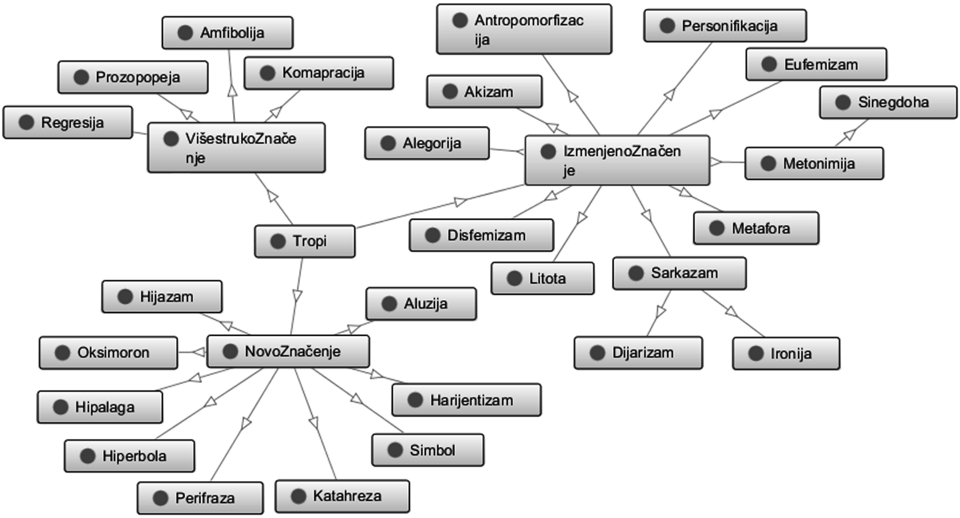


Figure 1: A part of taxonomic hierarchy of classes in an ontology *RetFig* concerning the figures of tropes

Explaining the case of figure *irony*, we will explain rules in the SWN ontology by which an occurrence of the corresponding figure in the text can be recognized. Verbal *irony* is the usage of words or phrases to say something opposite to the true meaning. The true meaning of the word is disguised and is opposite to the meaning of the used word or phrase. Examples of verbal irony are based on several types of semantic relations (Bagić, 2012), and commonly used forms in Serbian language based on the usage of an adjective instead its opposite (antonym) adjective, as in the examples:

- "*He's just brilliant!*" ("*The hidden meaning of the claim that someone is really stupid.*")
- "*See how skinny he is!*" ("*Hidden claim that someone is fat.*")

Another form consists of a noun and an adjective, but the adjective whose hidden meaning is opposite to the meaning of an adjective that is commonly associated with that noun.

- "*Fast as a turtle.*"

(Web Ontology Language for Description Logic) and RML language (Rule Markup Language).

– "*Brave as a rabbit!*"

The paper (Veale and Hao, 2009) used the term "*ironic comparisons*" to describe a specific type of *irony* derived from the figure simile. In previous examples, the adjective *fast* has a hidden meaning, the meaning of its antonym *slow* which is a natural feature of a noun *turtle*, while the adjective *brave* has a hidden meaning of the adjective *fearful*, which is a natural feature of a noun *rabbit*. Semantic relationships that exist in the ontology SWN and which can be used for generating candidates for the detection of figure *irony* are relations *specificOf* / *specifiedBy* and a relation *near_ antonym*. Relations *specificOf* / *specifiedBy* represent a pair of inverse relations (Mladenović et al., 2016a) which link an instance of a noun synset class with that instance of an adjective synset class which represents natural feature (is specific of) of the noun instance. The relation *near_ antonym* links two instances of an adjective synsets class which are direct antonyms mutually. Therefore, the rule for generating implicit relation representing *irony* is a relation between an instance of a noun synset class and an instance of adjective synset class. In ontology SWN it can be expressed as an OWL rule

$$\{?n : specifiedBy\ ?p1.\ ?p1 : near\_antonym\ ?p2\} \Rightarrow \{?n : Irony\ ?p2\}$$

or as SWRL rule

$$specifiedBy(?n, ?p1), near\_antonym(?p1, ?p2) \rightarrow Irony(?n, ?p2) \quad (1)$$

The result of inferencing by the rule (1) is equivalent to the result obtained by SPARQL query (Figure 2) in the ontology SWN.

Query result gives the candidates of figure *irony* in the form of a pair (adjective2, noun), for example, (*hrabar-brave*, *zec-rabbit*),(*brz-fast*, *kornjača-turtle*), (*brz-fast*, *puž-snail*), (*spor-slow*, *strela-arrow*), (*spor-slow*, *ideja-brainstorm*), (*lak-lightweight*, *slon-elephant*), (*vruć-hot*, *kamen-stone*), but also gives candidates of figure *simile* in the form of a pair (adjective1, noun), e.g. (*fearful*, *rabbit*), which is equivalent to OWL rule

$$\{?n : specifiedBy\ ?p1\} \Rightarrow \{?n : Simile\ ?p1\}$$

or

$$specifiedBy(?n, ?p1) \rightarrow Simile(?n, ?p1) \quad (2)$$

Query result in Figure 2 is analogous to the result of reasoning based on rules (1), but it can be observed that in natural language is not common to use some of the candidates for generating figure *irony*. Based on intuition, which at this point is not proven, it can be said that natural candidates for figure *irony* are

```
SPARQL query:

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX swn30: <http://www.mmiljana.com/swn30#>
SELECT ?adjective1 ?adjective2 ?noun
        WHERE { ?adjective1 swn30:near_antonym ?adjective2.
                ?adjective2 swn30:specificOf ?noun.
}
order by ?adjective1
```

| adjective1 | adjective2 | noun |
|---|---|---|
| hrabar | plašljiv | zec |
| brz | spor | kornjača |
| brz | spor | puž |
| spor | brz | strela |
| spor | brz | ideja |
| lak | težak | slon |
| vruć | hladan | kamen |

Figure 2: Examples of figures *irony* and *simile* obtained by SPARQL query in ontology SWN

those in which the sentiment polarity of the subject (?adjective1) in RDF triplets[6] (?adjective1 swn30:*near_antonym* ?adjective2) is positive. For example, more natural examples of irony are claims *fast as a turtle* or *lightweight as an elephant* than *slow as arrow* or *heavy as a feather*. Therefore, the query can be enhanced with an additional condition that take into account only those RDF triples in which the value of negative sentiment polarity of the subject is zero. Since each synset in the Serbian WordNet is labeled by a rate of positive and negative sentiment polarity (Mladenović and Mitrović, 2014) using the resource SentiWordNet (Baccianella et al., 2010), the SPARQL query can be extended to control the data property swn30:sentimentNegative and selected (filter) only those elements in which the value of that property is zero (Figure 3).

SWRL rule which includes an additional condition shown in Figure 3, is given below

$$\begin{aligned} specifiedBy(?n, ?p1), near\_antonym(?p1, ?p2), \\ sentiment\_negative(?p1, ?sent), swrlb : equal(?sent, 0) \ \rightarrow Irony(?n, ?p2) \end{aligned} \tag{3}$$

[6] RDF is a data model that uses a form of triplets (subject-predicate-object) to describe the semantic web resources, provide storage of data in graph databases and knowledge representation in ontological models.

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX swn30: <http://www.mmiljana.com/swn30#>
SELECT ?adjective1 ?adjective2 ?noun
            WHERE { ?adjective1 swn30:near_antonym ?adjective2.
                    ?adjective2 swn30:specificOf ?noun.
                    ?adjective1  swn30:sentimentNegative ?sent.
                    FILTER regex( str(?sent), "0,00000")
}
            ORDER by ?adjective1
```

| adjective1 | adjective2 | noun |
|---|---|---|
| hrabar | plašljiv | zec |
| brz | spor | kornjača |
| brz | spor | puž |
| vruć | hladan | kamen |

Figure 3: Improving search for examples of figure *irony*

Let us now consider the structure of rhetorical figure *oxymoron* which, by definition, represents merging of terms with opposite meanings into a new concept. Examples of *oxymoron* found in SWN ontology are:
primerRecvirtualna stvarnost (*virtual reality*), *glasna tišina* (*loud silence*), *luda pamet* (*crazy clever*), *živi fosil* (*living fossil*), *vatreni led* (*fire-ice*), etc. The rules for generating candidates for figure *oxymoron* in ontology SWN can be expressed as

$$\{?p1 : near\_antonym\ ?p2.\ ?p2 : derived\text{-}pos\ ?n\} \Rightarrow \{?p1 : oxymoron\ ?n\}$$
$$\{?p1 : near\_antonym\ ?p2.\ ?p2 : be\_in\_state\ ?n\} \Rightarrow \{?p1 : oxymoron\ ?n\}$$

or as SWRL rules:

$$near\_antonym(?p1, ?p2),\ derived\text{-}pos(?p2, ?n) \rightarrow oxymoron(?p1, ?n) \tag{4}$$
$$near\_antonym(?p1, ?p2),\ be\_in\_state(?p2, ?n) \rightarrow oxymoron(?p1, ?n) \tag{5}$$

The rules (4) and (5) can also be obtained by SPARQL query (Figure 4) in the SWN ontology. Query result gives candidates for generating an *oxymoron* figure in the form of a pair (adjective1, noun), for example (*glasan-loud*, *tišina-silence*). At Figure 4 we can see that RDF triples (adjective1, adjective2,noun) containing adjectives that are mutual antonyms become candidates for the instantiation of this figure when one of adjectives is in relationship by any of relations *derived-pos* and *be_in_state* with the corresponding noun. In SWN, relations *derived-pos* and *be_in_state* are lexical relations between synsets belonging to different part-of-speech (cross-part of speech relations — XPoS), in this case between adjectives and nouns (Koeva et al., 2008).

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX swn30: <http://www.mmiljana.com/swn30#>
SELECT ?adjective1 ?adjective2  ?noun
        WHERE {
        ?adjective1 swn30:near_antonym ?adjective2.
        {
            { ?adjective2 swn30:derived-pos ?noun.        }
             UNION
            { ?adjective2 swn30:be_in_state ?noun.        }
        }
        }
```

| adjective1 | adjective2 | noun |
|---|---|---|
| ružan | lep | lepota |
| virtualan | stvaran | stvarnost |
| spor | brz | brzina |
| vruć | hladan | temperatura |
| glasan | tih | tišina |

Figure 4: Examples of a figure oxymoron obtained by SPARQL query in ontology SWN

Beside the lexical-semantic relations, as an indicator of the existence of certain figures of tropes, an implicit relation *synonymy* can also be used. For example, we can find candidates for the figure *periphrasis* (*circumlocutions*) (Mitrović, 2014), when the term is described or replaced with more words using some essential properties of this concept (Figure 5). So the word *Paris* can be replaced by its synonyms existing in SWN: *Grad svetlosti* (*The City of Lights*), *Prestonica Francuske*, *Glavni grad Francuske*[7] (*The capital of France*), so pairs like (*Pariz — Grad svetlosti — The City of Lights*) become candidates for the instancing of this figure.

## 4 SWN Ontology Learning

In the previous section we described the rules defined in the SWN ontology which automatically generate candidates for some rhetorical figures belonging to group of figures tropes. However, the main problem in engineering an ontology, according to (Devedžić, 2010), is the ability of changing the ontology simultaneously with

[7] All examples extracted from SWN are presented in its original form, in Latin.

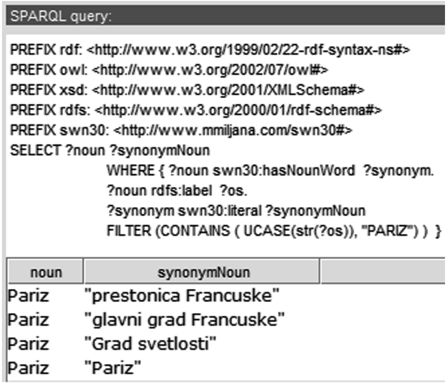


Figure 5: Examples of the figure *periphrasis* obtained by SPARQL query in SWN ontology

changing the knowledge in the domain. Ontology learning using automatic and semi-automatic methods of extraction and annotating new knowledge in the domain and integrating it with existing knowledge in the ontology, largely depends on the tools that are applied for that purpose. Since we use ontology SWN to recognize different rhetorical figures, it is very important to learn different forms of figurative speech from examples used in a natural language. Since that recognition, as we described in previous section, depends on relations in the SWN ontology (*specificOf*, *specifiedBy*, *near_ antonym*, *derived-pos*, *be_ in_ state* and *synonym*) and entities that are linked by them, by ensuring automatic or semi-automatic intensive and continuous increase of the number of these relations and linked entities will get the ontology to be more complete and therefore more effective in the task of recognition. Diagram of the SWN ontology learning is shown in Figure 6.

SWN ontology learning, based on a method of automatic enlargement of semantic network Serbian WordNet, is proposed in (Mladenović et al., 2016a). The proposed method uses the annotated part of the digital Corpus of the contemporary Serbian language[8] (Utvić, 2014) to generate relations *specificOf* and *specifiedBy* between synsets existing in SWN. In this paper we propose a semi-automated method of enlargement of SWN which uses Web tools to facilitate the connection of synsets using some of the lexical-semantic relations defined in the SWN. The user interface is shown in Figure 7 and represents a tool that can link two synsets, previously chosen by a user, by one of the available relations. On the left side of Figure 7 for the entered word (e.g. lep — beautiful), underneath appear all the meanings of the

[8] http://korpus.matf.bg.ac.rs/prezentacija/korpus.html

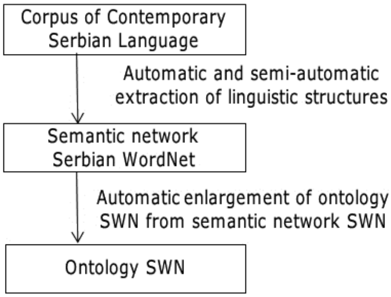


Figure 6: SWN ontology learning from Corpus of contemporary Serbian language

word that are defined in SWN. On the right side of the same figure for the entered word (e.g. slika — picture), also below, all its meanings defined in SWN are shown. The user selects the appropriate meaning of each of the entered word and connects them by choosing a relation from a list of relations given in the middle column. If the user selects a relation for which there is an inverse one (e.g. *SpecificOf*-*SpecifiedBy*) it is automatically generated and the corresponding inverse relation between synsets of concepts that are selected according to meaning.

## 5 Evaluation of ontology-based method for recognition of rhetorical figure *simile*

In order to evaluate recognition of the rhetorical figure *simile* in texts in Serbian, we carried out testing using the collection of various texts consisting of 10 digitized writings of various genres: children's songs, fairy tales, comedies, novels and essays, all listed in Appendix A. To this collection were added two datasets (named *news* and *films*) of articles which are used in (Mladenović et al., 2016b) as testing sets in sentiment analysis in Serbian, composed of film reviews and news from news portals in Serbian language (bars labeled as *vesti*-news and *filmovi*-movies in Figure 8). We built a tool that, with the help of regular expressions, parses the input text and extract linguistic structures in the form of candidates to identify a simile. We used regular expressions "РЕЧ *као* РЕЧ {1,2}", "РЕЧ *попут* РЕЧ {1,2}" (WORD *as* WORD {1,2}). From the extracted examples, we manually removed all those who are not in the form "РЕЧ *као* ИМЕНИЦА" (WORD *as* NOUN), "РЕЧ *као* ПРИДЕВ ИМЕНИЦА" and "РЕЧ *попут* ПРИДЕВ ИМЕНИЦА" (WORD *as* ADJECTIVE NOUN). Finally, we built a

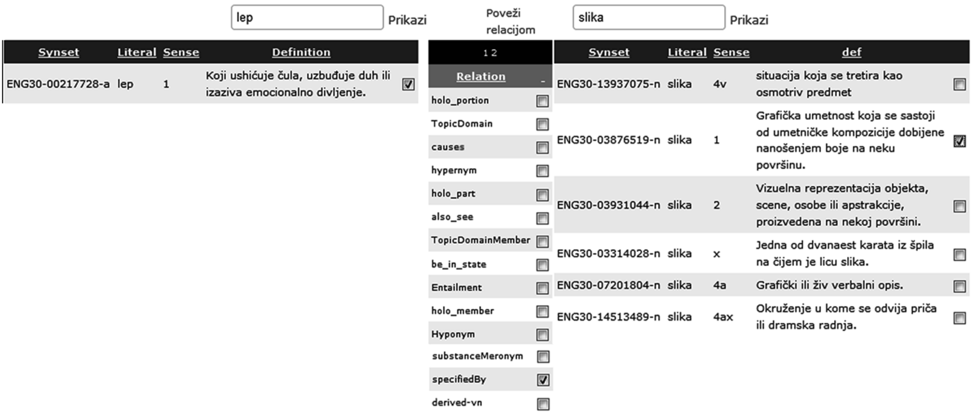


Figure 7: A Web tool for semi-automatic enlargement of SWN; in this case an adjective *lep* (*beautiful*) and a noun *slika* (*picture*) are linked by a pair of inverse relations *specificOf* / *specifiedBy*

tool which uses the rule (2) (see Section 3) in the ontology SWN to inference if the extracted linguistic structure represents the figure *simile*.

Linguistic structures extracted from the text collection by regular expressions indicate that comparisons which are uncommon in everyday speech can be found in poetic texts. For example, in the songs of Danojlić, structures like *as good as an elephant*, *calm like a cow*, *shaking like an arrow*, etc. occur. In songs of Ršumović, the following can be found: *blue as the carpet*, *green as carpet*, *ugly as a brush*, etc. At the end of this test, we analyzed linguistic structures extracted from datasets named *news* and *films*, which are used in SAFOS — a system for sentiment analyzing in Serbian language (Mladenović et al., 2016b), to assess the possibility of a further improvement in terms of recognizing the figurative speech. The result showed a low but uniform recognition accuracy[9] of about one third of the total number of appearances of the figure *simile*. In structures that were extracted from these datasets such as: *okarakterisan kao triler* (*characterized as a thriller*), *poznat kao zemlja* (*known as a country*), *urađen kao nastavak* (*made as a sequel*), etc., it can be seen that some of them satisfy the rule (2), but they are not similes and they could be detected as false positive candidates for instancing the figure *simile*. However, the recognition of the figures was performed using the knowledge of the ontology SWN, which is the reason

[9] Accuracy is one of the measures of classification quality and represents the ratio of the sum of true positive and true negative classified items and the sum of all items in the test collection.

why high precision was obtained, and false positive candidates were rejected[10]. On the other hand, unrecognized structures like *oštrim poput brijača* (*sharp as a razor*), *hrabrim kao SUPERMEN* (*brave as Superman*), *pući kao vidik* (*expanded as a view*), *razmnožavati se kao vinska mušica* (*multiplied as Drosophila*) indicate a low level of classification recall and the need for learning different linguistic structures besides those described, taking into account other types of nouns (named-entities, abbreviations, etc.) and parts of speech (verbs, for example). Classification accuracy of figure *simile* was carried out in all texts in a collection made of 10 digitized writings and the two datasets used as testing sets in sentiment analysis in Serbian. The results are shown in Figure 8.

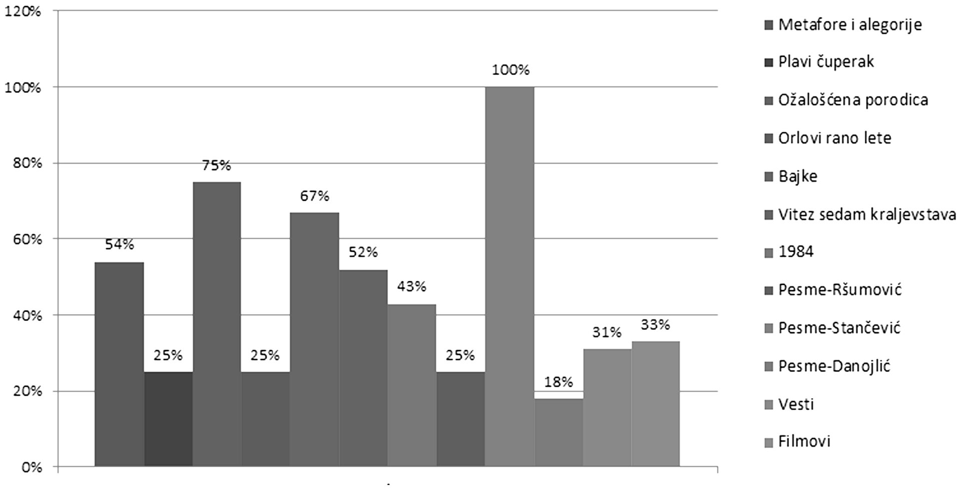


Figure 8: Accuracy of recognition of the figure *simile* in texts in Serbian using the ontology SWN

The evaluation of classification is executed in the process of classifying the figure *simile* in George R.R. Martin's novel "A Knight of the Seven Kingdoms", because that is the largest integral text in the collection. By applying regular expressions we found a total of 135 structures in a form "РЕЧ *као* РЕЧ{1,2}", "РЕЧ *попут* РЕЧ{1,2}" (WORD *as* WORD{1,2}). From these, a total of 107 structures represented the figure simile and 28 did not, like in the examples: "*Tu sam pecao kao dečak*"

[10] In the ontology SWN there is not a relation *specifiedBy* between nouns and adjectives, unless it is verified by methods of SWN ontology learning (see Section 4).

("*I caught fish there as a boy*"), "*Lim je imao naoštren kolac koji bi mogao poslužiti kao koplje*" ("*Lem had a sharpened stick that might serve for a spear*"), "*Dve godine je služio kao paž*" ("*He served two years as a page*"), etc. This classification is done by hand, and then a total of 135 structures subjected to ontological classification. This means that for each of 135 pairs (word, noun), a check is performed in order to see if they satisfy the rule (2). Some examples of recognized similes are: "*Bele kao kost bile su put i kosa Brindena Rečnog...*" ("*White as bone were the skin and hair of Brynden Rivers*"), "*Bila su hladna kao kamen, ali ih je bilo divno videti*" ("*They were hard as stone, but beautiful to look upon*"), "*...bilo je crveno kao krv*" ("*was red as blood*"). The results of ontological classifications are shown in Table 1, in the confusion matrix.

| figure simile classification | | manual | |
|---|---|---|---|
| | | yes | no |
| by classifier | yes | $tp = 42$ | $fp = 0$ |
| | no | $fn = 65$ | $tn = 28$ |

Table 1: Confusion matrix that represents classification of linguistic structures extracted by regular expressions from George R.R. Martin's novel "A Knight of the Seven Kingdoms" to those that represent figure *simile* and those that do not.

Based on the confusion matrix, estimates of the system are given in Table 2, and statistical assessment of a classifier by ROC curve is presented in Figure 9, where it is obtained that the $AUC > 0.5$ for a confidence interval of 95%.

| Evaluation measures of classification figure *simile* | $Precision = \frac{tp}{tp+fp}$ | $Recall = \frac{tp}{tp+fn}$ | $F_1 = \frac{2PR}{P+R}$ | Accuracy |
|---|---|---|---|---|
| Measures | 1.000 | 0.393 | 0.564 | 0.518 |

Table 2: Evaluation measures of classification of the figure *simile* in George R.R. Martin's novel "A Knight of the Seven Kingdoms"

Receiver operating characteristic curve (ROC) is one of the methods of evaluating the classifier that uses a graphical representation of the relationship of sensitivity ($\frac{tp}{tp+fn}$) and specificity ($\frac{tn}{tn+fp}$) for each possible score on the test. Figure 9 shows the (solid line) ROC curve of classification figure simile, while the diagonal line (shown in dashed line, called diagonal accidental outcomes) represents the outcome of a random classification. The area under the ROC curve ($AUC$) is interpreted as the probability that a randomly selected item labeled as positive (being figure simile), has a higher score than the one labeled as negative. $AUC$ is a measure of the accuracy of the classifier and may have a value of $AUC = 0.5$, when the ROC curve coincides with the diagonal of random outcomes, to $AUC = 1$ in the case of absolute separation of classes. Classification of figure simile in our experiment achieves $AUC = 0.696$.

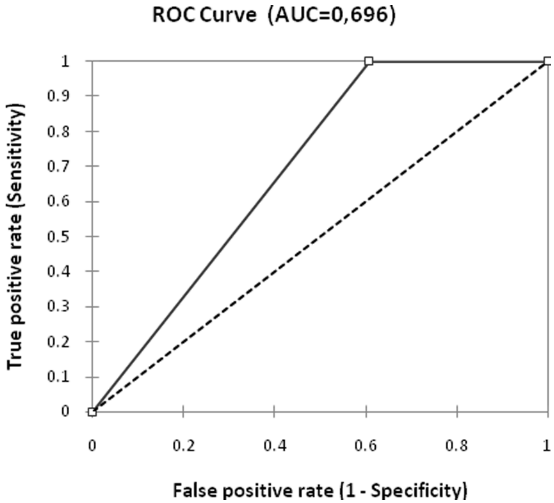


Figure 9: Statistical evaluation of the classification of the figure *simile* in George R.R. Martin's novel "A Knight of the Seven Kingdoms"

The results of the classification show low grade of recall. There are three groups of structures that are not recognized by the proposed system:

1. structures containing personal nouns as given in examples: *pametan kao princ Eris* (*as clever as Prince Aerys*), *visok kao Dank* (*almost of a height with Dunk*), etc;
2. structures that do not contain an adjective and a noun, but a verb and a noun as in examples: "*Onda se okrenu kao vihor, i jurnu u noć*" ("*Whirling, he darted back out into the night*") and "*Jaje ulete kao bez duše*" ("*Egg burst in panting*");
3. structures containing an adjective and a noun, but the order is such that it can not be unambiguously detected by propesed system, and examples of such struc-

tures are found in sentences: "*...gole kao od majke rođene, brašnjave od glave do pete*", "*...čije su vode svetlucale crveno i zlatno, sjajne kao ploča kovanog bakra*", "*Zbog blede kože i kao kreč bele kose ličio je na živi leš*".

Structures containing personal noun can be recognized by including tools for Named Entity Recognition and a pattern in the form of "PRIDEV *kao* VLASTITA_IMENICA" (ADJECTIVE *as* PROPER_NOUN). Those structures containing a verb and a noun can be identified by introducing new paterns like "GLAGOL *kao* IMENICA" (VERB *as* NOUN), and for the third mentioned group of structures, if we would use patterns like "*kao* IMENICA [VEZNIK|PREDLOG]? PRIDEV" (*as a* NOUN [CONJUNCTION| PROPOSAL]? ADJECTIVE).

## 6 Conclusion

In this paper, we propose a method for recognizing rhetorical figures that belong to the group of figures of twisted meaning (tropes). The method uses rules defined in the ontology based on the semantic network Serbian WordNet to identify whether a linguistic structure extracted from text satisfies some of the rules to be labeled as a rhetorical figure. The method was tested by detecting rhetorical figure simile in a collection of ten digitized writings and the two textual datasets which were previously used in sentiment analysis in Serbian. An evaluation of the system was conducted in the text of the George Martin's novel "Knight of the Seven Kingdoms". Evaluation results ($AUC = 0.696$) showed that the classification of rhetorical figure achieved middle grade, but precision of 100% indicated that further learning of SWN ontology, based on the proposed semi-automatic method, can improve the classification of this figure.

In future work, in the case of classification of the figure *simile*, we will introduce patterns described in section 5 to identify structures which are not recognized by the proposed system. In the case of the other figures, we will work on building of datasets for testing classifiers of rhetorical figures *irony*, *sarcasm* and *oxymoron*. We also plan to use the proposed method in a process of building features in sentiment analysis in Serbian. We will also try to expand the set of ontological rules to include procedures of generalization and specifications of concepts that are already included by the rules laid down by this system.

## References

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In Calzolari et al. (2010). ISBN 2-9517408-6-7. URL http://www.lrec-conf.org/proceedings/lrec2010/summaries/769.html.

Krešimir Bagić. *Rječnik stilskih figura.* Školska knjiga, Zagreb, 2012.

Francesco Barbieri, Francesco Ronzano, and Horacio Saggion. UPF-taln: SemEval 2015 Tasks 10 and 11. Sentiment Analysis of Literal and Figurative Language in Twitter. In Daniel M. Cer, David Jurgens, Preslav Nakov, and Torsten Zesch, editors, *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2015, Denver, Colorado, USA, June 4-5, 2015*, pages 704–708. The Association for Computer Linguistics, 2015. ISBN 978-1-941643-40-2. URL http://aclweb.org/anthology/S/S15/S15-2119.pdf.

Daniel G. Bobrow. A Question-Answering System for High School Algebra Word Problems. In *Proceedings of AFIPS Conference, 26. FJCC, Part I*, pages 591–614, New York, 1964. ACM. doi: 10.1145/1464052.1464108.

Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors. *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta*, 2010. European Language Resources Association. ISBN 2-9517408-6-7. URL http://www.lrec-conf.org/proceedings/lrec2010/index.html.

Paula Carvalho, Luís Sarmento, Mário J. Silva, and Eugénio de Oliveira. Clues for Detecting Irony in User-generated Contents: Oh...!! It's "So Easy" ;-). In *Proceedings of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion*, pages 53–56, New York, 2009. ACM.

Paula Carvalho, Luís Sarmento, Jorge Teixeira, and Mário J. Silva. Liars and Saviors in a Sentiment Annotated Corpus of Comments to Political Debates. In Lin et al. (2011), pages 564–568. ISBN 978-1-932432-88-6. URL http://www.aclweb.org/anthology/P11-2099.

Vladan Devedžić. *Semantic Web and Education.* Integrated Series in Information Systems. Springer-Verlag New York Inc., 2010. ISBN 9781441942012.

Richard Farkas, Eszter Simon, Gyorgy Szarvas, and Daniel Varga. Gyder: Maxent Metonymy Resolution. In Eneko Agirre, Lluís Màrquez, and Richard Wicentowski, editors, *SemEval '07: Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 161–164, Stroudsburg, PA, USA, 2007. Association for Computational Linguistics.

Dan Fass. met*: A Method for Discriminating Metonymy and Metaphor by Computer. *Computational Linguistics*, 17(1):49–90, 1991. ISSN 0891-2017.

Christiane Fellbaum. *WordNet: An Electronic Lexical Database.* Language, Speech, and Communication. A Bradford Book, 1998. ISBN 978-0262061971.

Elena Filatova. Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on*

*Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, 2012. European Language Resources Association (ELRA). ISBN 978-2-9517408-7-7.

Jakub Gawryjolek, Chrysanne Di Marco, and Randy A. Harris. An Annotation Tool for Automatically Detecting Rhetorical Figures — System Demonstration. In *Proceedings of the IJCAI-09 workshop on Computational Models of Natural Argument*, Pasadena, CA, 2009.

Roberto I. González-Ibáñez, Smaranda Muresan, and Nina Wacholder. Identifying Sarcasm in Twitter: A Closer Look. In Lin et al. (2011), pages 581–586. ISBN 978-1-932432-88-6. URL http://www.aclweb.org/anthology/P11-2102.

Thomas R. Gruber. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2):199–220, 1993. ISSN 1042-8143. doi: 10.1006/knac.1993.1008. URL http://dx.doi.org/10.1006/knac.1993.1008. Special issue: Current Issues in Knowledge Modeling.

Yanfen Hao and Tony Veale. An Ironic Fist in a Velvet Glove: Creative Mis-Representation in the Construction of Ironic Similes. *Minds and Machines*, 20 (4):635–650, 2010. doi: 10.1007/s11023-010-9211-1. URL http://dx.doi.org/10.1007/s11023-010-9211-1.

Andrew Hardie, Veronika Koller, Paul Rayson, and Elena Semino. Exploiting a Semantic Annotation Tool for Metaphor Analysis. In Davies, Matthew and Rayson, Paul and Hunston, Susan and Danielsson, Pernilla, editor, *Proceedings of the Corpus Linguistics 2007 Conference*, Birmingham, 2007. University of Birmingham.

Randy Harris and Chrysanne Di Marco. Constructing a Rhetorical Figuration Ontology. In *Symposium on Persuasive Technology and Digital Behavior Intervention, Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB)*, pages 47–52, Edinburgh, 2009.

Daniel Devatman Hromada. Initial Experiments with Multilingual Extraction of Rhetoric Figures by means of PERL-compatible Regular Expressions. In Irina P. Temnikova, Ivelina Nikolova, and Natalia Konstantinova, editors, *Proceedings of the Student Research Workshop associated with The 8th International Conference on Recent Advances in Natural Language Processing, RANLP 2011, 13 September, 2011, Hissar, Bulgaria*, pages 85–90. RANLP 2011 Organising Committee, 2011. URL http://www.aclweb.org/anthology/R11-2013.

Ashley R. Kelly, Nike A. Abbott, Randy Allen Harris, Chrysanne DiMarco, and David R. Cheriton. Toward an Ontology of Rhetorical Figures. In Júnia Coutinho Anacleto, Renata Pontin de Mattos Fortes, and Carlos J. Costa, editors, *Proceedings of the 28th Annual International Conference on Design of Communication, SIGDOC 2010, São Carlos, São Paulo state, Brazil, September 26-29, 2010*, pages 123–130. ACM, 2010. ISBN 978-1-4503-0403-0. doi: 10.1145/1878450.1878471. URL http://doi.acm.org/10.1145/1878450.1878471.

Alistair Kennedy and Diana Inkpen. Sentiment Classification of Movie Reviews Using Contextual Valence Shifters. *Computational Intelligence*, 22(2):110–125, 2006.

doi: 10.1111/j.1467-8640.2006.00277.x. URL http://dx.doi.org/10.1111/j.1467-8640.2006.00277.x.

Svetla Koeva, Cvetana Krstev, and Duško Vitas. Morpho-Semantic Relations in WordNet - a Case Study for two Slavic Languages. In *Proceedings of Global WordNet Conference*, pages 239–253, Szeged, 2008. University of Szeged, Department of Informatics.

Veronika Koller, Andrew Hardie, Paul Rayson, and Elena Semino. Using a Semantic Annotation Tool for the Analysis of Metaphor in Discourse. *Metaphorik.de*, 15: 141–160, 2008.

Johannes Leveling. FUH (FernUniversität in Hagen): Metonymy Recognition Using Different Kinds of Context for a Memory-Based Learner. In Eneko Agirre, Lluís Màrquez, and Richard Wicentowski, editors, *SemEval '07: Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 153–156, Stroudsburg, PA, USA, 2007. Association for Computational Linguistics.

Dekang Lin, Yuji Matsumoto, and Rada Mihalcea, editors. *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA - Short Papers*, volume 2, 2011. The Association for Computer Linguistics. ISBN 978-1-932432-88-6.

Katja Markert and Malvina Nissim. Metonymy Resolution As a Classification Task. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP '02, pages 204–213, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics. doi: 10.3115/1118693.1118720. URL http://dx.doi.org/10.3115/1118693.1118720.

Zachary J. Mason. CorMet: A Computational, Corpus-based Conventional Metaphor Extraction System. *Computational Linguistics*, 30(1):23–44, 2004. ISSN 0891-2017. doi: 10.1162/089120104773633376.

Ruslan Mitkov. *Anaphora Resolution*. Longman, Cambridge, UK, 2002.

Jelena Mitrović. Electronic Tools and Resources for Multi-Word Unit Detection and Research in Serbian. The 2th General Meeting of The IC1207 COST Action, PARSEME, Athens, Greece, 10–11 March, 2014, 2014. URL http://typo.uni-konstanz.de/parseme/images/Meeting/2014-03-11-Athens-meeting/PosterAbstracts/Mitrovic.pdf.

Jelena Mitrović, Miljana Mladenović, and Cvetana Krstev. Adding MWEs to Serbian Lexical Resources Using Crowdsourcing. The 5th PARSEME General Meeting, Iaşi, Romania, 23–24 September 2015, 2015. URL http://typo.uni-konstanz.de/parseme/images/Meeting/2015-09-23-Iasi-meeting/WG1-MITROVIC-MLADENOVIC-KRSTEV-poster.pdf.

Miljana Mladenović and Jelena Mitrović. Ontology of Rhetorical Figures for Serbian. In Ivan Habernal and Václav Matousek, editors, *Text, Speech, and Dialogue* -

*16th International Conference, TSD 2013, Pilsen, Czech Republic, September 1-5, 2013. Proceedings*, volume 8082 of *Lecture Notes in Computer Science*, pages 386–393. Springer, 2013. ISBN 978-3-642-40584-6. doi: 10.1007/978-3-642-40585-3_49. URL http://dx.doi.org/10.1007/978-3-642-40585-3_49.

Miljana Mladenović and Jelena Mitrović. Semantic Networks for Serbian: New Functionalities of Developing and Maintaining a WordNet Tool. In Gordana Pavlović-Lažetić, Cvetana Krstev, Ivan Obradović, and Duško Vitas, editors, *Natural Language Processing for Serbian: Resources and Applications*, pages 1–11, Belgrade, 2014. University of Belgrade, Faculty of Mathematics. ISBN 978-86-7589-088-1.

Miljana Mladenović, Jelena Mitrović, and Cvetana Krstev. A Language-Independent Model for Introducing a New Semantic Realation between Adjectives and Nouns in a WordNet. In Verginica Barbu Mititelu, Corina Forăscu, Christiane Fellbaum, and Piek Vossen, editors, *The Proceedings of Eighth Global WordNet Conference 2016*, pages 218–225, Bucharest, Romania, 2016a. ISBN 978-606-714-239-6.

Miljana Mladenović, Jelena Mitrović, Cvetana Krstev, and Duško Vitas. Hybrid Sentiment Analysis Framework for a Morphologically Rich Language. *Journal of Intelligent Information Systems*, 46(3):599–620, 2016b. doi: 10.1007/s10844-015-0372-5. URL http://dx.doi.org/10.1007/s10844-015-0372-5.

Cristina Nicolae, Gabriel Nicolae, and Sanda Harabagiu. UTD-HLT-CG: Semantic Architecture for Metonymy Resolution and Classification of Nominal Relations. In Eneko Agirre, Lluís Màrquez, and Richard Wicentowski, editors, *SemEval '07: Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 454–459, Stroudsburg, PA, USA, 2007. Association for Computational Linguistics.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs Up?: Sentiment Classification Using Machine Learning Techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, volume 10 of *EMNLP '02*, pages 79–86, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics. doi: 10.3115/1118693.1118704. URL http://dx.doi.org/10.3115/1118693.1118704.

Massimo Poesio and Ron Artstein. Anaphoric Annotation in the ARRAU Corpus. In Calzolari et al. (2010), pages 1170–1174. ISBN 2-9517408-6-7. URL http://www.lrec-conf.org/proceedings/lrec2010/index.html.

Vassiliki Rentoumi, Stefanos Petrakis, Manfred Klenner, George A. Vouros, and Vangelis Karkaletsis. United we Stand: Improving Sentiment Analysis by Joining Machine Learning and Rule Based Methods. In Calzolari et al. (2010). ISBN 2-9517408-6-7. URL http://www.lrec-conf.org/proceedings/lrec2010/summaries/41.html.

Antonio Reyes and Paolo Rosso. Building Corpora for Figurative Language Processing: The Case of Irony Detection. In *Proceedings of the 4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals (in conjunction with LREC 2012)*, pages 94–98, 2012a.

Antonio Reyes and Paolo Rosso. Making Objective Decisions from Subjective Data: Detecting Irony in Customer Reviews. *Decision Support Systems*, 53(4):754–760, 2012b. doi: 10.1016/j.dss.2012.05.027. URL http://dx.doi.org/10.1016/j.dss.2012.05.027.

Ekaterina Shutova, Simone Teufel, and Anna Korhonen. Statistical Metaphor Processing. *Computational Linguistics*, 39(2):301–353, 2013. ISSN 0891-2017. doi: 10.1162/COLI_a_00124.

Oren Tsur, Dmitry Davidov, and Ari Rappoport. ICWSM - A great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews. In William W. Cohen and Samuel Gosling, editors, *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010*. The AAAI Press, 2010. URL http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1495.

Miloš Utvić. *Izgradnja referentnog korpusa savremenog srpskog jezika*. Doktorska disertacija, Univerzitet u Beogradu, Filološki fakultet, Beograd, 2014.

Tony Veale. Detecting and Generating Ironic Comparisons: An Application of Creative Information Retrieval. In *Artificial Intelligence of Humor, Papers from the 2012 AAAI Fall Symposium, Arlington, Virginia, USA, November 2-4, 2012*, volume FS-12-02 of *AAAI Technical Report*. AAAI, 2012. URL http://www.aaai.org/ocs/index.php/FSS/FSS12/paper/view/5557.

Tony Veale and Yanfen Hao. Support Structures for Linguistic Creativity: A Computational Analysis of Creative Irony in Similes. In *Proceedings of CogSci 2009, the 31st Annual Meeting of the Cognitive Science Society*, pages 1376–1381, 2009.

Lowri Williams, Christian Bannister, Michael Arribas-Ayllon, Alun D. Preece, and Irena Spasić. The Role of Idioms in Sentiment Analysis. *Expert Systems with Applications*, 42(21):7375–7385, 2015. doi: 10.1016/j.eswa.2015.05.039. URL http://dx.doi.org/10.1016/j.eswa.2015.05.039.

# A Appendix

Collection of 10 digitized writings of various genres (children's literature, novels, fairy tales, comedies and essays) used for testing a proposed method for recognition of rhetorical figure simile.

1. Antić, Miroslav. (1965). Plavi čuperak, accessed 25.02.2016, http://www.kotraza-dragacevo.org/Knjige.html
2. Bajke Za Decu, accessed 25.02.2016, http://www.zadecu.com/Bajke
3. Ćopić, Branko. Orlovi rano lete, accessed 25.02.2016, http://www.kotraza-dragacevo.org/Knjige.html

4. Danojlić, Milovan. (1966). Kako spavaju tramvaji, accessed 25.02.2016, http://borovo1234.weebly.com/uploads/1/2/6/2/12626321/milovan_danojlic_-_kako_spavaju_tramvaji.pdf
5. Martin, George R. R. Vitez sedam kraljevstava. Beograd: Laguna, 2014.
6. Nušić, Branislav. (1935). Ožalošćena porodica, accessed 25.02.2016, https://www.scribd.com/doc/60214229/Branislav-Nusic-Ozaloscena-Porodica
7. Orwell, Gorge. 1984. Beograd: Laguna, 2014.
8. Petrović, Mihailo. Metafore i alegorije. Beograd: Srpska književna zadruga, 1967.
9. Ršumović, Ljubivoje. Ma šta mi reče. Beograd: Laguna, 2013.
10. Stančević, Sreten. „Dečiji osmeh". Klub pesnika "Orfej", accessed 25.02.2016, http://www.orfejsu.com/deciji-osmeh-sreten-stancevic-zbirka-decijih-pesama/