

РАДОСЛАВА ТРНАВАЦ\*  
Филозофски факултет Универзитет у Новом Саду  
Нови Сад, Србија

## СИСТЕМ ЕВАЛУАЦИЈЕ У РУСКОМ ЈЕЗИКУ – ПРОГРАМ ИСТРАЖИВАЊА

У раду се разматра пројекат у оквиру кога би се креирао велики дигитални корпус руског језика са анотацијама евалуативних израза. Предлаже се анотација корпуса која би укључивала следећу информацију: изразе става, евентуалности, неверидикалности, граматичке обрасце субјективности, лексикон оцене, дискурсивне феномене, имплицитну евалуацију, фигуративни језик и интенцију аутора текста.

**Кључне речи:** анотација корпуса, евалуација, сентимент анализа, дискурсивна и прагматичка информација

Током година, релативно бројне студије су се појавиле на тему разлике између објективног и субјективног језика (Wiebe 1994; Martin and White 2005; Trnavac 2006; Hunston and Thompson 2000, Pang and Lee 2008), али до сада није дошло до стварања унифициране теорије о томе како су евалуација и субјективност изражени у тексту. Циљеви истраживања будућег пројекта на руском језику би могли бити дефинисани на следећи начин:

- (1) Са теоријске тачке гледишта потребна је кохерентна теорија евалуације која може бити примењена на међујезичком нивоу;
- (2) Са апликативног становишта потребан је (а) велики дигитални корпус руског језика који би био анотирани у смислу субјективног израза; (б) софтвер који би био креиран за руски језик ради калкулације сентимента у необрађеним текстовима.

Импликације оваквог пројекта могу бити вишеструке. У научном смислу, студије субјективног израза почев од индивидуалних речи па до дискурсивног нивоа ће интензивирати теоријске и когнитивне перспективе

---

\* [radoslava.trnavac@ff.uns.ac.rs](mailto:radoslava.trnavac@ff.uns.ac.rs)

изучавања евалуације као језичког феномена. Рачунарски приступи језику ће такође имати користи од оваквог истраживања, посебно област сентимент анализе и екстракције мишљења аутора који су, на пример, повезани са анализом друштвених мрежа, анализом политичких тема или пословних података. Такав пројекат би такође допринео области корпусне лингвистике са аотираним корпусом субјективног израза, с обзиром да руски језик још увек заостаје у дигитализацији у односу на неке друге велике језике, типа енглеског, шпанског, кинеског, итд. Најзад, резултати датог истраживања би могли да буду употребљени у области методике наставе при обучавању страним језицима.

Нама није познато ниједно истраживање овог типа које је било спроведено на материјалу руског језика. Са информатичке тачке гледишта, у руском језику постоје радови који се баве сентимент анализом (Chetviorkin and Loukachevitch 2012), али теоријско-лингвистичка истраживања која би се бавила, на пример, односом дискурсивне структуре и евалуације, према нашим сазнањима, до сада нису спроведена. Сентимент анализа има снажну повезаност са социјалним медијима (Farzindar and Inkpen 2015), као и са моделирањем људског понашања, односно спровођењем сентимент анализе у реалности. Сентимент се одређује према детекцији намере, области која пружа лингвистичким објектима предиктивну моћ. На пример, у оквиру рачунарске лингвистике се може вршити предвиђање понашања гласача, као и резултата избора (Yano et al. 2013; Qiu et al. 2015), предикција лажи (Fitzpatrick et al. 2015), или интенција да се нешто купи (Ding et al. 2015). Предикције такође могу бити одређене на бази екстралингвистичких извора информације, као што су то, на пример, карактеристике аутора и њихове онлајн интеракције (Qiu et al. 2015). У будућности ће предикције везане за текстуалне кориснике морати да интегришу моделе из области вештачке интелигенције са технологијама природног језичког процесирања да би се пронашли специфични сигнали намере, као што су то промене у аргументативном низу или социјалном односу између учесника дискурса, промене топика или веровања корисника, изненадно коришћење сентимента или емоција одређеног типа, корелација између жанра и коришћења специфичних лингвистичких инструмената. У даљем тексту биће речи о томе како је евалуативни језик укључен у процес сентимент анализе.

Евалуативни аспект језика изражава осећања, оцене људи, као и оцене о ситуацијама и објектима. У лингвистици различити правци се баве евалуацијом, као што су теорија оцене (Martin and White 2005), става (Biber and Finegan 1989), евалуације (Hunston and Thompson 2000) и неверидикалности (Taboada and Trnavaц 2013). У оквиру информатике савремена истраживања се фокусирају на аутоматску екстракцију мишљења

у оквиру три нивоа (Liu 2012) – документа, реченице и домена. Први ниво класификује целокупна документа као позитивне и негативне. Други се бави субјективном оријентацијом и поларитетом мишљења (позитивног или негативног), који су представљени редоследом речи субјективног карактера у реченици. Трећи ниво се заснива на екстракцији мишљења базираног на одређеним текстуалним доменима или жанровима. Методе за екстракцију које се употребљавају у сваком од наведена три нивоа се заснивају на различитим приступима, укључујући употребу репрезентација базираних на субјективним и граматичким везама. Овакве врсте модела се баве комплексним језиком који укључује негацију, спекулацију, као и различите феномене условљене контекстом. Постоји велики број прегледа евалуативне литературе у области рачунарске и примењене лингвистике. Ханстон и Томпсон (2000) демонстрирају у својој књизи Евалуација у тексту како евалуативни изрази могу бити анализирани на лексичком, граматичком и текстуалном нивоу. Недавна студија Томсона и Албе Јуаз (Thompson and Alba-Juez 2014) се фокусира на теоретске и емпиријске студије евалуативног текста у оквиру фонолошког, лексичког и семантичког нивоа, као и на различите текстуалне жанрове и контексте. Рачунарски приступи евалуативном тексту, који се сврставају у ред сентимент анализе, највише су обрађени у студијама Панга и Лија (Pang and Lee (2008) и Лија (Liu, 2010, 2012).

У даљем тексту описаћемо лингвистичке аспекте и теорије евалуативног језика које би требало размотрити пре анотације корпуса на руском језику. Такође ћемо изложити како дискурсивна и прагматичка информација могу да помогну да се изађе изван оквира локалних контекстуалних феномена уз ослањање на лексикон поларитета, као и на површинску односно дубинску синтаксичку анализу. У наредном тексту ћемо поменути следеће теме:

- Комплексну лексичку семантику евалуативних израза који укључују семантичку категоризацију;
- Контекстуалне ефекте који настају захваљујући негацији, модалности и невердикалности;
- Дискурсивну структуру и друге блиске дискурсивне методе;
- Комплексне евалуативне језичке феномене као што су имплицитна евалуација, фигуративни језик (иронија, сарказам) и детекција намере;
- Екстралингвистичке информације, као што су то, на пример, друштвене мреже.

Као што се види из претходно наведеног текста, обухватићемо тематику како феномена који се тичу локалног и реченичног контекста, тако и

главног дискурса и контекста за који верујемо да треба да буде поменут да би се адекватно описала тематика евалуације.

### 1.1. Став

Један од најранијих покушаја да се опише евалуативни језик се може наћи код Бајбера, Финегана и др. У различитим радовима (Biber and Finegan 1988, 1989; Conrad and Biber 2000), ови аутори описују став као израз ауторовог односа, осећања и судова, као и његову приврженост поруци коју изражава. Став, у овом случају, обухвата средства која изражавају евентуалност, као и ауторову приврженост поруци и афект (позитивну и негативну евалуацију). Иницијални фокус (Biber and Finegan 1988) је био на адвербијалима (речима, као што су искрено, наравно, очигледно). Касније, Бајбер и Финеган (1989) су додали придеве, глаголе и модалне речи као маркере евентуалности и афекта. У њиховим радовима се врши класификација маркера на основу група ставова. Дате групе могу изражавати „емфатичку експресију афекта”, „експресију сумње” или „безличност”. Сваки од ставова се употребљава у одређеној врсти текста, као што су, на пример, лична писма или литерарни романи. С друге стране, текстови који одражавају став „безлично” могу се наћи у следећим жанровима: академска проза, преса, радио емисије, биографије, официјелни документи и писма, итд. Поменута врста жанровске класификације је креирана на основу чињенице колико је аутор текста укључен у структуру текста. Ако вршимо анализу веома субјективног жанра, онда присуство субјективних и евалуативних речи у њему није изненађујуће. С друге стране, за жанрове које Бајбер и Финеган називају „безличним”, висок ниво евалуативне лексике је далеко значајнији када се појави. Термин став у раду Бајбера и Финегана је близак нашој употреби термина евалуативни језик.

### 1.2. Евентуалност

Евентуалност обухвата истраживање лингвистичког кодирања односа према знању (Chafe and Nichols 1986). Она изражава три основна значења (Chafe 1986; Boye and Harder 2009): (1) поузданост информације (означава се прилозима као што су *можда*, *ситурно*), (2) начин на који се долази до сазнања (уверење које се изражава конструкцијама као што су *мислим*, *прећийостављам*); (3) индукцију која се исказује облицима, као што су *морати*, *изгледа*, *очигледно*; и (4) дедукцију изражену облицима требати, моћи, вероватно. Извор информација може бити формулисан глаголима који указују на различите канале перцепције

(видети, чути, осећати). Феномену евентуалности није посвећена велика пажња у литератури која се бави субјективним изразом, иако феномен евентуалности обухвата скуп средстава којим се утврђује поузданост размишљања и припада домену спекулације.

### 1.3. Невередикалност

У појединим радовима ови облици се називају иреалисом. Иреалис у глобалу означава изразе који указују на догађаје који нису фактички. Невередикалност је нешто шира по свом значењу јер укључује све контексте који нису веридикални, односно који се не ослањају на истинитост или који у реалности не постоје (Giannakidou 1995; Zwarts 1995). Оператори који изражавају неверидикалност укључују негацију, модалне глаголе, менталне глаголе (*веровајши, мислиши, желеши*), императиве, питања, протазе у оквиру кондиционала, као и коњуктив у језицима који га поседују (Trnavac and Taboada 2012). Невередикалност је другачија од евентуалности у смислу тога да маркери који изражавају евентуалност могу кодирати неверидикална значења, али такође и различите нијансе у оквиру веридикалних пропозиција. Говорник може да представља нешто као чињеницу (дакле, веридикално значење), али у исто време може да се дистанцира од поузданости изјава кроз маркере евентуалности (на пример, Студенти кажу да је предавање било добро). Невередикалност је релевантна у сентимент анализи јер евалуативни изрази у оквиру неверидикалног оператора немају стабилно значење. Они понекада могу да представљају супротни поларитет од њиховог основног, у ситуацијама када су евалуативни изрази у оквиру домена негације, односно могу да примају утицаје од неверидикалних оператора. Присуство кондиционала као неверидикалних оператора је навело лингвисте на идеју да истражују однос између кохерентности и евалуативног језика (Trnavac and Taboada 2012; Asher et al. 2009).

### 1.4. Субјективност

Истраживачи који користе овај термин проучавају домене деикси-са, локативних израза, као и употребу модалних глагола у оквиру других феномена (Langacker 1990). Везе са сентимент анализом су очигледне јер епистемички модали преносе субјективно значење и повезани су са евентуалношћу. Вајт (2004) је писао о субјективности као о тачки гледишта, и део његовог истраживања се одвијао у оквиру модела теорије оцено. Вајт расправља о опозицији између објективних и субјективних израза у дискурсу медија, те како субјективни изрази могу бити исказа-

ни кроз асоцијације, метафоре и инференце. Анализа оваквих облика је заиста важна у сентимент анализи, али се до ње тешко долази пошто се ти облици ослањају на информације које долазе из света који нас окружује. Вибе и колеге су посветили велику пажњу индикаторима субјективности у реченицама (Wiebe and Riloff 2005; Wiebe et al. 2004), те су предложили скуп сигнализатора субјективности, од којих су неки лексички а неки синтактички. Међу лексичким сигнализаторима посебно су значајни глаголи који указују на психолошка стања, као и глаголи и придеви који указују на оцену личности или предмета.

### 1.5. Евалуација и граматика образаца

Под термином евалуације може се разматрати разграната област истраживања која проучава евалуацију и мишљење у оквиру функционалног модела. Најбољи пример такве врсте истраживања је зброник Ханстона и Томсонове (2000). Ханстон и Томсонова (2000) у предговору ове књиге предлажу два аспекта евалуативног језика – модалност и евалуацију, оцену или перспективу. Модалност, по мишљењу ових аутора, изражава мишљење о пропозицијама, као што је то, на пример, вероватноћа (Могла би да пада киша). Модалност је углавном граматикализована. Евалуација, са друге стране, изражава мишљење о ентитетима, и већином се изражава придевима. Ханстон и Томсонова су анализирали неке од главних приступа теми субјективности, и запазили су да постоје правци који раздвајају модалност од евалуације (Halliday and Matthiessen 2014; Martin 2000; White 2003). Други истраживачи комбинују два израза размишљања, обично под једним називом, кроз термине став односно перспектива (Biber and Finegan 1989; Conrad and Biber 2000). У сентимент анализи модалност је укључена као један од феномена који утиче на евалуативни језик. Ханстон и Томсонова (2000) користе комбиновани приступ и предлажу јединствени термин за оба аспекта евалуације. Аутори се залажу да евалуација има три главне функције: (1) да изрази мишљење писца или говорника, (2) да изгради и одржи релације између говорника/писца и слушаоца/читаоца, и (3) да организују дискурс.

О првим двама функцијама се доста говорило у оквиру сентимент анализе са нагласком на прву тему, и појединим истраживањима у оквиру друге теме, у смислу тога како писац представља, манипулише и резимира информацију. То је посебно истражено у оквиру система ангажмана који представља део модела оцене (Martin & White 2005). Трећи аспект је добио најмању пажњу у оквиру обраде евалуације језика. Ханстон и Томсонова (2000) представљају дату функцију као систем аргументације. Писац не само да изражава мишљење, ангажује читаоца,

већ и представља аргументе одређеним редоследом. Евалуација на крају дискурсивних јединица, као што су то параграфи, указује да је главни смисао аргументације изречен, и да писац сматра да читалац прихвата ту тачку гледишта. Генерално, Ханстон и Томсонова се залажу да је евалуација изражена колико текстом, толико и индивидуалним лексичким изразима и граматиком. Ми сматрамо да поред локације у тексту и других текстуалних и дискурсивних карактеристика и односи кохерентности играју улогу у интерпретацији евалуације.

Рад на језику евалуације се посебно интензивира у неколико последњих година. Поменућемо зборник Томсона и Алба-Језа (2014). Овај зборник ставља још већи акценат на евалуацију која се базира на целом тексту. Наслов зборника је Евалуација у контексту, а радови у њему демонстрирају свепрожимајући утицај евалуације, као и њену везу са емоцијама. У уводу зборника, Алба-Јез и Томсон подвлаче да се евалуација провлачи кроз све домене језика, укључујући фонолошки ниво који је изражен кроз интонацију и висину тона, морфолошки ниво који је исказан суфиксима у језицима који имају развијену морфологију, затим лексички ниво који је детаљно обрађен у сентимент анализи, синтактички ниво, као и систем модалности и неверидикалности, али и редослед речи и структурални аспекти текста. Семантички и прагматички ниво такође играју значајнију улогу. Поменути нивои се посебно односе на примере евалуативних израза који се значењски не могу разложити и који могу зависити од контекста. Зборник јасно указује да фокус само на лексичком нивоу не може донети задовољавајуће резултате у смислу укупне дескрипције евалуације у језику.

## 1.6. Модел оцене

Модел оцене припада системско-функционалној традицији коју је започео Халидеј (Halliday and Matthiessen 2014) и коју су развијали Џим Мартин, Питер Вајт и колеге (Martin 2000; Martin and White 2005; White 2012; Martin 2014). Мартин (2000) карактерише оцену као скуп средстава који се користе да би се изражавале емоције, судови и вредности заједно са средствима за амплификацију и за изражавање ангажмана говорника. Мартин, такође, сматра да средства за изражавање оцене креирају сопствени систем у оквиру језика (систем у оквиру системско-функционалне лингвистике), и да је систем оцене подељен на три различита подсистема, а то су: став, градација и ангажман.

Централни аспект модела оцене представља подсистем става који је опет подељен на сопствена три подсистема – афект, суд и естетску евалуацију. Афект се користи да би се окарактерисао емотивни став говорника или нечија туђа реакција (срећа, туга, страх). Суд симболизује

евалуацију морала особе која није говорник (етичан, лажан, храбар). Естетска евалуација обухвата естетски квалитет објеката или природних феномена (изванредан, елегантан, иновативан). Рачунарска обрада модела оцене обично укључује експлицитно присутне речи у тексту са позитивним или негативним значењем. Постоје евалуације које су у тексту имплицитно изражене. На пример, фраза *шаленџовано деџе* исказује експлицитну оцену. С друге стране, суд *деџе које чиџа мноџо* представља имплицитну оцену. С обзиром да је лакше експлицитну оцену идентификовати у рачунарским апликацијама, рачунарска лингвистика се углавном њоме и бави. О имплицитној евалуацији ћемо неколико речи рећи у следећем одељку, у који ћемо укључити и информације о метафори, иронији и сарказму. Осим централне теме која је представљена моделом става, Мартин и Вајт (2005) сматрају да друга два система играју круцијалну улогу у изражавању мишљења. Систем градиције је одговоран за говорникову способност да интензивира или ослаби снагу сопственог мишљења, те има силу или фокус као подсистеме. Сила подразумева интензивирање или слабљење речи чије се значење може окарактерисати кроз меру, док фокус репрезентује начин на који говорници могу да изоштре или омекшају речи које изражавају објекат који има значење немерљиве количине. Интензивирање се може демонстрирати примером *џрилично инџересанџан* док се процес слабљења може илустровати примером *џомало џужан*. У оквиру фразе *џрави џријаџељ* оригинално значење речи *џријаџељ* не укључује у себи семантику мере, али се кроз придев *џрави* то значење ставља у фокус. Са друге стране, фраза *неџџо џоџуџ џријаџеља* умекшава значење речи *џријаџељ*.

Систем ангажмана представља скуп лингвистичких опција које дозвољавају да појединац преноси степен његове или њене привржености мишљењу које је презентовано.

Постоји фундаментална разлика између хетероглосичних и моноглосичних израза који следе класификацију Бахтина (1981). У оквиру хетероглосичних израза, међусубјективна позиција је отворена јер искази препознају, одговарају, предвиђају, ревидирају или проблематизују масу конвергентних и дивергентних алтернативних исказа (Martin and White 2005; White 2012, 2003). Друга опција је представљена моногловијом где не постоји алтернативна перспектива, односно постоји отвореност да се прихвати постојећа перспектива. Моноглосични искази се презентују као чињенице.

Систем оцене је детаљно представљен у великом броју радова на енглеском језику у којима се анализа врши у оквиру различитих текстуалних жанрова. Овај модел пружа веома богату дескрипцију различитих аспеката евалуације и субјективног позиционирања. Почев од раних фаза развоја сентимент анализе истраживачима је било јасно да модел



оцене може бити од помоћи да се идентификују типови мишљења и да се укаже на допринос фраза које интензивирају или редукују иницијално значење речи (Polanyi and Zaenen 2006). Велико интересовање за модел оцене је условило појаву публикација које се баве питањем како да се аутоматски идентификују изрази који би карактеризовали три подтипа система става (афект, суд и естетску оцену). Потенцијалне могућности овог модела су велике. Позитиван или негативан евалуативни израз је информативан сам по себи, али још више ако ми знамо да ли се односи на личне емоције, мишљење о другима или на евалуацију објеката.

### 1.7. Сентимент анализа

У овом одељку ћемо описати како се сентимент, односно евалуација, изражава почев од нивоа речи завршно са контекстом. Такође ће бити речи како се евалуација исказује на нивоу речи, фраза и на реченичном нивоу, као и на нивоу дискурса и контекстуалних феномена. Проналазак субјективних речи (као и њиховог основног поларитета) представља активну сферу истраживања у којој су методолошки приступи базирани на корпусу и на лексикону. Постоји неколико ручно или аутоматски креираних лексичких средстава. Већина њих поседује ограничену специфику домена и језика и групише евалуативне изразе дуж бинарне позитивне или негативне осе. Под претпоставком да слична терминологија има сличну емотивну или субјективну оријентацију, технике лексичке експанзије повећавају иницијални скуп субјективних семених речи кроз различиту метрику семантичке сличности. Ове технике такође могу да активирају односе са речима као што су синоними, антоними, и хипероними кроз општа лингвистичка средства попут WordNeta или синтаксичке особине попут односа зависности. Свега неколико студија предлаже да се увећа бинарна категоризација евалуативних средстава. Радови у овој области могу бити подељени на начин на који се категоризују лексички изрази – кроз интензитет емоција или кроз синтаксичке или семантичке принципе класификације, укључујући ту и модел оцене.

Један начин да се побољша детерминација поларитета јесте да се повеже било која субјективна јединица са евалуационим поенима или нивоом интензитета. SentiWordnet (Baccianella et al. 2010) представља продужетак WordNeta (Fellbaum 1998) који има могућност да придружи сваком смислу три сентимент поена: позитиван (P), негативан (N), и објективан (O). Дакле, различите врсте смисла које покривају исти термин могу имати различити број сентимент бодова. Иако је SentiWordNet био успешно искоришћен да би се добила оријентација на нивоу документа (Manoussos et al. 2014), алгоритми који служе за разграничење

смисла се често употребљавају да би се нашао прави смисао за потребни термин.

Други начин да се побољша информација о сентименту у тексту сем кодирања поларитета или интензитета субјективне речи јесте и кодирање информације о семантичкој категорији евалуативног израза. Категорије могу бити одређене на основу класификација базичних емоцијама као што су бес, страх, изненађење или љубав (Osgood et al. 1957; Izard 1971; Russell 1983; Ekman 1984; Ortony et al. 1988). Ворднет-афект (WordNet-Affect) је такође средство предвиђено за лексичку репрезентацију афективног знања (Strapparava and Valitutti 2004). У оквиру овог система сваки афективни термин прати класификацију модела емоција које су понудили Ортони и др. (1998). Други извори афективне лексике су „сентикнет” (SenticNet, Poria et al. 2013) и база информација „емотинет” (EmotiNet, Balahur et al. 2011) који повезују поларитет и афективну информацију са афективним ситуацијама, као што су достизање циља, падање на испиту и др. Моделирање таквих ситуација је од посебне важности за препознавање имплицитних информација (Balahur et al. 2012). Поред очигледне корисности афективне лексике при детекцији емоција, употреба истих у сентимент анализи је од посебног значаја, као и при детекцији фигуративног језика (Reyes and Rosso 2012).

Додатне категорије су предложене у литератури које се баве комплексном лексичком семантиком евалуативних израза. Поједине категорије су синтаксички и семантички базиране, док су друге искључиво ослањају на семантику. Левин (1993) класификује преко 3,000 енглеских глагола према заједничком значењу и синтаксичком понашању. Ова класификација истражује понашање глагола у односу на велики скуп синтаксичких алтернација које рефлектују значење глагола. Неколико класа су значајне за сентимент анализу, као што су глаголи са семантиком оцене или глаголи који описују психолошко стање. Матијеу (2005) нам предлаже семантичку класификацију сентимента у којима су глаголи и именице подељени у оквиру 38 семантичких категорија у француском језику, према свом значењу (љубав, страх, чуђење, итд.). Аутор датог истаживања истиче да синтаксичке структуре утичу на интерпретацију евалуативних израза, те диференцира три класе глагола: (1) оне које означавају искуство или каузацију непријатног осећања, (2) пријатног осећања, (3) као и оне које су неутралног карактера. Семантичке класе су повезане значењем, интензитетом, и анатомијом односа. Оне повезују скуп лингвистичких особина са речима или класама и граде семантичке репрезентације. Матијеу и Фелбаум (2010) су проширили дату класификацију на енглеске глаголе који описују емоције. СентиФрејмнет (SentiFrameNet, Ruppenhofer and Rehbein 2012) је вероватно најбољи пример синтаксичко-семантичке категоризације евалуативних

израза. Он представља проширену верзију фрејмнета (FrameNet, Baker et al. 1998), повезујући изворе мишљења, усмеравајући се на семантичке улоге и додајући семантичке карактеристике, као што су поларитет, интезитет, дејство које описује промену стања учесника догађаја. Евалуативни језик није самостално описан у фрејмнету, али неколико фрејмова је значајно за анализу сентимента, као што су ОЦЕНА, МИШЉЕЊЕ, УСМЕРЕНА ЕМОЦИЈА и семантичке улоге попут ЕКСПЕРИЕНЦЕРА. Сваки фрејм се повезује са скупом лексичких јединица које се састоје од речи које евоцирају дати фрејм.

За суптилинији рад на сентимент анализи свакако су значајни тзв. „мењачи валентности” (valence shifters) који су први пут поменути у раду Полањи и Заенена (2006). Дати елементи описују како евалуативни изрази могу бити модификовани уз помоћ контекста, у шта се укључују екстрапропозициони аспекти значења, као што су интензификација, смањивање вредности, пресупозиције, дискурс и иронија. У наредном тексту ћемо посветити пажњу „мењачима валентности” који могу да утичу на евалуативне изразе на реченичном или субреченичном нивоу. Рад са овим елементима укључује три задатка: идентификацију ових израза и њиховог поља деловања, анализирање њиховог утицаја на евалуацију, као и израчунавање њиховог композитивног сентимента у комбинацији са евалуативним изразима уз балансирање тог ефекта са претходним поларитетом субјективне речи.

Ефекат мењача валентности на сентимент анализу може бити двојак – интензификација и смањење евалуативне вредности субјективне речи. Опши термин „интензификатор” се користи за средства која мењају интензитет индивидуалне речи, независно од тога да ли подижу или смањују вредности евалуатора. Многа средства интензивирају евалуацију. На пример, придеви могу да интензивирају или да „спусте” вредност евалуације именице коју прате (*велики усџех*). Перифрастични изрази могу такође да утичу на интензитет других речи. На пример, они могу да се појаве као додаци или модификатори глагола (*јонашао се лејо*), модификатори именица (*само јензионери*), придеви (*веома ојасан јуш*), прилози (*веома лејо*) и клаузе (*несумњиво да је био у јраву*). Прилози степена и начина су највише изучавани у литератури јер су они веома релевантни за субјективност у тексту. Њихово узимање у обзир је констатно показивало да се побољшава продуктивност система за анализу (Venamara et al. 2007; Taboada et al. 2011). Ефекат интензификатора или „умањивача” евалуативног језика је генерално обликован у смислу линеарног модела који користи систем сабирања или одузимања. На пример, ако позитиван придев има вредност 2, позитивно амплифициран придев може имати вредност 3, а његова умањена вредност може износити 1. Метод који се обично користи да би се модификовале дате разлике је

метод множења вредности. На пример, Табоада и др. (2011) користе за интензификаторе процентуалну скалу, предлажући следеће вредности: за суперлатив + 100%, за модификатор прави (*прави јосјодин*) +25%, за прилог заиста (*заиста слабо*) + 15%. Постоје и аспекти значења који преносе информацију која се односи на натпропозиционални садржај клауза или реченица, али који још не припадају домену синтаксе нити дискурса. Неверидикалност (Giannakidou 1995, Zwarts 1995), о којој је било речи у претходном тексту, припада таквој појави. Она може да се користи да би се изразила могућност, неопходност, дозвола, облигација или жеља, и граматички је изражења кроз адвербијале (*можда, сигурно*), кондиционал, стање глагола, неке модале (*морати, моћи*), као и поједине менталне глаголе (*мислити, веровати*). Придеви и именице такође могу изражавати модалност (*вероватан начин*). Општеприхваћена је чињеница у оквиру сентимент анализе да су неверидикалност и иреалис непоздани у изражавању сентимента у реченицама које их изражавају (Trnavaц & Taboada 2012). Недовољно истраживања је спроведено у области која би испитивала утицај невердикалних маркера на евалуативне изразе. Лиу и др. (2014) аутоматски путем одређују да ли су мишљења изражена модалношћу у реченицима позитивна, негативна или неутрална. Негација је такође лингвистички феномен који утиче на евалуативне изразе на локалном нивоу (Trnavaц & Taboada 2012). Негација може да буде коришћена да би се оповргли искази или ентитети. Она, на пример, може бити граматички изражена префиксима (не-човек) или кроз изразе који сигнализују негативни поларитет, као што су било који, било шта, икад. Речи које указују на негативни поларитет се обично појављују у негативним реченицама, али не и у њиховим кореспондирајућим позитивним примерима. Негативна евалуација може бити исказана коришћењем именица или глагола које имају негацију као део њихове лексичке семантике (*елиминисати*). Такође може бити изражена имплицитно без икаквог коришћења негативних речи, као у следећој реченици:

(1) Овај софтвер је испод мојих очекивања.

Негативни елементи се могу груписати на више начина. У неким језицима, вишеструка негација отказује ефекат негације (*Ово јело никад не може да вас разочара*), док у другим језицима вишеструка негација обично интензивира ефекат негације. У поређењу са облицима који обележавају негацију, речи које указују на негативни поларитет контекста и вишеструку негацију представљају мање истражену област у литератури која се бави сентимент анализом.

Табоада и др. (2011) анализирају изразе који указују на негативни поларитет (као и на модалност) у енглеском језику, сматрајући их блокторима иреалиса јер игноришу семантичку оријентацију субјективних речи у њиховом пољу. С друге стране, Бенамара и др. (2012) сматрају да изрази који сигнализирају негативни поларитет у француском језику појачавају снагу субјективних речи који се налазе у пољу њиховог дејства.

Још један значајан аспект везан за негацију је њена маркираност. Негативни искази се обично посматрају као маркиранији него њихови афирмативни дупликати и у психолошком и прагматичком смислу (Horn 1989). Истраживања у сентимент анализи су показала да прецизно одређивање негативног сентимента је теже јер се користи мање негативних термина и због тога што се у многим језицима негативна евалуација формулише у виду позитивне терминологије (Pang and Lee 2008). Поједини радови третирају негацију као егзактни преокрет позитивног израза (Polanyi and Zaenen 2006). По другим ауторима (Taboada et al. 2011), негација не може да буде редукована на егзактно опозитни поларитет. На пример, ако узмемо да је бодовање придева одличан +3, онда бодовање истих у реченици Тај рад није одличан, не може бити -3. Та реченица само означава да рад није довољно добар. Могућа солуција би била да се ефектом негативне речи промени негирани термин на скали за одређени степен, али без креирања тоталне опозиције оригиналној речи (Taboada et al. 2011; Chardon et al. 2013a).

У следећим одељцима ћемо кратко описати пет контекстуалних феномена који ће представљати кључне појаве у систему сентимент анализе у наредном периоду: дискурс, имплицитна евалуација, фигуративни језик, екстралингвистичка информација и детекција интенције.

#### а. Дискурсивни феномени

Текстови и конверзација не представљају егзактне еквиваленте речима и реченицама. Они су структурно организовани у дискурсивне јединице које су повезане једна са другом да би обезбедиле дискурсивну кохерентност и кохезију. Кохерентност се односи на логичку структуру дискурса у којој сваки део текста има функцију и улогу у односу на друге делове у тексту (Taboada and Mann 2006). Кохерентност се тиче семантичких или прагматичких релација међу јединицама да би се произвело свеукупно значење у дискурсу (Hobbs 1979; Grosz et al. 1995). Утисак кохерентности у тексту (*да је организован и да се држи заједно*) је такође потпомогнут кохезијом, односно повезивањем ентитета у дискурсу (Halliday and Hasan 1976). То се дешава граматичким и лексичким средствима, таквим као што су анафора и лексички односи (синонимија,

меронимија, хипонимија). У сентимент анализи дискурсивна структура обезбеђује круцијалну везу између локалног реченичног нивоа и целог документа (чланка, конверзације, итд.) и потребна је за боље разумевање изражавања субјективности у тексту. Дискурс посебно може да делује у оквиру три правца: (1) кроз идентификацију субјективности и оријентацију поларитета евалуативних израза; (2) кроз издвајање релевантних сигнала за препознавање имплицитне евалуације и (3) кроз оцену комплетне перспективе текста. Два правца делују у оквиру дискурсивних феномена – макроорганизација текстуалних образаца високог нивоа и конструисање комплексних дискурсивних јединица које су начињене од елементарних дискурсивних јединица у рекурзивном облику. Први правац се бави дискурсивним сегментима као јединицама већим од реченице (параграфи, јединице са топицом) и фокусира се на креирање топикализоване или функционалне структуре (Stede 2011). Организација дискурса преко топика се одвија током текстуалне сегментације дискурса у оквиру линеарних израза од којих се сваки фокусира на посебни субтопик који се појављује у контексту једног или више главних топика. Сегментација топика се дешава уз помоћ локалног дискурсивног континуитета или уз помоћ лексичке кохезије која претпоставља да су топик и лексичка употреба (понављање речи, дискурсивни везници и парадигматички односи) веома повезани.

Аргументација представља још један аспект који може да игра важну функционалну улогу у евалуативним текстовима. Наиме, евалуативни језик може да послужи у креирању аргумената. Ханстон и Томсонова (2000) предлажу да евалуација помаже да се организује дискурс, поред улоге коју има у презентацији мишљења. Аргументација је процес у коме аутор гради аргументе да би појаснио или одбранио своја мишљења. Аргумент је генерално дефинисан као скуп премиса које обезбеђују доказ за или против неких закључака. Премисе тих аргумената могу бити уведене у текстове специфичним маркерима као што су али, или, јер. Праћење аргумената у тексту доприноси идентификацији аргументативне структуре, укључујући идентификацију премисе, закључака и везе између њих, односа аргументације и контрааргументације. Екстракција аргумената је релативно нова област у природном језичком процесирању. Постоје три основна приступа при екстракцији аргумената (Lawrence and Reed 2015). Први се ослања на листу дискурсивних маркера који повезују премисе које су подељене у две групе захваљујући свом ефекту на аргументацију – маркери који је подржавају и маркери који је оспоравају. На пример, везници који указују на супротност као што су али, и иако повезују аргументе који се оспоравају, док су везници као и, или, и онда повезани независним аргументима који имају исти циљ. Други приступ аргументацији користи класификацију

по којој се изрази класификују према томе да ли или нису аргументи. Сваки аргумент може бити класифициран као премиса или закључак, или да одговара предефинисаној аргументативној схеми која преузима облик многобројних премиса које заједно подржавају или оспоравају закључак. Најзад, последњи приступ користи промену топика као индикатора за промену аргументативне линије. На пример, ако је топик неке пропозиције сличан оној која је анализирана у претходним пропозицијама, онда се може закључити да су ове пропозиције повезане и да прате исту линију резоновања.

Други правац у оквиру дискурсивне анализе који се користи у рачунарске сврхе обухвата конструисање комплексних дискурсивних јединица које се образују од елементарних дискурсивних јединица у рекурзивном облику. Овај правац наглашава идентификацију реторичких односа који се успостављају међу елементарним јединицама, клаузама које се углавном не преклапају, али могу бити и веће јединице од клауза. Идентификација реторичких односа представља важан корак у дискурсивној анализи. Неке релације су експлицитно маркиране, док су друге имплицитне за чију идентификацију је потребно детаљно познавање лексике и њеног домена. Напори да се инкорпорира дискурсивна информација у оквиру сентимент анализе могу бити груписани у две категорије: оне које се ослањају на локалне дискурсивне односе на интерсентенцијалном или интрасентенцијалном нивоу и на оне које се ослањају на структуру целог документа. Главна идеја је да су међу различитим релацијама само неке релевантне у циљу одређивања сентимента. Најједноставнији начин за њихову идентификацију је да се узму у обзир дискурсивни везници. Полањи и Заенен (2006) су били први који су приметили да дискурсивни маркери могу да обрну поларитет и да могу да измене тип субјективности. На пример, у оквиру односа КОНЦЕСИЈЕ у реченици Иако је ова књига добра за широку публику, слабије се продаје у стручним круговима. позитивни поларитет придева добар се неутрализује прилогом слабије у главној клаузи. Однос УСЛОВА такође лимитира степен позитивне евалуације, као што је то истакнуто у раду Трнавац и Табоаде (2012) у корпусној студији која користи модел оцене у рецензијама књига и филмова. На пример, у реченици То је сјајна књига ако волиш ратну тематику. позитивна конотација придева сјајна је ублажена условом на основу кога читаоци морају да формирају и прилагоде своје мишљење.

Табоада и др. (2008) прво ексципирају реторичку структуру из текстова, класификују делове текста који су у статусу нуклеуса (важнијих делова текста) или сателита (мање важних делова текста). Аутори оцењују да је учинак овог приступа у поређењу са класификатором топика који издваја реченице са топиком већи. Резултати показују да

усмеравање фокуса на релевантне реченице доводи до побољшавања калкулације везане за семантичку оријентацију у односу на метод који је базиран само на сентимент речима и који разматра целокупни текст подједнако. Па ипак, овај метод користи нуклеусе без обзира на тип релације између нуклеуса и сателита. Познато је да контрастрирајући сегмент може да игра другачију улогу у преношењу целокупног сентимента него, на пример, елобарација информације у нуклеусу. Хеершоп и др. (2011) даље развијају овај правац пронашавши да коришћење реторичких релација на нивоу реченица побољшава комплетну класификацију сентимента у тексту за 4.5%.

#### б. Имплицитна евалуација

Поредећи са експлицитном евалуацијом, имплицитна евалуација захтева да читаоци креирају прагматичку инференцу која излази ван домета буквалног схватања текста. Иако су људска бића много боља од аутоматских система у разумевању имплицитне евалуације, њима је доста често тешко да њу прозру. Бенамара и др. (2016) илуструју да имплицитна мишљења боље корелирају са глобалним сентиментом текста онда када се негативна мишљења изражавају. Овакав закључак изражава тенденцију прикривања негативног сентимента као релативно објективне изјаве, која може бити повезана са социјалним конвенцијама (љубазношћу, на пример). Грајс (1975) је направио јасну разлику између онога што је изречено као исказ (значење ван контекста) и онога шта је имплицирано или се подразумева под исказом (значење у контексту). У својој теорији конверзацијске импликатуре, Грајс (1975) тврди да услов да се разуме значење онога што говорник каже јесте то да слушалац мора да се ослони на значење реченице која је употребљена, контекстуалне претпоставке и принцип кооперативности за који се очекује да га говорник прати. Принцип кооперативности је изражен кроз четири максиме које би учесници требало да поштују. Максиме требају од говорника да каже оно шта он мисле да је истина (квалитет), да буде информативан онолико колико је то могуће (квантитет), те да говорник изражава исказ у правом моменту интеракције (релевантност) и на одговорајући начин (начин).

Пример (2) представља типичну илустрацију нарушавања принципа релевантности:

- (2) а. Хоћеш ли на пиће вечерас?
- б. Морам да учим за испит.

Наиме, особа Б у дијалогу каже да не прихвата позив на пиће од особе А иако то није изречено у буквалном смислу.



Генерално, постоје три начина да евалуација буде имплицитна. Први начин је да се опишу жељене или нежељене ситуације (стања или догађаји). Вилсон (2008) указује да су ово поларизоване чињенице, односно да су ово чињенице (које су супротстављене мишљењима), али оне преносе поларитет јер се таква стања или догађаји конвенционално асоцирају са позитивном или негативном евалуацијом. Поларитет таквих чињеница се може одредити користећи информацију из контекста. Ситуације могу бити приказане као глаголске или именичке фразе, као у следећим примерима:

- (3) Филм није лош иако су неке особе напустиле салу.
- (4) Пукотина се појавила у њиховом односу.

Други тип имплицитне евалуације се тиче речи које указују на објекте или ситуације који имају позитивну или негативну конотацију. На пример, обратимо пажњу на подвучене речи у следећем примеру:

- (5) Никола нема сталну адресу.
- (6) Никола је бескућник.

Први пример указује на релативну неутралну констатацију док друга реченица има негативну конотацију. Поједине именице и глаголи имају и неутралну и субјективну конотацију, као у следећим двома реченицама:

- (7) Наставник је инспирисао децу да развијају свој таленат. (позитивна конотација)
- (8) Филм је инспирисан истинитим догађајима. (неутрална конотација)

Иста реч може мењати своју конотацију у зависности од промене тзв. домена или контекста у коме се налази. Иако конотација има снажан утицај на сентимент анализу, највећи број савремених субјективних лексикона садржи речи које су у својој суштини позитивне или негативне. Трећи начин по коме долази до имплицитне евалуације јесте када неко изражава евалуацију према имплицитном аспекту ентитета. Таква ситуација често долази до изражаја у сентимент анализи која је аспектуално базирана. На пример, реченица *Мобилни је шежак*. имплицитно исказује негативно мишљење о аспекту тежине. Неке од ових имплицитних евалуација настају због различитих конотација, неке због постојања полисемије, а неке од њих због везе са одређеним доменом. Иако недавне студије ословљавају глаголске изразе који имплицирају негативна мишљења (Li et al. 2015), идентификација имплицитних аспеката који нису сигнализирани сентимент речима представља још увек проблем који треба решити.

## в. Фигуративни језик

Фигуративни језик дочарава преносна и неконвенционална значења. Он укључује широки спектар феномена, као што су то метафора, оксиморони, идиоматски изрази, иронија, сарказам, итд. (Taboada, Trnavac & Goddard 2017). Метафоре поистовећују два различита ентитета, концепта или идеје, које се називају извором и метом. Традиционално се метафора посматра као домен изражавања експресивног и поетичног језика, али су студије из области когнитивне лингвистике показале да су оне веома честе и у разговорном језику. Когнитивна лингвистика има свеобухватни поглед на метафору и посматра је као фундаментални феномен за наш концептуални систем. Метафоре у језику представљају просто рефлексију нашег концептуалног модела, по коме концептуализујемо један домен користећи језик који је базичнији. На пример, политичке или друге врсте дебата често позајмљују језик рата и конфликта да окарактеришу своју антагонистичку природу (одбранити своју тачку гледишта). Лејкоф и Џонсон (1980) представљају фундаментални рад у овој области. Сем што је фигуративни језик присутан у књижевном тексту, метафора и фигуративни језик су присутни у свим осталим жанровима и на свим нивоима формалности језика. По тврдњама Шутове и Теуфела (2010) отприлике једна у три реченице у регуларном тексту садржи метафорички израз. Иронија и сарказам могу да се посматрају као облици метафоричког и фигуративног језика јер преносе значења која представљају нешто више од онога што је буквално исказано. Иронија представља сложени лингвистички феномен који је широко проучаван у филозофији и лингвистици (Sperber and Wilson 1981; Utsumi 1996; Attardo 2000). Занемарујући разлике између различитих праваца, иронија може бити представљена као неподударност између буквалног значења свих изрази и њихових намереваних значења. На пример, да би се изразило негативно мишљење о филму, неко може да употреби буквално значење користећи реч са негативном конотацијом као у реченици Овај филм је права катастрофа. или да употреби преносни облик значења користећи речи са позитивном конотацијом, као на пример *Какав диван филм!* За многе истраживаче, иронија се преклапа са многим другим средствима фигуративног језика, као што су сатира, пародија и сарказам (Gibbs 2000). У рачунарској лингвистици, иронија се често користи као термин који укључује и сарказам, иако неки истраживачи праве разлику између ироније и сарказма, сматрајући да сарказам има тенденцију да буде грубљи и агресивнији (Clift 1999). Постоје два начина да се одреди да ли је у питању иронија или сарказам тексту. Први начин је да говорник намерно ствара експлицитну јукстапозицију инкомпатибилних радњи или речи које имају супротне поларитете или

могу бити семантички невезане. Експлицитна опозиција може такође да настане од експлицитног позитивног или негативног контраста између субјективних пропозиција и ситуације која описује нежељену активност или стање. Иронија настаје на основу претпоставке да писац и читалац деле заједничку информацију о датој ситуацији која се дефинише као негативна кроз културне и социјалне норме. Ради детекције ироније у експлицитним и имплицитним опозицијама савремени приступи се ослањају на велики низ стилистичких црта, као и карактеристика које су засноване на особинама речника базираних на сентименту и афекту (González-Ibáñez et al. 2011).

За сентимент анализу може бити значајна екстралингвистичка информација. То може бити, пре свега, демографска информација и структура социјалне мреже. Демографска информација се односи на статистичке податке који се користе у маркетингу и бизнису да би се публика класификовала према узрасту, полу, раси, приходима, локацији, политичкој оријентацији и другим категоријама. Неколико студија је пронашло снажну корелацију између изражавања субјективности и пола и искористила ове корелације за идентификацију пола (Volkova et al. 2015). На пример, жене чешће користе више емотивних сигнала (емотикона) него мушкарци. Недавно су Волкова и др. (2014) предложили приступ који би исказао разлике у половима да би се побољшала мултилингвална сентимент класификација у социјалним медијима. Тај метод се ослања на претпоставку да ће неколико субјективних речи бити коришћено од стране мушкарца, али не и од жена, или обрнуто. На поларитет се може утицати и на основу присуства пола. Поред пола, Persing и Ng (2014) истражују 15 других демографских информација да би предвидели гласове из коментара који су представљени на вебсајту за популарно социјално гласање. Ова информације се може наћи на профилу корисника и укључује, на пример, политичке погледе, брачни статус, и да ли корисник пије или пуши.

#### г. Интенција

Дискурсивни и различити прагматички контексти могу да побољшају системе за сентимент анализу. Међутим, познајући шта агент воли или не воли представља само први корак у процесу одлучивања. Погледајмо следеће реченице:

- (9) Не волим јефтине пакет аранжмане и никада нећу купити ниједан.
- (10) Волео бих да купим лепу кућу са базеном.
- (11) Долазим следећег месеца.

Ако анализирамо наведене примере с тачке гледишта сентимент анализе, само први пример би био класификован као негативан, док би се други пример сматрао објективним. Ипак, као додатак негативном мишљењу, писац у примеру (9) експлицитно изражава своју интенцију да не купује јефтине аранжмане. У примеру (10) писац жели промену своје постојеће ситуације, али не постоји гаранција да ће ова жеља водити формирању намере да се купи нова кућа у будућности. У примеру (11) је исказана имплицитна интенција. Анализа такве интенције покушава да се фокусира на детекцију будућних стања које би агент желео да постигне. Термин намера се користи као шири термин који покрива жеље, преференце и намере, представљајући ментални став који доприноси рационалном понашању агента. Дати ставови играју мотивацијску улогу. Пре него што се одлучимо на неку акцију, агент разматра различите жеље, које представљају стање ствари које би код агента у идеалном свету довело до њене реализације. Жеље могу бити у конфликту и као такве су предмет неконзистентности. Међу тим жељама само неке могу бити потенцијално задовољене. Изабране жеље на које се агент обавезао се зову намере (Perugini and Vagozzi 2004). Намере не могу да буду у конфликту једна са другом и морају бити конзистентне. Ово се сматра важном разликом између жеља и интенција. Жеље могу бити засноване на преференцама. На пример, преференца агента може бити дефинисана у односу на активности као што су куповина новог аутомобила. Међу тим резултатима неки су прихваљиви за агента – наиме, агент је спреман да дејствује да би их испунио, док други резултати нису прихватљиви за њега. Преференце не представљају мишљења. Док се мишљења дефинишу као перспектива, уверења, сентимент или суд који агент може имати о објекту или о личности, преференце су релацијске и упоредиве. Мишљења укључују апсолутне судове према објектима или личностима (позитивна, негативна или неутрална), док се преференце тичу релативних судова према активностима. Следећи примери илуструју изречено.

(12) Књига није лоша.

(13) Први део филма је бољи него други.

(14) Волела бих да идем у позориште. Хајде да погледамо „Злочин и казну”.

Пример (12) изражава директно позитивно мишљење према књизи, али не знамо да ли је та књига најомиљенија. Пример (13) изражава компаративно мишљење о два дела филма. Ако је гледање ова два дела филма јасно из контекста, овакво компаративно мишљење би имплицирало преференцу, односно давање првенства првом делу у односу на други. На крају, пример (14) изражава две преференце од којих једна зависи

од друге. Резоновање о преференцама се такође разликује од резоновања о мишљењима. Преференце агента одређују редослед резултата који предвиђа како ће агент у будућности поступати. Ово не важи за мишљења. Мишљења највише имају индиректну везу са активношћу, што значи следеће – можда апсолутно не волим шта тренутно радим, али то ипак радим јер више волим такав исход у односу на било коју другу алтернативу.

Верујемо да дискурсивни и прагматични феномени играју веома важну улогу у интерпретацији евалуативног језика и да треба да буду узети у обзир у циљу тачног описа сентимента, као и да треба да буду обухваћена аотацијским процесом приликом креирања корпуса са субјективним изразима. Динамичка дефиниција сентимента која је била презентована укључује различите контекстуалне аспекте који би били обухваћени калкулацијом сентимента евалуативних речи и израза, који могу бити примењени на све нивое језика. На пример, контекстуалне карактеристике могу бити од значаја да би се детектовала имплицитна евалуација, и да би се прецизно описало значење фигуративних израза. Укључивање дискурсивне информације у анализу размишљања је од велике користи. Дискурс је такође успешно ангажован у машинском превођењу (Hardmeier 2013), генерисању природног језика (Ashar and Indukhya 2010) и језичкој технологији уопште (Taboada and Mann 2006).

## ЛИТЕРАТУРА

- Ashar, J., & N. Indukhya (2010) A unifying view of computational discourse and natural language generation. *ACM Computing Surveys*, pages 1–30.
- Asher, N., F. Benamara, & Y. Mathieu (2009) Appraisal of opinion expressions in discourse. *Linguisticae Investigationes*, 32(2): 279–292.
- Attardo, S. (2000) Irony as relevant inappropriateness. *Journal of Pragmatics*, 32(6): 793–826.
- Baker, C. F., C. J. Fillmore, & J. B. Lowe. (1998) The Berkeley FrameNet Project. In *Proceedings of the International Conference on Computational Linguistics* – Volume 1, pages 86–90, Montréal.
- Bakhtin, M. (1981) Discourse in the novel. In Michael Holquist, editor, *The Dialogic Imagination: Four Essays by M. M. Bakhtin*. University of Texas Press, Austin, pages 259–422.
- Balahur, A., J. M. Hermida, A. Montoyo, & R. Muñoz (2011) EmotiNet: A knowledge base for emotion detection in text built on the Appraisal theories. In *Natural Language Processing and Information Systems*, volume 6716 of Lecture Notes in Computer Science. Springer, pages 27–39.

- Balahur, A., J. M. Hermida, & A. Montoyo (2012) Detecting implicit expressions of emotion in text: A comparative analysis. *Decision Support Systems*, 53(4): 742–753.
- Baccianella, S., A. Esuli, & F. Sebastiani (2010) SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Conference on International Language Resources and Evaluation*, pages 2200–2204, Malta.
- Benamara, F., C. Cesarano, A. Picariello, D. Reforgiato, & V. S. Subrahmanian (2007) Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In *Proceedings of the International Conference on Weblogs and Social Media*, Boulder, CO.
- Benamara, F., B. Chardon, Y. Mathieu, & V. Popescu (2011) Towards context-based subjectivity analysis. In *Proceedings of the International Joint Conference on Natural Language Processing*, pages 1180–1188, Chian Mai.
- Benamara, F., B. Chardon, Y. Y. Mathieu, V. Popescu, & N. Asher (2012) How do negation and modality impact opinions? In *Proceedings of the ACL-2012 Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics*, pages 10–18, Jeju Island.
- Benamara, F., N. Asher, Y. Mathieu, V. Popescu, & B. Chardon (2016) Evaluation in discourse: A corpus-based study. *Dialogue and Discourse*, 7(1): 1–49.
- Biber, D., & Finegan, E. (1989) Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text* 9 (1), pp. 93–124.
- Boye, K., & P. Harder (2009) Linguistic categories and grammaticalization. *Functions of Language*, 16(1):9–43.
- Chafe, W. (1986) Evidentiality in English conversation and academic writing. In W. Chafe and J. Nichols, editors, *Evidentiality: The Linguistic Coding of Epistemology*. Ablex, Norwood, NJ, pages 261–272.
- Chafe, W., & J. Nichols (1986) *Evidentiality: The Linguistic Coding of Epistemology*. Ablex, Norwood, NJ.
- Chardon, B., F. Benamara, Y. Mathieu, V. Popescu, & N. Asher. (2013) Sentiment composition using a parabolic model. In *Proceedings of the International Conference on Computational Semantics*, pages 47–58, Potsdam.
- Chetviorkin I., & N. Loukachevitch Extraction of Russian Sentiment Lexicon for Product Meta-Domain. In *Proceedings of COLING 2012*, Mumbai, India, 2012, pp. 593–610.
- Conrad, S., & D. Biber (2000) Adverbial marking of stance in speech and writing. In S. Hunston and G. Thompson, editors, *Evaluation in Text: Authorial Distance and the Construction of Discourse*. Oxford University Press, Oxford, pages 56–73.
- Ding, X., T. Liu, J. Duan, and J. Y. Nie (2015) Mining user consumption intention from social media using domain adaptive convolutional neural network. In *Proceedings of the Conference on the American Association of Artificial Intelligence*, pages 2389–2395.
- Ekman, P. (1984) Expression and the nature of emotion. In K. Scherer and P. Ekman, editors, *Approaches to Emotion*. Erlbaum, Hillsdale, NJ, pages 319–344.

- Farzindar, A. and Diana I. (2015) *Natural Language Processing for Social Media. Synthesis Lectures on Human Language Technologies*. Morgan & Claypool publishers.
- Giannakidou, A. (1995) On the semantic licensing of polarity items. In Anastasios-Phoevos Christidis, M. M. R., & A. Argyris (Eds.), *Studies in Greek Linguistics 15: Proceedings of the Annual Meeting of the Department of Linguistics*, pages 406–418, University of Thessaloniki.
- Gibbs, R. W. (2000) Irony in talk among friends. *Metaphor and Symbol*, 15(1–2): 5–27.
- González-Ibáñez, R., S. Muresan, & N. Wacholde (2011) Identifying sarcasm in Twitter: A closer look. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 581–586, Portland, OR.
- Grice, P. (1975) Logic and conversation. In Cole, P.; Morgan, J. *Syntax and semantics*. 3: Speech acts. New York: Academic Press. pp. 41–58.
- Grosz, B. J., A. K. Joshi, & S. Weinstein (1995) Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 21(2): 203–225.
- Halliday, M.A.K., & R. Hasan (1976) *Cohesion in English*. London: Longman.
- Halliday, M. A. K., & C. M. I. M. Matthiessen (2014) *An Introduction to Functional Grammar*. Arnold, London, 4th edition.
- Hardmeier, C. (2013) Discourse in statistical machine translation: A survey and a case study. *Discours*, 11.
- Heerschop, B., F. Goossen, A. Hogenboom, F. Frasinca, U. Kaymak, & F. de Jong (2011) Polarity analysis of texts using discourse structure. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, pages 1061–1070, Glasgow.
- Hobbs, J. (1979) Coherence and coreference. *Cognitive Science* (3), 8: 67–90.
- Horn, L. (1989) *A Natural History of Negation*. University of Chicago Press.
- Hunston, S., & G. Thompson (2000) *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press.
- Izard, C. E. (1971) *The Face of Emotion*. Appleton Century Crofts, New York.
- Lakoff, G. and M. Johnson (1980) *Metaphors We Live By*. University of Chicago Press, Chicago.
- Langacker, R. W. (1990) Subjectification. *Cognitive Linguistics*, 1(1): 5–38.
- Lawrence, J. and C. Reed (2015) Combining argument mining techniques. In *Working Notes of the 2nd Argumentation Mining Workshop*, pages 127–136, Denver, CO.
- Levin, B. (1993) *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press, Chicago.
- Liu, Bing 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool.
- Liu, Bing 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press, Cambridge.
- Liu, F., W. Dong, L. Bin, & L. Yang (2010) Improving blog polarity classification via topic analysis and adaptive methods. In *Proceedings of Human Language*

- Technologies: The Annual Conference of the North American Chapter of the ACL*, pages 309–312, Los Angeles, CA.
- Liu, Y., X. Yu, B. Liu, & Z. Chen. (2014) Sentence-level sentiment analysis in the presence of modalities. In *Computational Linguistics and Intelligent Text Processing*, volume 8404 of CICLing 2014, Kathmandu, pages 1–16.
- Manoussos, I. Katakis, I. Varlamis, & G. Tsatsaronis (2014) Pythia: Employing lexical and semantic features for sentiment analysis. In *Machine Learning and Knowledge Discovery in Databases*, volume 8726 of Lecture Notes in Computer Science, pages 448–451. Springer, Berlin.
- Martin, J. R. (2000) Beyond exchange: Appraisal systems in English. In S. Hunston and G. Thompson, editors, *Evaluation in Text: Authorial Distance and the Construction of Discourse*. Oxford University Press, Oxford, pages 142–175.
- Martin, J. R. (2014) Evolving Systemic Functional Linguistics: Beyond the clause. *Functional Linguistics*, 1(3): 1–24.
- Martin, J. R. & P. R. R. White (2005) *The Language of Evaluation*. New York: Palgrave.
- Mathieu, Y. Y., & C. Fellbaum (2010) Verbs of emotion in French and English. In *The 5th International Conference of the Global WordNet Association*, Mumbai.
- Ortony, A., G. L. Clore & A. Collins (1988) *The cognitive structure of emotions*. New York: Cambridge University Press.
- Osgood, C. E., G. J. Suci, & P. H. Tannenbaum (1957) *The Measurement of Meaning*. University of Illinois Press, Urbana.
- Pang, B., & L. Lee, (2008) Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2): 1–135
- Persing, I., & V. Ng (2014) Vote prediction on comments in social polls. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1127–1138, Doha.
- Perugini, M., & R. P. Bagozzi (2004) The distinction between desires and intentions. *European Journal of Social Psychology*, 34(1): 69–84.
- Polanyi, L. & A. Zaenen (2006) Contextual valence shifters. In James G. Shanahan, Yan Qu, and Janyce Wiebe, editors, *Computing Attitude and Affect in Text: Theory and Applications*. Springer, Berlin, pages 1–10.
- Poria, S., A. Gelbukh, A. Hussain, D. Das, & S. Bandyopadhyay (2013) *Enhanced SenticNet with affective labels for concept-based opinion mining*. *IEEE Intelligent Systems*, 28(2): 31–38.
- Qiu, M., Y. Sim, N. A. Smith, & J. Jiang (2015) Modeling user arguments, interactions and attributes for stance prediction in online debate forums. In *SIAM International Conference on Data Mining*, pages 855–863, Vancouver.
- Reyes, A. and P. Rosso (2012) Making objective decisions from subjective data: Detecting irony in customer reviews. *Decision Support Systems*, 53(4): 754–760.
- Ruppenhofer, J. and I. Rehbein (2012) Semantic frames as an anchor representation for sentiment analysis. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, pages 104–109, Jeju Island.
- Russell, J. A. (1983) Pancultural aspects of the human conceptual organization of emotions. *Journal of Personality and Social Psychology*, 45: 1281–1288.



- Shutova, E., & S. Teufel. (2010) Metaphor corpus annotated for source-target domain mappings. In *Proceedings of the International Language Resources and Evaluation*, LREC 2010, pages 3255–3261, Malta.
- Sperber, D., & D. Wilson (1981) Irony and the use-mention distinction. *Radical Pragmatics*, 49: 295–318.
- Stede, M. (2011) *Discourse Processing*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.
- Strapparava, C., & A. Valitutti (2004) WordNet-Affect: An affective extension of WordNet. In *Proceedings of the International Conference on Language Resources and Evaluation*, pages 1083–1086, Lisbon.
- Trnavac, R. (2006) *Aspect and subjectivity in modal constructions*. PhD thesis. Leiden: LOT.
- Taboada, M., & W. C. Mann (2006) Rhetorical Structure Theory: looking back and moving ahead. *Discourse Studies*, 8(3): 423–459.
- Taboada, M., K. Voll, & J. Brooke (2008) Extracting sentiment as a function of discourse structure and topicality. *Technical report 2008-20*, School of Computing Science.
- Taboada, M., J. Brooke, M. Tofiloski, K. Voll, & M. Stede (2011) Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics* 37(2), 267–307.
- Taboada, M., & R. Trnavac (eds.) (2013) *Nonveridicality and evaluation: Theoretical, computational and corpus approaches*. Leiden: Brill.
- Trnavac, R., & M. Taboada (2012) The contribution of nonveridical rhetorical relations to evaluation in discourse. *Language Sciences*, 34 (3): 301–318.
- Taboada, M., R. Trnavac, & C. Goddard (2017) On being negative. *Corpus Pragmatics* 1(1): 57–76.
- Thompson, G., & L. Alba-Juez, editors (2014) *Evaluation in Context*. John Benjamins.
- Utsumi, A. (1996) A unified theory of irony and its computational formalization. In *Proceedings of the 16th Conference on Computational Linguistics*, pages 962–967, Sydney.
- White, P. R. R. (2003) Beyond modality and hedging: A dialogic view of the language of intersubjective stance. *Text*, 23(2):259–284.
- White, P. R. R. (2004) Subjectivity, evaluation and point of view in media discourse. In C. Coffin, A. Hewings, & K. O'Halloran (eds) *Applying English Grammar: Corpus and Functional Approaches*. Arnold, London, pages 229–246.
- White, P. R. R. (2012). An introductory course in Appraisal analysis. <http://languageevaluation.info/appraisal/>.
- Wiebe, J. M. (1994) Tracking point of view in narrative. *Computational Linguistics*, 20(2): 233–287.
- Wiebe, J., T. Wilson, R. Bruce, M. Bell, & M. Martin (2004) Learning subjective language. *Computational Linguistics*, 30(3):277–308.
- Wiebe, J., & E. Riloff (2005) Creating subjective and objective sentence classifiers from unannotated texts. In *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*, volume 3406 of Lecture Notes in Computer Science, pages 486–497.

- Wilson, T. (2008) Annotating subjective content in meetings. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, pages 2738–2745, Marrakech.
- Volkova, S., G. Coppersmith, & B. Van Durme (2014) Inferring user political preferences from streaming communications. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 186–196, Baltimore, MD.
- Volkova, S., Y. Bachrach, M. Armstrong, & V. Sharma (2015) Inferring latent user properties from texts published in social media. In *Proceedings of the Twenty-Ninth Conference on Artificial Intelligence*, pages 4296–4297, Austin, TX.
- Yano, T., D. Yogatama, & N. A. Smith (2013) A penny for your tweets: Campaign contributions and Capitol Hill microblogging. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*, pages 737–740, Boston, MA.
- Zwarts, F. (1995) Nonveridical contexts. *Linguistic Analysis*, 25(3/4): 286–312.

Radoslava Trnavac

#### THE SYSTEM OF EVALUATION IN RUSSIAN – RESEARCH PROGRAM

##### Summary

The objective of this paper is to describe a project for creating a digital corpus of Russian with annotations of evaluative expressions. The suggested annotation of corpus includes the following information: the expressions of attitude, eventuality, nonveridicality, grammatical patterns of subjectivity, the evaluative lexicon, discourse phenomena, implicit evaluation, figurative language and intention of the speaker.

*Key words:* corpus annotation, evaluation, sentiment analysis, discourse and pragmatic information