

Онтолошко препознавање реторичких фигура

УДК 811.163.41'322.2

САЖЕТАК: Аутоматско препознавање реторичких фигура (компарације, ироније, сарказма, хумора, метафоре и сл.) све чешће се користи у задацима обраде природног језика, пре свега за унапређење система класификације текста према осећањима, машинског превођења, али и система који анализирају језичке структуре на различитим нивоима. У овом раду предложена је метода аутоматског препознавања и класификације реторичких фигура из групе тропи која користи правила онтолошког закључивања у онтологији Српски ворднет (SWN). Евалуација методе изведена је над реторичком фигуром *компарација*, а статистичка оцена бинарног класификатора ROC кривом ($AUC = 0,696$) указује да се он може успешно користити у решавању ове врсте задатака. За даље обучавање онтологије SWN, предложена је полу-аутоматска метода учења онтологије повећањем броја и врста релација које могу помоћи у откривању фигуративног говора у текстовима на српском језику.

КЉУЧНЕ РЕЧИ: реторичке фигуре, компарација, класификација помоћу онтологије, WordNet

РАД ПРИМЉЕН: 25. фебруар 2016.

РАД ПРИХВАЋЕН: 9. мај 2016.

Миљана Младеновић

ml.miljana@gmail.com

Универзитет у Београду,

Математички факултет

1. Увод

Аутоматско препознавање реторичких фигура (фигуративног говора) у тексту и њихова анотација није ново поље интересовања у обради природног језика. Једно од првих истраживања у области аутоматског препознавања и интерпретације реторичких фигура које се бави идентификацијом *анафоре* (anaphora resolution) дато је 1964. у раду „A question-answering system for high school algebra a word problems“ (Bobrow, 1964), док је 1991. у раду „met*: A method for discriminating metonymy and metaphor by computer“ (Fass, 1991) предмет аутоматског препознавања била *метафора* и *метонимија*. У литератури су

описане методе аутоматског аотирања *метафоре* (Mason, 2004; Hardie et al., 2007; Koller et al., 2008; Shutova et al., 2013), *анафоре* (Mitkov, 2002; Poesio and Artstein, 2010), *метонимије* (Farkas et al., 2007; Leveling, 2007; Nicolae et al., 2007), *ироније* (Carvalho et al., 2009; Veale, 2012), *сарказма* (Tsur et al., 2010; González-Ibáñez et al., 2011) и др. Међутим, методе анализе осећања у тексту (енг. sentiment analysis — SA) највећим делом разматрају текстове који садрже дословно значење или занемарују утицај фигуративног говора на поларитет осећања у тексту.

У раду (Hao and Veale, 2010) анализа употребе фигуративног говора показала је да одређене реторичке фигуре у тексту делују попут модификатора значења чија је улога у промени поларитета осећања речи или фраза приказана у раду „Sentiment Classification of Movie Reviews Using Contextual Valence Shifters“ (Kennedy and Inkpen, 2006). У општем случају, *модификатори* су речи и изрази који смањују, повећавају или у потпуности мењају поларитет осећања речи или фраза уз које стоје (нпр. поларитет речи *срећан* се може у потпуности променити речју *нимало*). На сличан начин делују и одређене групе реторичких фигура чијом се применом може утицати на поларитет осећања речи или фраза у тексту. По својој дефиницији, фигуре *иронија* и *сарказам* мењају поларитет, *дисфемизам* и *хипербола* појачавају постојећи ниво сентименталне изражајности, док га *литота* и *еуфемизам* умањују. *Метафора*, *метонимија*, *оксиморон* и *компарација* имају сложеније механизме деловања у оба правца промене и јачине и поларитета осећања.

Стога, поступак аутоматске или полуаутоматске идентификације реторичких фигура у тексту може побољшати процес анализе осећања, али и машинског превођења и других апликација заснованих на обради природних језика. На пример, традиционални систем класификације текста на основу осећања (Pang et al., 2002) би на исти начин интерпретирао реченице „Он је брз као зељ“ и „Он је брз као пуж“, док би се укључењем процеса препознавања фигуративног говора у другој реченици учила реторичка фигура иронија, чиме би њен поларитет осећања био у потпуности промењен у односу на дословни. Истраживања објављена у (Reyes and Rosso, 2012b) показала су да се тачност класификације у задатку анализе осећања у тексту може знатно побољшати (са 54% до максималних 89,05%) када се укључе предиктори којима се препознаје фигуративни говор, у односу на скуп предиктора који текст третирају у дословном значењу. Метода машинског учења којом се решава проблем SA унапређена је интегисањем са методом заснованом на правилима којом се препознаје фигуративна употреба језика (Rentoumi et al., 2010). У раду (Williams et al., 2015) разматрана је улога идиома у анализи осећања у тексту. У два експеримента аутори су показали побољшање F-мере класификатора са 45% на 64% и са 46% на 61% када

се као предиктори класификације користе и идиоми који носе фигуративно значење. У раду (Carvalho et al., 2011) резултати анализе осећања изражених у коментарима објављеним на веб порталима новинских вести показују да би 11% коментара посматраног скупа било некоректно означено позитивним поларитетом, уколико би изостала анализа и препознавање појаве реторичке фигуре иронија. Због тога је анализа осећања у текстовима у којима је изражен фигуративни говор нови изазов на пољу обраде природног језика. У 2015. је први пут, на глобалном плану, постављен један семантички задатак ове врсте — задатак семантичке анализе фигуративног говора на Твитеру (International Workshop on Semantic Evaluation — SemEval-2015).¹

Истраживања у области обраде фигуративног говора крећу се у два правца:

- ка побољшању метода класификације поларитета осећања (Reyes and Rosso, 2012b; Rentoumi et al., 2010);
- ка бољем разумевању језичких структура — у раду (Veale and Hao, 2009) упоредне анализе ироничних компарација на енглеском и кинеском језику показале су да је *иронија* лингвистички и културолошки феномен, јер се из посматраног скупа компарација које се примењују у енглеском свега 3–4% могу применити и у кинеском језику.

У истраживању фигуративног говора важну улогу имају семантичке мреже као што је ворднет (енг. WordNet) (Mason, 2004; Barbieri et al., 2015), онтологије (Harris and Di Marco, 2009; Kelly et al., 2010) и лексички ресурси као што су корпуси (Mason, 2004; Hao and Veale, 2010; Reyes and Rosso, 2012a), специјализовани речници жаргона, емотикона и интерпункцијских знакова (Carvalho et al., 2009; González-Ibáñez et al., 2011; Barbieri et al., 2015) и лексикони попут SentiWordNet-a (Rentoumi et al., 2010; González-Ibáñez et al., 2011; Barbieri et al., 2015).

У овом раду најпре је предложена метода аутоматског препознавања реторичких фигура из групе троши која се ослања на онтолошко закључивање, а затим и метода учења онтологије у смислу повећања броја и врста релација у SWN које могу помоћи у откривању таквих фигура у текстовима на српском језику.

2. Методе аутоматског препознавања реторичких фигура

Аутоматско препознавање реторичких фигура у тексту зависи од природе фигуре и језичке структуре коју фигура користи. Да би могао бити изграђен ефикасан систем аутоматског препознавања фигуративног говора, потребно је,

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

најпре, формално дефинисати и описати реторичке фигуре. Поступак изградње прве формалне доменске онтологије реторичких фигура у српском језику *Ret-Fig* приказан је у раду „Ontology of Rhetorical Figures for Serbian“ (Mladenić and Mitrović, 2013). Онтологија *RetFig* описује 98 реторичких фигура. За сваку од њих дефинисано је којој реторичкој и којој лингвистичкој групи припада. Дефинисани су лингвистички опсег, објекти и елементи који учествују у креирању дате реторичке фигуре, међусобни однос објеката и елемената као и лингвистичке операције које учествују у процесу креирања реторичке фигуре. Све фигуре подељене су у четири групе: фигуре наглашавања (дикције), фигуре конструкције, фигуре речи (тропи) и фигуре проширења мисли. Фигуре наглашавања се заснивају на дејству одређених гласова у говору. Понављање одређених гласова или група гласова у говору, односно слова или група слова у тексту, њихово изостављање или уметање на неочекиваним местима, опонашање одређених звукова и шума из природе утиче на појачавање или смањење значаја језичких структура над којима се изводе. Ова врста фигура нема утицаја на значење структура над којима се граде и не мења их, већ само наглашава њихово основно значење. Фигуре конструкције настају мењањем распореда речи у реченици или у некој другој већој целини (одломку, стиху и сл.) у односу на уобичајени тј. подразумевани, па се могу посматрати као фигуре поретка или распореда. И ова врста фигура не мења основно значење језичких структура над којима се граде. Насупрот овим групама, фигуре речи (тропи) имају задатак да промене основно значење неке речи. Фигурама проширења мисли такође се мења основно значење језичке структуре која је комплекснија од речи.

Аутоматско препознавање реторичких фигура наглашавања (дикције) и конструкције у тексту могуће је спровести применом формализованих образаца како је то примењено у радовима (Gawryjolek et al., 2009; Hromada, 2011). На пример, *антиметабола* је фигура конструкције коју карактерише понављање речи или фраза из првог дела реченице и у другом, али обрнутим редоследом². У раду (Gawryjolek et al., 2009) описан је модел препознавања као и одговарајући алат за аутоматску анотацију (Java Annotation Tool Of Rhetoric — JANTOR) у коме се за *антиметаболу* дефинише синтаксички образац облика $[W_a] \dots [W_b] \dots [W_b] \dots [W_a]$ и којим се проналази појава ове фигуре у тексту, где су W_a and W_b речи или фразе које се у посматраној реченици у другом делу реченице појављују у инверзном поретку у односу на први. Резултати примене JANTOR-а показују да се реторичке фигуре из група фигура конструкције и наглашавања успешно проналазе на овај начин, јер је могуће дефинисати регуларне изразе који их једнозначно одређују. Обе ове класе фигура засноване су на синтаксичким или морфолошким операцијама у склопу граматичких

² Нпр. „Једемо да живимо, а не живимо да једемо.“; „Пиши као што говориш, говори као што пишеш.“

правила природног језика. Међутим структура реторичких фигура у класама фигура тропи и проширења мисли је таква да оне немају једнозначан образац примене на нивоу синтаксе.

У процесима аутоматског препознавања и анотације реторичких фигура које утичу на промену значења текста успешно се користе методе машинског учења. У радовима који су се бавили препознавањем *метонимије* коришћене су методе надгледаног машинског учења као што су листе одлучивања (Markert and Nissim, 2002), максимална ентропија (Farkas et al., 2007), к-најближих суседа (Leveling, 2007)³, стабла одлучивања, логистичка регресија (Nicolae et al., 2007) и др. Када је реч о *иронији* и *сарказму*, у раду (Tsur et al., 2010) коришћена је метода полу-надгледаног учења ради препознавања и класификације саркастичних твитова и коментара са сајта за електронску трговину Амазон. Метода потпорних вектора и логистичка регресија коришћене су у процесу класификације твитова на саркастичне и позитивно и негативно поларисане несаркастичне твитове у (González-Ibáñez et al., 2011).

С друге стране, показано је да се *иронија* и *сарказам* могу успешно детектовати и другим техникама. У (Carvalho et al., 2009) генерисно је осам образаца (као нпр. $P_{laugh} = (LOL|AH|EMO+)$, $P_{quote} = (ADJ_{pos}|N_{pos})\{1,2\}$) који представљају структуралне индикаторе постојања *ироније* у коментарима објављиваним на новинским порталима на португалском језику. Метода групне расподеле рада (енг. crowd sourcing) у процесу детекције *ироније* и *сарказма* у приказима производа на сајту Амазон примењена је у (Filatova, 2012), док је у (Veale, 2012) дефинисана аутоматска метода екстракције семантичког знања из примера примене фигуре компарације ради препознавања *ироније*. У (Mitrović et al., 2015; Mladenović et al., 2016a) коришћена је метода групне расподеле рада у процесу детекције *компарације* у аотираном корпусу и предложен је алгоритам за аутоматску екстракцију детектованих кандидата на основу фреквентности појављивања у корпусу.

У овом раду у поступку детекције реторичких фигура из групе фигура речи (тропи) користи се метода онтолошког закључивања над онтологијом проистеклом из семантичке мреже Српски ворднет (SWN)⁴.

3. Онтолошко препознавање фигура речи (тропи)

Онтологија је један од видова репрезентације знања. Уколико је онтологија дата на неком формалном језику и складишти се у формату који је рачунарски

³ Коришћен је софтверски пакет TiMBL заснован на имплементацији алгоритма Memory-Based Learning који је проистекао из методе „k-nearest neighbor“.

⁴ <http://sm.jerteh.rs/MemberZone/eW3.aspx?id=miljanam>

читљив, реч је о формалној онтологији. Схваћена као „формална спецификација дељене концептуализације“ (Gruber, 1993), онтологија указује на онај вид знања који је могуће пренети, разменити и употребити. Према (Devedžić, 2010), главна сврха онтологија састоји се у дељењу и вишеструкој употреби знања од стране различитих интелигентних агената и апликација. Зависно од тога који део стварности описују, онтологије могу бити:

- онтологије највишег нивоа (енг. top level ontologies) — када моделирају опште концепте, а знање које репрезентују је свеобухватно, систематизовано и применљиво у великом броју апликација;
- доменске онтологије (енг. domain ontologies) — када се знања која репрезентују тичу једног домена или класе проблема;
- онтологије задатака (енг. task ontologies) или апликацијске онтологије (енг. application ontologies) — садрже само она знања неопходна за извршавање дате класе задатака.

У овом раду предлажемо методу идентификације реторичких фигура из класе фигура речи (тропи) коју, по дефиницији, чине оне фигуре чија се улога заснива на промени основног значења речи или фразе унутар једне језичке целине која може бити: реч, фраза, стих или реченица. Реторичке фигуре из ове класе своју улогу заснивају на:

- замени значења посматране речи тако што се она замени неком другом речи или фразом (нпр. фигура *иронија* се у примеру „*Красан си ми ти пријатељ*“ састоји у замени речи *лош* речју *красан* у посматраној реченици);
- додели вишеструког значења посматраној речи или фрази (нпр. фигура *синегдоха* где у примеру „*Добио је кров над главом*“, фраза *кров над главом* може добити и значење речи *сигурност*, *уточниште* и њихових синонима);
- генерисању новог значења посматране речи или фразе (фигура *оксиморон* која спаја појмове супротних значења у нови појам, нпр. *виртуелна стварност*).

У овим случајевима није могуће дефинисати морфолошко-синтаксичке шаблоне, али је могуће утврдити какве семантичке релације постоје између учесника који граде одређену фигуру и на основу тога дефинисати правила закључивања која се могу применити над онтологијом каква је проистекла из семантичке мреже ворднет (Fellbaum, 1998). Фигуре које припадају класи тропи добили смо SPARQL упитом над доменском онтологијом *RetFig* (Mladenović and Mitrović, 2013).

```
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
```

Тако је добијен скуп од 26 реторичких фигура (слика 1) које имају заједничку особину да припадају групи фигура тропи (?nazivRetGrupe="FIGURE_TROPI") и које карактерише неки вид промене уобичајеног значења једне речи (?nazivObjekta="REC"). Класа на врху овог дела онтологије *RetFig* означена је као Tropi. Она садржи три подкласе (VišestrukoZnačenje, IzmenjenoZnačenje, NovoZnačenje) којима се дефинише улога појединачних фигура, односно вид промене уобичајеног значења речи над којом се фигура примењује. Идеја је да се, уколико је могуће, свака од тих фигура дефинише применом правила (изражених помоћу Semantic Web Rule Language — SWRL)⁵ над онтологијом SWN.

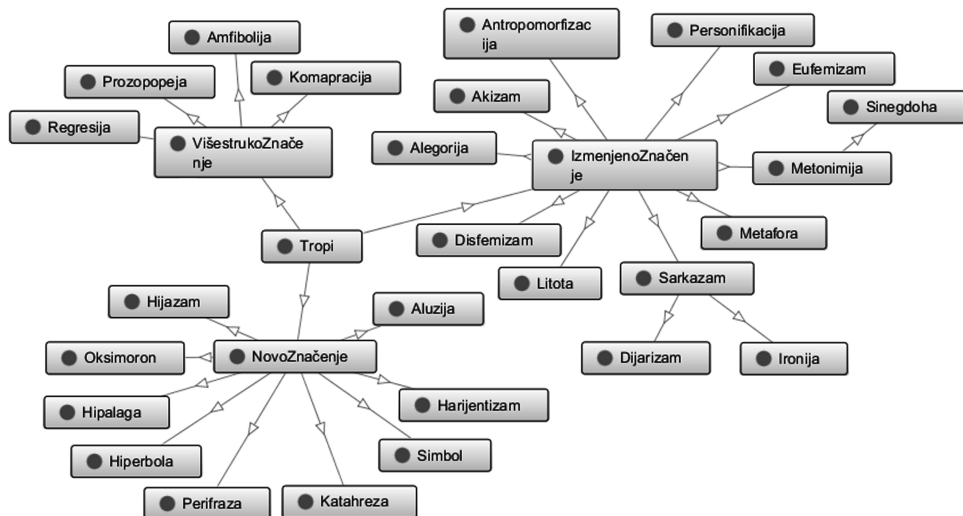
На примеру фигуре *иронија* (енг. irony) биће објашњен начин изградње правила над онтологијом SWN помоћу којих се може открити појава одговарајуће фигуре у тексту. Вербална иронија је појава речи или израза којим се једно каже, а супротно мисли и разуме. Право значење речи прикривено је и супротно исказаном значењу израза. Примери вербалне ироније заснивају се на неколико врста семантичких релација (Bagić, 2012), а често коришћени облици у српском језику заснивају се на употреби придева коме се даје значење њему подразумеваног супротног (антонимског) придева као у примерима:

- „Он је баш генијалан!“ (Скривено значење је тврдња да је неко заправо глуп.)
- „Види се да је жгољав!“ (Скривена тврдња да је неко дебео.)

Друга форма се састоји од везе именице и придева, али се придеву додељује скривено значење оног придева који се уобичајено везује уз наведену именицу.

- „Брз је као корњача.“
- „Храбар је као зец!“

⁵ Semantic Web Rule Language — SWRL је језик који се користи у области семантичког веба за представљање формалних логичких израза, а добијен је комбиновањем особина језика OWL DL (Web Ontology Language for Description Logic) и језика RML (Rule Markup Language).



Слика 1: Део таксономије онтологије изведене из онтологије *RetFig* који се односи на класу фигура речи (тропи)

У раду (Veale and Hao, 2009) користи се назив „иронична компарација“ како би се описала специфична врста ироније изведена из компарације. У наведеним примерима придев *брз* добија скривено значење придева који представља природну карактеристику именице *корњача*, а то је придев *спор*, док придев *храбар* добија скривено значење придева који је природна карактеристика именице *зеи*, а то је придев *плашљив*. Семантичке релације које постоје у онтологији SWN и које се могу користити при генерисању кандидата за детекцију фигуре *иронија* су релације *specificOf/specifiedBy* и релација *near_antonym*. Релације *specificOf/specifiedBy* представљају пар инверзних релација (Mladenović et al., 2016a) којима се повезују инстанце класе именичких синсетова са инстанцама синсетова њима специфичних придева. Релација *near_antonym* међусобно повезује инстанце класа синсетова придева који су директни антоними. Стога се правило за генерисање ироничних релација, примењено над онтологијом SWN, може изразити као OWL правило

$$\{?n : \textit{specifiedBy} ?p1. ?p1 : \textit{near_antonym} ?p2\} \Rightarrow \{?n : \textit{Irony} ?p2\}$$

или у форми SWRL правила

$$\textit{specifiedBy}(?n, ?p1), \textit{near_antonym}(?p1, ?p2) \rightarrow \textit{Irony}(?n, ?p2) \quad (1)$$

Правило (1) може се добити и SPARQL упитом (слика 2) над онтологијом SWN.

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.mmijana.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

Слика 2: Кандидати за генерисање фигура *иронија* и *компарација*

Резултат упита даје кандидате фигуре *иронија* у форми пара (adjective2, noun), нпр. (*храбар, зец*), али даје и кандидате фигуре *компарација* (енг. simile) које чине парови (adjective1, noun), нпр. (*плашљив, зец*), што је еквивалентно OWL правилу

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

или

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

Резултат упита са слике 2 аналоган је резултату закључивања на основу правила (1), али се из упита може уочити да у природном језику није уобичајено коришћење неких од кандидата за генерисање фигуре *иронија*. На основу интуиције која се у овом тренутку не доказује, може се рећи да су природни кандидати фигуре *иронија* они код којих је поларитет осећања субјекта (?adjective1) у RDF тројци⁶ ?adjective1 swn30:near_antonym ?adjective2 — позитиван.

⁶ RDF је модел података који користи форму тројке или триплета „субјекат-предикат-објекат“ (subject–predicate–object) да би описао ресурсе семантичког веба, обезбедио

На пример, природнији примери ироније исказани су тврдњама: „*Брз је као корњача*“ или „*Лак је као слон*“ него „*Спор је као стрела*“, или „*Тежак је као перце*“. У том смислу, претрага се може унапредити додатним условом који узима у обзир оне RDF тројке у којима је вредност негативног поларитета осећања субјекта једнака нули. Како сваки синсет у семантичкој мрежи Српски ворднет има етикету (Mladenović and Mitrović, 2014), као и вредност, позитивног и негативног сентимент поларитета, што је добијено успостављањем једнозначних релација са ресурсом SentiWordNet (Baccianella et al., 2010), то се, у циљу идентификације поларитета осећања субјекта у RDF тројци, упит може проширити контролом својства податка (dataproperty) `swn30:sentimentNegative`, а на основу вредности тог својства бирају се (филтрирају) само они елементи код којих је вредност тог својства једнака нули (слика 3).

```
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.mmijana.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

Слика 3: Побољшање претраге кандидата за генерисање фигуре *иронија*

Одговарајуће SWRL правило, које укључује додатни услов приказан на слици 3, дато је у наставку

$$\begin{aligned}
 & \text{specifiedBy}(?n, ?p1), \text{near_antonym}(?p1, ?p2), \\
 & \text{sentiment_negative}(?p1, ?sent), \text{swrlb:equal}(?sent, 0) \rightarrow \text{Irony}(?n, ?p2)
 \end{aligned} \tag{3}$$

Размотримо сада и структуру реторичке фигуре *оксиморон* која, по дефиницији, представља спајање појмова супротних значења у нови појам и чији

складиштење података у графовским базама података и представљање знања у онтолошким моделима.

су примери у SWN бази: *виртуелна стварност*, *гласна тишина*, *луда памет*, *живи фосил*, *ватрени лед*, и др. Правила за генерисање фигуре *оксиморон* примењена над онтологијом SWN могу се изразити као:

$$\{?p1 : near_antonym ?p2. ?p2 : derived-pos ?n\} \Rightarrow \{?p1 : oxymoron ?n\}$$

$$\{?p1 : near_antonym ?p2. ?p2 : be_in_state ?n\} \Rightarrow \{?p1 : oxymoron ?n\}$$

или у форми SWRL правила:

$$near_antonym(?p1, ?p2), derived-pos(?p2, ?n) \rightarrow oxymoron(?p1, ?n) \quad (4)$$

$$near_antonym(?p1, ?p2), be_in_state(?p2, ?n) \rightarrow oxymoron(?p1, ?n) \quad (5)$$

Правила (4) и (5) могу се добити и SPARQL упитом (слика 4) над онтологијом SWN. Резултат упита даје кандидате за генерисање фигуре *оксиморон* у форми пара (adjective1, noun), нпр. (*гласан*, *тишина*). Са слике 4 може се уочити да RDF тројке (adjective1, adjective2, noun) које садрже антонимске придеве постају кандидати за инстанцирање ове фигуре, када задовољавају услов постојања релација *derived-pos* или *be_in_state*, које представљају лексичке релације између синсетова различитих врста речи (енгл. cross-part of speech relations — XPoS), у овом случају између придева и именица (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.mmijana.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

Слика 4: Кандидати за генерисање фигуре *оксиморон*

Поред лексичко-семантичких релација које могу бити индикатор постојања одређених тропи фигура, може се користити и имплицитна релација синонимије којом се могу проналазити кандидати за инстанцирање фигуре *перифраза* (Mitrović, 2014), фигуре којом се један појам описује или замењује са више речи коришћењем неких битних особина тог појма (слика 5). Тако се реч *Pariz* може заменити синонимима из ворднета: *Grad svetlosti*, *Prestonica Francuske*, *Glavni grad Francuske*⁷, па парови попут (*Pariz*, *Grad svetlosti*) постају кандидати за инстанце ове фигуре.

noun	synonymNoun
Pariz	"prestonica Francuske"
Pariz	"glavni grad Francuske"
Pariz	"Grad svetlosti"
Pariz	"Pariz"

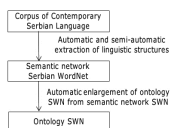
Слика 5: Кандидати за генерисање фигуре *перифраза*

4. Учење онтологије SWN

У претходном одељку описана су правила, дефинисана у онтологији SWN, којима се аутоматски генеришу инстанце неких реторичких фигура које припадају групи фигура тропи. Међутим, главни проблем у инжењерингу једне онтологије према (Devedžić, 2010) је њено оспособљавање да се мења истовремено са увећањем и изменом доменских знања. Способност онтологије да непрестано „учи“ (енг. ontology learning) тако што се аутоматским и полу-аутоматским поступцима екстрахују и аотирају нова знања из посматраног домена и интегришу са постојећим знањима у онтологији, у многоме зависи

⁷ Сви примери екстраховани из ворднета приказани су у свом оригиналном облику, односно на латиници.

од алата који се у ту сврху примењују. Како је задатак онтологије SWN да препознаје различите реторичке фигуре, врло је важно да она „учи“ различите облике фигуративног говора из примера на природном језику. С обзиром да се процес тог учења заснива на препознавању учесника у свим семантичким релацијама онтологије SWN, а посебно у релацијама као што су: *specificOf*, *specifiedBy*, *near_antonym*, *derived-pos*, *be_in_state* и *synonym*), уколико се обезбеди аутоматско и интензивно ручно и континуирано проширење онтологије SWN овим релацијама и ентитетима који се повезују овим релацијама, добиће се онтологија која ће бити потпунија и самим тим ефикаснија у решавању постављених задатака. Дијаграм учења онтологије SWN приказан је на слици 6.



Слика 6: Учење онтологије SWN из Корпуса савременог српског језика

Учење онтологије SWN засновано на аутоматској методи проширења семантичке мреже Српски ворднет (Mladenović et al., 2016a) користи анотирани део електронског (дигиталног) корпуса савременог српског језика⁸ (Utvić, 2014) ради генерисања кандидата за проширење релацијама *specificOf* и *specifiedBy*. У овом раду се предлаже полу-аутоматска метода проширења семантичке мреже Српски ворднет заснована на употреби веб алата којим се олакшава повезивање синсетова помоћу неке од лексичко-семантичких релација дефинисаних у SWN. Корисничко радно окружење приказано је на слици 7 и представља алат којим се повезују два жељена синсета једном од релација које су на располагању. На левој страни слике 7, за унету реч (нпр. *lep*) приказују се, испод ње, сва значења ове речи која су дефинисана у Српском ворднету. На десној страни исте слике, за унету реч (нпр.

primerRecsslika), такође испод, приказују се сва њена значења дефинисана у Српском ворднету. Корисник бира одговарајуће значење сваке од унетих речи и повезује их лексичко-семантичком релацијом коју бира из скупа релација датих у средњој колони. Уколико корисник одабере релацију за коју постоји инверзна релација (нпр. *SpecificOf-SpecifiedBy*), аутоматски се генерише и одговарајућа

⁸ <http://korpus.matf.bg.ac.rs/prezentacija/korpus.html>

инверзна релација између синсетова концепата који су одабрани на основу значења.

lep

Prikazi

Poveži relacijom

slika

Prikazi

Synset	Literal Sense	Definition	1 2	Synset	Literal Sense	def	
ENG30-00217728-a	lep	1	Koji ushićuje čula, uzbuđuje duh ili izaziva emocionalno divljenje. <input checked="" type="checkbox"/>	ENG30-13937075-n	slika	4v	situacija koja se tretira kao osmotriv predmet <input type="checkbox"/>
			Relation <input type="checkbox"/>				
			TopicDomain <input type="checkbox"/>				
			causes <input type="checkbox"/>	ENG30-03876519-n	slika	1	Grafička umetnost koja se sastoji od umetničke kompozicije dobijene nanošenjem boje na neku površinu. <input checked="" type="checkbox"/>
			hypernym <input type="checkbox"/>				
			holo_part <input type="checkbox"/>				
			also_see <input type="checkbox"/>	ENG30-03931044-n	slika	2	Vizuelna reprezentacija objekta, scene, osobe ili apstrakcije, proizvedena na nekoj površini. <input type="checkbox"/>
			TopicDomainMember <input type="checkbox"/>				
			be_in_state <input type="checkbox"/>	ENG30-03314028-n	slika	x	Jedna od dvanaest karata iz špila na čijem je licu slika. <input type="checkbox"/>
			Entailment <input type="checkbox"/>	ENG30-07201804-n	slika	4a	Grafički ili živ verbalni opis. <input type="checkbox"/>
			holo_member <input type="checkbox"/>	ENG30-14513489-n	slika	4ax	Okrugljenje u kome se odvija priča ili dramska radnja. <input type="checkbox"/>
			Hyponym <input type="checkbox"/>				
			substanceMeronym <input type="checkbox"/>				
			specifiedBy <input checked="" type="checkbox"/>				
			derived-vn <input type="checkbox"/>				

Слика 7: Веб алат за полу-аутоматско проширење семантичке мреже Српски ворднет; у овом случају повезани су придев *lep* (lep) и именица *slika* (slika) паром међусобно инверзних релација *SpecificOf / SpecifiedBy*

5. Евалуација методе онтолошког препознавања реторичке фигуре *компарација*

У циљу оцене степена препознавања реторичке фигуре *компарација*, у текстовима на српском језику, спроведен је тест изведен над колекцијом различитих текстова која се састоји од 10 дигитализованих дела разноврсног жанра: дечје песме, бајке, комедије, романи и студије наведених у додатку А. Овом скупу придружени су и текстови који су у раду (Mladenović et al., 2016b) коришћени као скупови за тестирање класификатора текстова на основу осећања, а који су састављени од оцена филмова и вести са новинских портала на српском језику (ступци означени лабелама *vesti* и *filmovi* на слици 8). Изграђен је алат којим се, уз помоћ регуларних израза, улазни текст парсира тако да се екстрахују изрази у форми кандидата за препознавање фигуре *компарација*. Регуларним изразима из наведених текстова екстраховане су структуре облика „РЕЧ као РЕЧ {1,2}“, „РЕЧ *попут* РЕЧ {1,2}“. Из екстрахованих структура, ручно су уклоњене све оне које нису облика „РЕЧ као ИМЕНИЦА“,

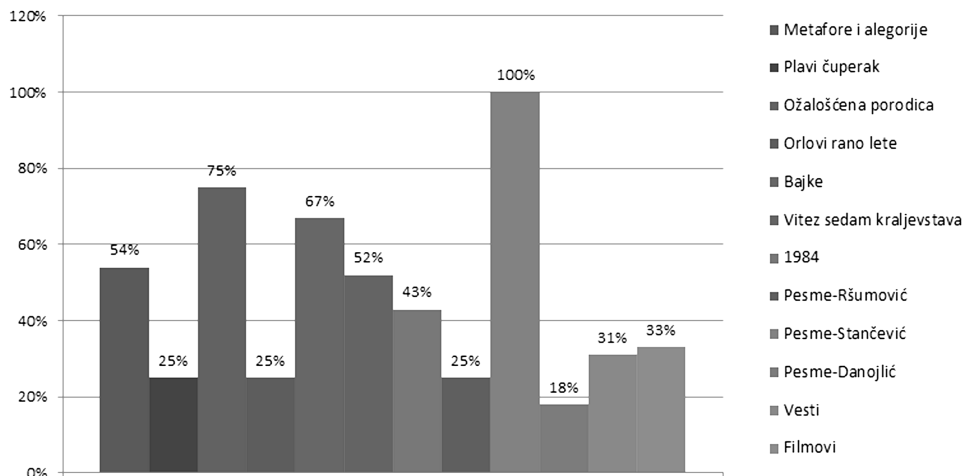
„РЕЧ као ПРИДЕВ ИМЕНИЦА“, „РЕЧ *попут* ПРИДЕВ ИМЕНИЦА“. Најзад, изграђен је и алат који, на основу правила (2) (в. одељак 3.) у онтологији SWN закључује да ли дата структура представља фигуру компарација.

Језичке структуре које су добијене овим тестом указују на то да се у поетским текстовима могу наћи поређења неуобичајена у свакодневном говору. На пример, у песмама Данојлића, се јављају структуре *добар као слон, спокојан као крава, устрептала као стрела* и сл. У песмама Ршумовића могу се наћи: *плаво као ћилим, зелено као тепих, ружан као четка*, итд. На крају овог теста, анализирана је структура скупова за тестирање *filmovi* и *vesti* којима је тестиран SAFOS — систем за анализу осећања текстова на српском језику (Mladenović et al., 2016b), како бисмо оценили могућност за његову даљу доградњу у смислу препознавања фигуративног говора. Резултати су показали ниску, али уједначену тачност⁹ препознавања од око једне трећине укупног броја појављивања фигуре *компарација*. На примерима структура које су екстраховане из ових скупова као што су: *okaraterisan kao triler, poznat kao zemlja, uraden kao nastavak* и сл. може се видети да би у текстовима било могуће детектовати и лажно позитивне кандидате за инстанцирање фигуре *компарација*. Међутим, како је препознавање фигуре вршено помоћу знања у онтологији SWN, постигнута је потпуна прецизност, чиме су лажно позитивни кандидати одбачени¹⁰. С друге стране, непрепознате структуре облика: *oštrim poput brijača, hrabrim kao SUPERMEN, puknuti kao vidik, razmnožavati se kao vinska mušica* указују на низак ниво одзива класификације и потребу испитивања и другачијих језичких структура од описаних, узимањем у обзир и других врста сем именица (личних имена, израза, скраћеница и др.) са једне, као и глагола са друге стране језичке структуре, из којих би онтологија SWN могла учити. Резултати тачности класификације фигуре *компарација*, изведене над скупом улазних текстова који се састоји од 10 дигитализованих дела и два скупа текстова којим је тестиран класификатор текстова на основу осећања, приказани су на слици 8.

Евалуација система приказана је на поступку препознавања фигуре *компарација* у роману „Витез седам краљевстава“ Џорџа Мартина, с обзиром да се ради о најобимнијем интегралном тексту у скупу тестираних текстова. Применом регуларних израза пронађено је укупно 135 конструкција облика „РЕЧ (као|попут) РЕЧ{1,2}“. Од тога је 107 конструкција које представљају

⁹ Тачност (енг. *assiguacy*) представља једну од мера оцене квалитета класификације и представља однос збира тачно позитивно и тачно негативно класификованих примерака и збира свих примерака у тестираној колекцији.

¹⁰ У онтологији SWN не постоји релација *specifiedBy* између именице и придева, осим уколико то није придев верификован методама учења онтологије SWN (в. одељак 4.).



Слика 8: Тачност препознавања фигуре *компаратива* онтологијом SWN у текстовима на српском језику

фигуру *компаратива*, а 28 није носило то значење као у примерима: „*Tu sam pesao kao dečak*“, „*Lim je imao naoštren kolac koji bi mogao poslužiti kao koplje*“, „*Dve godine je služio kao paž*“ и сл. Ова класификација је извршена ручно, а затим је такав скуп подвргнут онтолошкој класификацији. То значи да је за сваки од 135 парова (реч, именица) добијених из датих језичких структура проверено да ли задовољавају правило (2), тј. да ли у посматраном уређеном пару реч представља придев, а затим да ли је уређени пар придев-именица повезан релацијом *specificOf*. Неки примери препознатих фигура компарације су: „*Bele kao kost bile su put i kosa Brindena Rečnog*“... „*Bila su hladna kao kamen, ali ih je bilo divno videti*“, „... *bilo je crveno kao krv*“, а резултати онтолошке класификације приказани су у Табели 1, матрицом конфузије.

На основу матрице конфузије, параметри оцене система дати су у табели 2, а статистичка оцена класификатора представљена је ROC кривом на слици 9, где је добијено да је $AUC > 0,5$ за интервал поверења 95%.

Receiver operating characteristic curve (ROC) је једна од метода евалуације класификатора која користи графички приказ односа сензитивности ($\frac{tp}{tp+fn}$) и специфичности ($\frac{tn}{tn+fp}$) за сваки могући резултат на тесту. На слици 9 приказана је (пуном линијом) ROC крива класификатора фигуре *компаратива*, док се дијагонална линија (приказана испрекиданом линијом) назива дијагоналом случајног исхода и представља исходе случајне класификације. Површина испод

класификација фигуре компарација		ручна класификација	
		да	не
оцена класификатора	да	$tp = 42$	$fp = 0$
	не	$fn = 65$	$tn = 28$

Табела 1: Матрица конфузије система класификације језичких структура које су екстраховане из романа „Витез седам краљевстава“ Џорџа Мартина на оне које представљају фигуру *компарација* и оне које то нису.

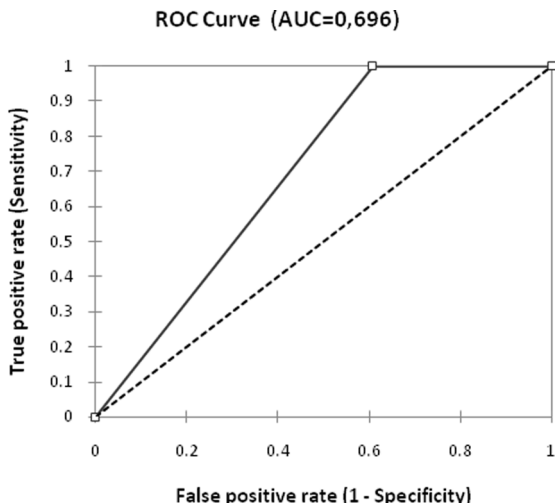
параметри оцене система класификације фигуре <i>компарација</i>	Прецизност = $\frac{tp}{tp+fp}$	Одзив = $\frac{tp}{tp+fn}$	$F_1 = \frac{2PR}{P+R}$	Тачност
Мере	1,000	0,393	0,564	0,518

Табела 2: Параметри оцене система класификације фигуре *компарација* у роману „Витез седам краљевстава“ Џорџа Мартина

ROC криве (енг. Area Under a Curve — *AUC*) интерпретира се као вероватноћа да случајно одабран позитивно означен примерак фигуре *компарација* има већи скор него примерак означен негативно. *AUC* представља меру тачности класификатора и може имати вредности од $AUC = 0,5$ (када се ROC крива поклапа са дијагоналом случајног исхода) до $AUC = 1$ (у случају апсолутног раздвајања класа). Класификација фигуре *компарација* постигла је вредност $AUC = 0,696$.

Анализирањем резултата класификације, могу се уочити разлози ниског одзива. Постоје три групе структура које нису препознате предложеним системом:

1. структуре које садрже личну именицу као у примерима: „*pametan kao princ Eris*“, „*visok kao Dank*“, и сл.;
2. структуре које не садрже придев и именицу, већ глагол и именицу као у примерима: „*Onda se okrenu kao vihor, i jurnu u noć*“ и „*Jaje ulete kao bez duše*“;
3. структуре које садрже придев и именицу, али је поредак такав да није могуће једнозначно детектовати фигуру на основу структура „ПРИДЕВ као/*попут* ИМИЕНИЦА“ и „ПРИДЕВ као/*попут* ПРИДЕВ ИМИЕНИЦА“, већ би се посматрани оквир морао проширити. Примери таквих структура су пронађени у реченицама: „...*gole kao od majke rođene, brašnjave od glave*“



Слика 9: Статистичка оцена система класификације фигуре *компарација* у роману „Витез седам краљевстава“ Џорџа Мартина

do pete“, „...*čije su vode svethlucale crveno i zlatno, sjajne kao ploča kovanog bakra*“, „*Zbog blede kože i kao kreč bele kose ličio je na živi leš.*“

Структуре које садрже личну именицу могу се препознавати укључивањем алата за препознавање именованих ентитета и обрасца претраге „ПРИДЕВ као ВЛАСТИТА_ИМЕНИЦА“. Оне које садрже глагол и именицу могу се идентификовати употребом алата за означавање врстама речи и обрасца „ГЛАГОЛ као ИМЕНИЦА“, а оне измењеног поретка могу се препознавати обрасцима који то узимају у обзир, нпр. „као ИМЕНИЦА [ВЕЗНИК|ПРЕДЛОГ]? ПРИДЕВ“.

6. Закључак

У овом раду предложена је метода препознавања реторичких фигура из групе фигура речи (тропи) које се не могу препознавати на основу синтаксичко-морфолошких образаца. Метода користи правила дефинисана у онтологији SWN, како би препознала да ли језичка структура екстрахована из текста задовољава постављено правило и на тај начин може бити означена као одговарајућа реторичка фигура. Метода је тестирана откривањем реторичке фигуре *компарација* над скупом улазних текстова који се састоји од 10

дигитализованих дела и два скупа текстова којима је претходно тестиран класификатор текстова на основу осећања, а статистичка оцена система спроведена је над текстом романа „Витез седам краљевстава“ Џорџа Мартина. Резултати статистичке оцене показали су да класификација овом методом ($AUC = 0,696$) у случају реторичке фигуре *компарација* показује осредње резултате, али прецизност класификације од 100% указује на то да се даљим учењем онтологије, на основу предложене полуаутоматске методе, могу побољшати резултате класификације ове фигуре.

У даљем раду, када се ради о фигури *компарација*, биће уведени обрасци описани у одељку 5. ради препознавања структура које нису препознате предложеним системом. Када је реч о другим фигурама, радиће се на изградњи скупова за тестирање класификатора реторичких фигура *иронија*, *сарказам* и *оксиморон*, како би позитивно класификоване структуре могле бити коришћене у процесу изградње предиктора класификатора текста на основу осећања. Радиће се и на проширењу скупа онтолошких правила тако да укључују поступке генерализације и спецификације концепата који већ задовољавају правила постављена овим системом.

Литература

- 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 Sagion. 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 Associ-

- ation. ISBN 2-9517408-6-7. URL <http://www.lrec-conf.org/proceedings/lrec2010/index.html>.
- Paula Carvalho, Luís Sarmiento, 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 Sarmiento, 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 Odiijk, 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 Misrepresentation 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, Iași, 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 Relation 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>.

А Прилог

Колекција текстова од 10 дигитализованих дела разноврсног жанра (дечје песме, романи, бајке, комедије и студије) којом је тестиран метод препознавања реторичке фигуре компарација.

1. Antić, Miroslav. (1965). Plavi čuperak, преузето 25.02.2016, <http://www.kotraza-dragacevo.org/Knjige.html>
2. Bajke Za Decu, преузето 25.02.2016, <http://www.zadecu.com/Bajke>
3. Ćopić, Branko. Orlovi rano lete, преузето 25.02.2016, <http://www.kotraza-dragacevo.org/Knjige.html>
4. Danojlić, Milovan. (1966). Kako spavaju tramvaji, преузето 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, преузето 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”, преузето 25.02.2016, <http://www.orfejsu.com/deciji-osmeh-sreten-stancevic-zbirka-decijih-pesama/>