
СентИроСа: корпус за детекцију ироније и сарказма у кратким текстовима на српском језику

Данка Јокић¹[0000-0002-7621-5847] , Милица Андрић²[0009-0008-0319-4548]

¹ Универзитет у Београду, Србија, danka.jokic@rcub.bg.ac.rs

² Универзитет у Београду, Филолошки факултет, Србија, milica@mail.ru

Научни рад

DOI: 10.18485/judig.2026.2.ch14

Апстракт

Препознавање ироније и сарказма у тексту може побољшати перформансе многих система за обраду природног језика (ОПЈ), укључујући сумирање текста, анализу сентимента, откривање увредљивог језика итд. Развој система за аутоматску детекцију ироније и сарказма захтева ручно обележене скупове података за обучавање и евалуацију модела. Међутим, за српски језик до сада није постојао јавно доступан ручно обележен корпус потребан за аутоматску детекцију ироније и сарказма. У овом раду представљамо СентИроСа, први јавно доступан ручно анотирани корпус за детекцију ироније и сарказма на српском језику. Корпус је формиран проширивањем постојеће колекције кратких текстова са платформе Твитер/Х, претходно анотираних ознакама сентимента, новим ручним анотацијама за иронију и сарказам. Састоји се од 1117 твитова која су обележавала три независна анотатора. У раду су описани поступак анотације и резултати квалитативне и квантитативне анализе корпуса, укључујући анализу слагања анотатора, лингвистичких маркера ироније и сарказма и повезаности ових феномена са сентиментом. СентИроСа представља нови језички ресурс за истраживања ироније, сарказма и других повезаних феномена у обради природног језика на српском језику.

Кључне речи: иронија, сарказам, корпуси

1. Увод

Аутоматско препознавање ироније и сарказма у текстовима је значајно за више области у обради природног језика (ОПЈ), јер се препознавањем ових реторичких фигура могу унапредити

ти перформансе ОПЈ система (Kühn et al., 2026). У ширем смислу, Wallace (2015) чак тврди, да сваки модел који се нада да ће разумети људску комуникацију или изражавање мора бити у стању да разликује иронично од искреног. Најпре, само боље разумевање текста, чему доприноси препознавање ироније и сарказма, помаже у његовој квалитетнијој сумаризацији. Поред тога, иронија и сарказам могу да промене сентимент текста – није свеједно да ли текст има дословно значење, или се у њему крију иронија и сарказам (Carvalho et al., 2009). Даље, овакви текстови могу имати и имплицитно увредљиво значење, због чега се иронија и сарказам доводе у везу са системима за детекцију имплицитног говора мржње и других облика индиректног увредљивог језика (Frenda et al., 2022; Cabrera et al., 2025). И на крају, препознавање ироније и сарказма може бити релевантно и за детекцију лажних вести и дезинформација, будући да су претходна истраживања указала на њихово присуство у оваквим садржајима (Rubin et al., 2016; Buzea et al., 2022). У говорном дискурсу, посебно у контексту разговора, обично можемо уочити низ спољашњих знакова (нпр. израз лица, интонацију, трајање паузе) који омогућавају препознавање ироније. У писаном тексту ови знакови нам нису на располагању, што додатно компликује детекцију ироније и сарказма.

Да би се аутоматски препознали иронија и сарказам, потребни су ручно анотирани корпуси. Прикупљање и анотација оваквих корпуса је обиман посао. Поред прављења или проналажења адекватног извора података, посебан изазов представља анотација таквог скупа. Наиме, још увек не постоји сагласност међу истраживачима из различитих домена, а овом темом се баве и лингвисти и психолози као и информатичари, како се формално могу дефинисати иронија и сарказам (Filatova, 2012; Wallace, 2015). Њихово прецизно дефинисање и разграничење додатно отежавају промене у употреби током времена, као и културне и регионалне варијације, посебно у употреби сарказма (Filatova, 2012). Као што Myers (1977) примећује, дефиниција ироније се мора проширити изван једноставне предиктивне опозиције (говорење супротног од онога што се мисли) која је своди на поље семантике. Поред тога, потребно је укључити прагматичка и евалуативна значења, анализу дискурса, као и активно учешће слушаоца у препознавању често имплицитног значења које стоји

иза дословних изјава. Ситуација са дефиницијом сарказма је још комплекснија. Док га одређени аутори сматрају посебним стилском фигуром, неки други га дефинишу као инструмент посебног облика тзв. вербалне ироније (Carvalho et al., 2009). Према Joshi et al. (2018), сарказам представља облик фигуративног језика код кога дословно значење исказа не одговара намераваном значењу говорника. Аутори истичу да је сарказам блиско повезан са иронијом и да његово разумевање често зависи од контекста и заједничког знања учесника комуникације. Зато се поставља питање како направити упутство за анотаторе са дефиницијама и скупом формалних критеријума за откривање ироније и сарказма у текстовима. Иако не постоје општеприхваћене формалне дефиниције ироније и сарказма, људи углавном имају интуитивно разумевање ових појмова и могу поуздано да препознају текстуалне исказе који садрже иронију или сарказам (Filatova, 2012). Међутим, препознавање присуства ових феномена не подразумева и њихово једнозначно међусобно разграничење. Истраживања показују да се иронија и сарказам у свакодневном разумевању могу доживљавати као веома слични или чак међусобно замениви појмови, док њихова употреба у конкретним ситуацијама указује и на одређене разлике. Карактеристике на основу којих се они могу разграничити нису једнозначне и могу зависити од језичког и културног контекста (Taylor, 2017).

Додатни изазови јављају се и приликом изградње корпуса. Наиме, претходна истраживања указују на проблеме недостатка података и неуравнотежене расподеле класа у корпусима, у којима су примери ироније и сарказма често знатно мање заступљени, као и на могућу пристрасност при избору примера услед фаворизовања јасних и лако препознатљивих примера у односу на граничне случајеве (Kühn et al., 2026). Поред избора примера, посебан изазов представља и субјективност самог процеса анотације. Casola et al. (2024) су уочили високу варијабилност у перцепцији ироније међу анотаторима различитих социодемографских група, што указује на важност узимања у обзир карактеристика и порекла анотатора за анализу ироније. Поред тога, аутори указују да препознавање ироније и даље остаје отворен истраживачки проблем, посебно код мање заступљених језика и у сценаријима без претходно означених података за обучавање модела (енг. *zero-shot*).

Истраживања ироније и сарказма на српском језику у контексту аутоматског препознавања ових феномена готово да нема. У радовима који се баве анализом сентимента, аутори препознају да ови феномени постоје и да утичу на промену поларитета текста и томе прилазе на два начина:

- Укључују препознавање ироније и сарказма у процес аотације, као критеријум или правило на које треба обратити пажњу приликом означавања текста (Batanović, 2020; Slijepčević Bjelivuk, 2023);
- Сматрају да су у домену њиховог истраживања ови феномени ретки (Grljević, 2016) или имају други фокус истраживања у том тренутку (Ljajić and Marovac, 2019)

Једино истраживање на ову тему урадили су Mladenović et al. (2017), који су представили два одвојена скупа твитова, један за иронију и други за сарказам, уз хибридни модел за детекцију ироније заснован на комбинацији правила и модела машинског учења. Ресурси из овог истраживања нису јавно доступни.

Важност теме за обраду природног језика и недостатак јавно доступних ресурса, мотивисали су нас да аотирамо скуп твитова на српском језику и учинимо га доступним другим истраживачима. Поред тога, као што су истакли Casola et al. (2024), препознавање ироније остаје изазов, нарочито за мање заступљене језике. У овом раду представимо процес аотације скупа твитова са ознакама за иронију и сарказам, те квалитативне и квантитативне резултате аотације и везу ироније и сарказма са сентиментом.

Доприноси овог рада су:

1. Аотација скупа од 1117 твитова на присуство ироније и сарказма у тексту,
2. Квалитативна анализа и идентификација лингвистичких маркера за препознавање ироније и сарказма у кратким текстовима на Твитеру/Х-у,
3. Анализа повезаности сентимента и ироније и сарказма.

Остатак рада је организован на следећи начин. У поглављу 2. представљамо претходна истраживања, који се баве детекцијом сарказма и ироније, и дајемо неколико примера корпуса

који су направљени за ове потребе, са посебним освртом на приступе анотацији и лингвистичке маркере коришћене за њихово препознавање. У поглављу 3. описујемо скуп података и поступак његове анотације. У поглављу 4. представљамо резултате квалитативне и квантитативне анализе корпуса, укључујући анализу унутараннотаторског и међуанотаторског слагања, повезаности сентимента са иронијом и сарказмом и уочених лингвистичких образаца. Коначно, у поглављу 5. сумирамо главне налазе истраживања, разматрамо ограничења представљеног корпуса и указујемо на правце будућих истраживања.

2. Претходна истраживања

У овом одељку дат је преглед релевантних корпуса за иронију и сарказам, приступа њиховој анотацији, као и лингвистичких маркера који се користе за опис и анализу ових феномена.

2.1. Корпуси и анотација ироније и сарказма

Carvalho et al. (2009) су изучаваајући ставове корисника на мрежама, прикупили коментаре са веб-сајта популарних португалских новина. Приметили су да је код 35% лажно позитивних ставова, узрок грешке била вербална иронија. Стога је циљ експеримента био да се идентификују обрасци који могу бити добри индикатори ироничних реченица које би, у супротном, биле класификоване као позитивне. Према ручној евалуацији реченица из корпуса, најпоузданије карактеристике које сигнализирају присуство ироније су: (i) емотикони и ономатопејски изрази за смех, (ii) наглашена интерпункција, (iii) знаци навода и (iv) узвици којима се изражава позитиван став, одобравање, подршка или захвалност, као што су „браво“, „само напред“или „хвала пуно“. Закључили су да се са релативно високом прецизношћу од 45% до 85% само на основу ових „површинских“ лингвистичких образаца може препознати ироничан садржај. Са друге стране, маркери засновани на дубљим лингвистичким информацијама, као што је присуство деминутива личних имена, били су много мање ефикасни.

Filatova (2012) је прикупила коментаре на производе са сајта Amazon.com у паровима – један коментар без ироније и сарказма (неутралан) и други са иронијом и сарказмом, користећи ангажовање спољних анотатора са платформе Amazon Turk

(енг. *crowdsourcing*). Анотаторима нису дата унапред упутства, рачунало се на то да ће они интуитивно моћи да препознају ове феномене. Стога је слагање између више анотатора коришћено као главни критеријум за прихватање или елиминацију коментара из скупа података. Након отклањања дупликата у подацима, уследила је квалитативна провера из три корака, да би се у скупу задржали само коментари око којих се сложило најмање три од пет анотатора. Ауторка је приметила да не само да постоји неслагање између различитих анотатора, него је иста особа у различитим циклусима исти текст означавала различито. Кроз анализу сентимента коментара са једне стране, и корисничких оцена производа изражених бројем звездица са друге, ово истраживање је емпиријски потврдило да иронија и сарказам међају позитивно дословно значење у негативно – коментари су звучали позитивно, док је оцена производа била ниска.

Значајан подстицај истраживањима ироније дала је SemEval-2018 кампања за детекцију ироније у твитовима (Van Hee et al., 2018). За потребе овог задатка развијен је ручно аотирани корпус твитова на енглеском језику, који је постао један од најчешће коришћених скупова података за евалуацију система за детекцију ироније. Анотација је спроведена у два корака: најпре је утврђивано да ли је твит ироничан или није, а затим су иронични твитови сврставани у једну од три категорије: вербална иронија заснована на сукобу поларности, ситуациона иронија и други облици вербалне ироније. У овом раду сарказам се посматра као подврста ироније. Корпус садржи 3.834 твита у скупу за обучавање и 784 твита у тестном скупу.

Новији приступ представља iSarcasm корпус, у коме су за разлику од већине претходних корпуса ознаке сарказма додељивали сами аутори објава, са циљем да се сарказам препозна у складу са намером аутора, а не на основу субјективне интерпретације спољних анотатора (Oprea and Magdy, 2020). Од аутора саркастичних објава тражено је и да објасне зашто је њихова објава саркастична и да формулишу несаркастичну верзију која преноси исто значење. Саркастични примери су додатно класификовани у категорије фигуративног говора, укључујући сарказам, иронију, сатиру, претеривање, потцењивање и реторичка питања. На овом приступу заснован је и SemEval-2022 задатак iSarcasmEval (Abu Farha et al., 2022).

Недавна истраживања све више наглашавају перспективистички приступ анотацији при изградњи корпуса, који полази од тога да неслагање анотатора не мора нужно представљати грешку у поступку анотације, већ може одражавати различите интерпретације субјективних феномена као што су иронија и говор мржње (Casola et al., 2024). MultiPICo (Casola et al., 2024) је вишејезични корпус за детекцију ироније који садржи 18.778 парова објава и одговора на њих са друштвених мрежа Reddit и Twitter/X на девет језика. Задатак анотације био је да се утврди да ли одговор представља ироничну реакцију на дату објаву. Корпус је аотирало 506 анотатора различитих демографских профила, при чему је сваки пример добио у просеку више од пет независних анотација. Посебна вредност овог корпуса огледа се у укључивању демографских карактеристика анотатора, што омогућава проучавање утицаја различитих перспектива на перцепцију ироније.

Што се тиче аутоматске детекције ироније и сарказма на српском језику, Mladenović et al. (2017) су представили хибридни модел за детекцију ироније заснован на правилима онтолошког закључивања коришћењем српског WordNet-a (SWN) и језичих ресурса, међу којима су лексикон сентимента и лексикон маркера за иронију. Из ових ресурса су дизајниране карактеристике за обучавање модела машинског учења. За потребе овог пројекта аотирана су два скупа података – један за иронију и други за сарказам. Највећа тачност развијеног класификатора за иронију била је 86,1%, док је за сарказам најбољи резултат био 72,8%. Ниска вредност метрике одзива класификатора указује на потребу за већим ресурсима да би се добили бољи резултати класификације, као и да је неопходно прилагодити ресурсе за сарказам, укључујући и лексикон маркера као и додавање речи које изражавају мржњу и увреде. Када је реч о корпусима на сродним јужнословенским језицима, Rašl et al. (2021) су за словеначки језик формирали скуп од 1.031 твита за потребе препознавања сарказма, при чему твитови нису ручно аотирани, већ су ознаке изведене на основу хештегова које су доделили сами корисници.

2.2. Лингвистички маркери ироније и сарказма

Први системи за откривање ироније и сарказма у тексту били су засновани на правилима. Ови системи су идентификова-

ли сарказам и иронију на основу присуства специфичних фраз, усмених или гестикулационих знакова у коментарима корисника, као што су емотикони, ономотопејски изрази за смех, наглашена интерпункција, знаци навода и узвици којима се изражава позитиван став, одобравање, подршка или захвалност (Carvalho et al., 2009). У многим случајевима, иронија и сарказам се могу схватити тек када се размотри ситуација или шири контекст исказа. На пример, реченица “Свака ти част!” може да се схвати и као похвала, али и као иронија у зависности од дискурса.

Reyes et al. (2013) су предложили вишедимензионални приступ детекцији ироније на друштвеним мрежама. Уместо ослањања на појединачне маркере, аутори су анализирали више група карактеристика, укључујући контраст између различитих делова исказа, неочекиване семантичке комбинације, стилске маркере и емоционалне обрасце. Резултати су показали да се иронија најпознатије препознаје комбиновањем различитих лингвистичких и прагматичких сигнала, што указује да она не може бити сведена на један појединачни маркер.

Претходна истраживања указала су на повезаност сарказма са сентиментом. Саркастични искази могу имати позитивно дословно значење, док истовремено изражавају негативан став или оцену (Filatova, 2012). Значајан допринос разумевању овог феномена дали су Riloff et al. (2013), који су приметили да се сарказам често манифестује као контраст између позитивног исказа и негативне ситуације на коју се исказ односи. Полазећи од ове претпоставке, аутори су развили приступ за аутоматско издвајање саркастичних образаца из текстова друштвених мрежа. Њихови резултати подржавају хипотезу да контраст између позитивно интонираног исказа и негативне ситуације представља важну карактеристику саркастичног говора. Овај налаз је посебно значајан јер повезује сарказам са анализом сентимента и указује на то да сентимент може представљати важан извор информација за његово препознавање.

3. Методологија

Процес изградње корпуса СентИроСа обухватао је избор извора података, дефинисање категорија за анотацију, израда упутства и ручно означавање твитова од стране анотатора. Након завршетка анотације извршена је консолидација ознака и про-

цена квалитета анотације израчунавањем унутар-анотаторског и међу-анотаторског слагања. Тако добијен корпус је коришћен за анализу расподеле ироније и сарказма, као и испитивање њихове повезаности са сентиментом.

3.1. Скуп података

У овом истраживању је коришћен један део скупа твитова описаног у раду (Ljajić and Marovac, 2019). Подаци коришћени за анотацију преузети су из јавно доступног GitHub репозиторијума¹, који садржи текстове твитова и припадајуће ознаке сентимента. Оригинални скуп је прављен за потребе анализе сентимента и истраживања утицаја негације на његову аутоматску класификацију. Као извор је коришћен Twitter/X, тако што су твитови прикупљени са временске линије 42 корисника – 15 званичних масовних медија, 10 музичара, 5 јавних личности из света забаве, 5 спортиста и 7 глумаца. Анотацију су радила три анотатора, који су означаваали сваки твит ознакама за сентимент – позитиван (ознака 4), неутралан (ознака 2) и негативан (ознака 0). У коначан скуп од 7663 твита су ушли само твитови око чије анотације су се сложили сви анотатори.

Подскуп који је коришћен у овом раду се састоји од 1150 твитова. Анализом скупа примећено је да је потребно елиминисати 33 твита из даље анализе. Разлози су били: (1) текст је написан на страном језику (8 твитова), (2) твит се састоји само из помињања (енг. *mentions*) и линкова ка сајтовима, нема текста (пронађено 7 оваквих случајева), (3) текст није комплетно скинут са Твитера (18 твитова). Просталих 1117 твитова је коришћено за даљу анализу и анотацију. У погледу дистрибуције поларитета, 374 твита има негативан (33.5%), 382 (34.2%) има позитиван сентимент, док се преосталих 361 (32.3%) сматра неутралним. Скуп садржи и линкове ка екстерним страницама, који нису били коришћени као извор информација при анотацији.

3.2. Анотација скупа

У процесу анотације учествовало је девет анотатора, од којих осам поседује филолошко образовање, а један анотатор се

¹ <https://github.com/Data-Science-kosta/Twitter-Sentiment-Analysis-RoBERTa/tree/main>

бави обрадом природног језика. Анотатори су имали претходно искуство са анотацијом ресурса. Обезбеђена им је обука и урађена је пробна анотација. Анотаторима је достављено упутство за анотацију, које је у себи садржало и примере за иронију и сарказам. При изради упутства за анотацију консултована су и анотациона упутства описана у раду [Mladenović et al. \(2017\)](#), као и релевантна литература из области анотације ироније и сарказма. У овом истраживању иронија и сарказам разматрани су у оквиру појединачног твита, без увида у претходни ток комуникације.

Приликом анотације коришћене су следеће оперативне дефиниције:

- Иронија је књижевно средство у коме се речи користе да би се формирало тврђење супротно од очекиваног или оног што је реално [Mladenović et al. \(2017\)](#).
- Саркастични искази су искази са позитивним дословним значењем, негативним намеравањем и јасним метама [Filatova \(2012\)](#).

При разграничењу сарказма од ироније, присуство јасне мете и заједљивости, поруге или вређања сматрано је показатељем сарказма.

Анотатори су сваки твит означили једном од категорија:

- 0 – ни саркастичан ни ироничан: текст твита не садржи ни иронију ни сарказам;
- 1 – иронија: текст је ироничан или сигнализира присуство ироније у реченици;
- 2 – сарказам: текст је саркастичан или сигнализира присуство сарказма у реченици;
- 3 – нејасно: контекст није довољан за доношење одлуке о томе да ли су иронија или сарказам присутни у твиту или не.

Како би се проценила доследност анотација на нивоу појединачних анотатора, у скуп за анотацију су намерно укључени дуплирати твитова. Исти твитови појављивали су се на различитим позицијама унутар скупа, а да анотатори нису знали да се ради о поновљеним примерима. На овај начин омогућена је процена унутар-анотаторског слагања (енг. *intra-annotator agreement*), односно степена у којем исти анотатор додељује исте ознаке идентичним текстовима током процеса анотације. За процену доследности коришћени су проценат поклапања и Коенова капа (κ).

Табела 1: Број анотација и различитих твитова по анотатору

Анотатор	Број анотација	Број различитих твитова
Анотатор 1	319	249
Анотатор 2	316	272
Анотатор 3	308	280
Анотатор 4	312	289
Анотатор 5	316	282
Анотатор 6	324	162
Анотатор 7	323	162
Анотатор 8	552	539
Анотатор 9	1117	1117

Број различитих твитова које је појединачни анотатор аотирао приказан је у Табели 1. Од укупно 1.117 твитова, 1.116 су независно аотирала три различита анотатора, док су један твит аотирала четири анотатора. Процес креирања финалног скупа података спроведен је кроз двостепени систем агрегације како би се обезбедила конзистентност анотација и формирале коначне ознаке за сваки твит. Комплетан поступак формирања коначног корпуса приказан је на Слици 1.

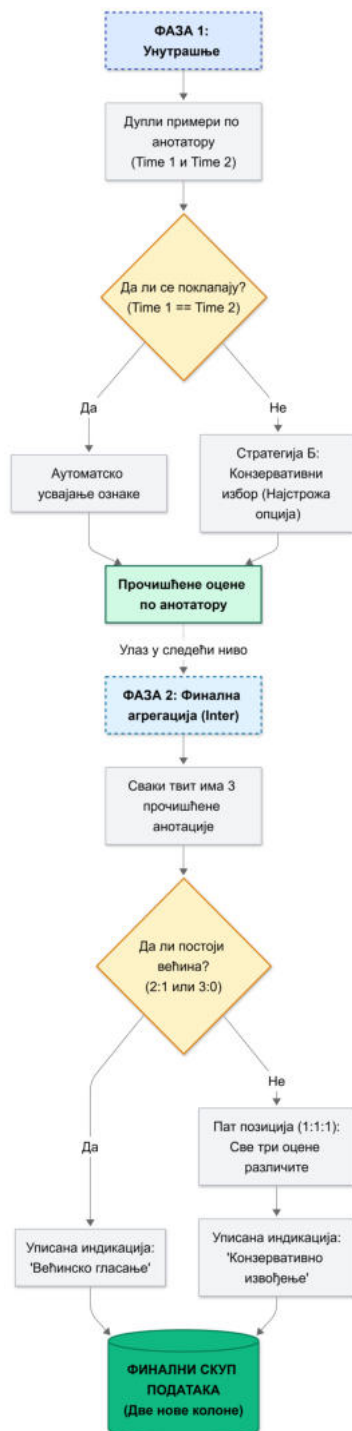
У првој фази вршена је консолидација поновљених анотација истог анотатора (енг. *intra-annotator consolidation*). За твитове које је исти анотатор аотирао више пута и у којима се појавило неслагање приоритет су имале ознаке по следећој шеми:

сарказам > иронија > нејасно > нема ироније и сарказма.

У другој фази вршена је консолидација анотација између три независна анотатора (енг. *inter-annotator consolidation*). Коначна ознака додељивана је већинским гласањем. У случајевима потпуног неслагања анотатора (однос 1:1:1), када није било могуће применити већински принцип, конфликти су разрешавани применом претходно дефинисане шеме приоритета. Оваквим приступом настојало се да се очувају потенцијални примери фигуративног језика и обезбеди додела јединствене коначне ознаке сваком твиту.

4. Резултати

У овом делу рада приказани су резултати анотације корпуса СентИроСа, укључујући анализу унутаранотаторског и међу-



Слика 1: Ток формирања коначног корпуса

анотаторског слагања, расподелу коначних ознака, као и анализу односа између сентимента, ироније и сарказма. Поред квантитативних показатеља, представљени су и квалитативна запажања о маркерима ироније и сарказма и њихова расподела у СентИроСа корпусу.

4.1. Унутаранотаторско слагање

Резултати анализе унутаранотаторског слагања показали су проценат поклапања од 66,36% и вредност Коенове капе од $\kappa = 0.487$, што указује на умерен ниво унутаранотаторског слагања. Ови резултати указују да су анотатори били релативно доследни у примени критеријума приликом означавања ироније и сарказма.

Резултати по појединачним анотаторима приказани су у Табели 2. Анализа резултата показује значајне разлике у степену доследности међу анотаторима. Највиши степен слагања остварили су Анотатор 3 ($\kappa = 0,800$) и Анотатор 5 ($\kappa = 0,727$), што указује на високу интерну конзистентност приликом поновљеног означавања истих примера. С друге стране, код Анотатора 2 забележена је негативна вредност Коенове капе ($\kappa = -0,133$), што указује на веома ниску унутрашњу доследност. Код Анотатора 8 забележена је готово занемарљива вредност капе ($\kappa = 0,015$), али овај резултат треба тумачити са опрезом јер је заснован на малом броју поновљених примера ($N = 13$).

Посебно је занимљив случај Анотатора 4, код којег је забележен висок проценат поклапања (86,96%), али релативно ниска вредност Коенове капе ($\kappa = 0,343$). Овај раскорак представља познати парадокс Коенове капе (енг. *prevalence problem*), који се јавља када је расподела класа неуравнотежена и када висока усаглашеност не мора бити праћена високим вредностима капе.

Додатна анализа образаца неслагања приказана је на Слици 2. Матрица конфузије показује да је највећи број промена забележен код примера који су приликом прве анотације означени као иронија (78): у поновљеној анотацији 34 примера означена су као нејасна, 23 као сарказам, а 21 као пример без ироније и сарказма. Најмање промена забележено код примера означених као нејасни (24). Најчешћа неслагања забележена су између категорија иронија и нејасно (43 случаја) и између ироније и примера

Табела 2: Резултати унутаранотаторског слагања

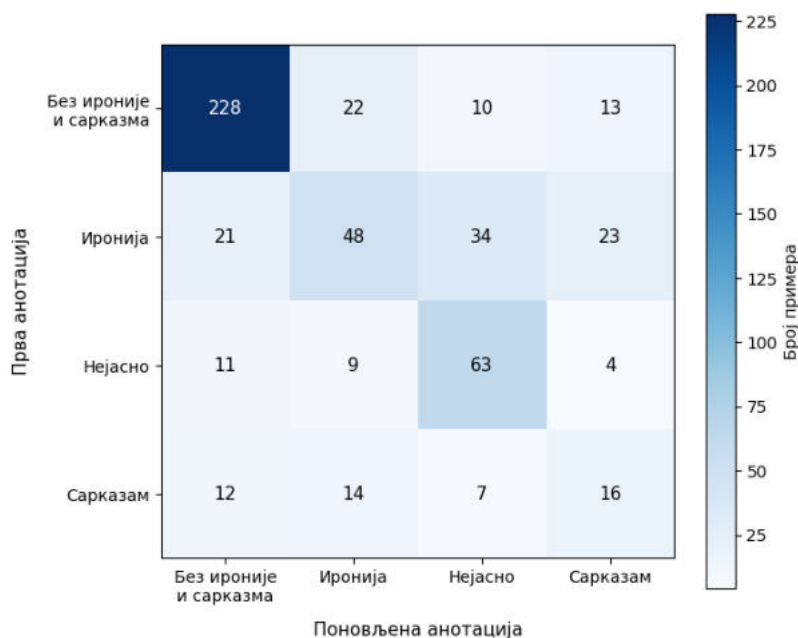
Анотатор	Поновљени твитови (N)	Поклапање (%)	Коенова капа (κ)	Ниво усагла- шености
Анотатор 1	70	67,14	0,400	Умерена
Анотатор 2	44	54,55	-0,133	Испод случајне
Анотатор 3	28	92,86	0,800	Значајна
Анотатор 4	23	86,96	0,343	Слаба (парадокс капе)
Анотатор 5	34	88,24	0,727	Значајна
Анотатор 6	162	67,28	0,311	Слаба
Анотатор 7	161	56,52	0,269	Слаба
Анотатор 8	13	61,54	0,015	Занемарљива

без ироније и сарказма (43 случаја), а затим између ироније и сарказма (37 случајева).

Ови резултати указују да је управо идентификација ироније представљала највећи изазов током анотације. Конфликти су се најчешће јављали у граничним случајевима, када није било јасно да ли твит садржи иронију, да ли припада категорији сарказма или га је потребно означити као нејасан. Истовремено, значајан број неслагања између категорија *иронија* и одсуства фигуративног језика указује на то да су анотатори у одређеним случајевима различито процењивали само присуство фигуративног значења. Сличне изазове у анотацији и интерпретацији ових феномена наводе и претходна истраживања, која истичу њихову контекстуалну зависност и тешкоће у разграничењу (Carvalho et al., 2009; Filatova, 2012). Стога одређени степен варијабилности у поновљеним анотацијама није неочекиван.

4.2. Међуанотаторско слагање

Након консолидације поновљених анотација, применом унапред дефинисане шеме приоритета, израчунато је међуанотаторско слагање три анотатора по једном твиту. Међуанотаторско слагање, мерено Флајсовом капом, износило је $\kappa = 0.161$. Ова вредност указује на низак ниво слагања између различитих анотатора, што није неочекивано имајући у виду природу задатка и сложеност разграничења између категорија ироније, сарказма и нејасних случајева. Низак ниво слагања делимично се може обја-



Слика 2: Матрица конфузије између прве и поновљене анотације

снити чињеницом да иронија није једноставан језички феномен који се препознаје на основу једног маркера, већ захтева тумачење контекста и имплицитног значења (Reyes and Rosso, 2014). Субјективна природа сарказма и ироније представља додатни изазов приликом анотације. Како наводе Joshi et al. (2018), вредности међуанотаторског слагања у истраживањима значајно варирају, што указује на различите интерпретације ових феномена. У поређењу са унутаранотаторским слагањем ($\kappa = 0.487$), овај резултат сугерише да су анотатори били релативно доследни у примени задатих критеријума, али се њихово тумачење критеријума значајно разликовало.

Овакав налаз је у складу са претходним истраживањима која указују да је перцепција ироније изразито субјективна и да може значајно варирати између различитих група анотатора (Casola et al., 2024). Casola et al. (2024) показују да су старосна доб и статус студента међу факторима који највише доприносе поларизацији анотација приликом препознавања ироније. Поред тога, сама граница између категорија ироније и сарказма није једнозначна и зависи и од културно и језичког контекста (Taylor,

2017), док њихово препознавање може зависити и од контекста и информација о говорнику, имплицитног значења и индивидуалне процене (Wallace, 2015). Стога се део ученог неслагања може посматрати не само као последица тешкоћа у анотацији, већ и као одраз различитих интерпретација и перспектива приликом препознавања ових феномена. Иако у овом раду није спроведена посебна анализа утицаја демографских карактеристика анотатора, добијени резултати указују да препознавање ироније и сарказма у великој мери зависи од контекста, интерпретације и индивидуалне процене.

Примена алтернативног приступа консолидацији, заснованог на задржавању прве анотације, довела је до готово идентичних вредности међуанотаторског слагања ($\kappa = 0.170$), што указује да резултати нису осетљиви на изабрани механизам консолидације. Коначни корпус садржи 817 твитова без ироније и сарказма, 95 ироничних твитова, 180 саркастичних твитова и 25 нејасних примера.

Табела 3: Однос између коначних ознака и сентимента

Финална ознака	Негативан	Неутралан	Позитиван	Укупно
Без ирон. и сарк.	262	267	288	817
Иронија	31	31	33	95
Сарказам	80	58	42	180
Нејасно	1	5	19	25
Укупно	374	361	382	1117

4.3. Повезаност сентимента са иронијом и сарказмом

Анализа односа између сентимента и коначних ознака за иронију и сарказам показала је статистички значајну повезаност између ових променљивих ($\chi^2 = 33.93$, $p < 0.001$).

Као што се може видети у Табели 3 категорија сарказма била је најзаступљенија у негативно интонираним твитовима (80 од 180 примера, 44,4%), док је број саркастичних твитова опадао ка неутралним (58 примера) и позитивним садржајима (42 примера). Овај налаз указује на повезаност сарказма са негативним ставовима, критиком и исмевањем. Добијени резултати су у складу са запажањем Riloff et al. (2013) да се сарказам у твитовима често исказује контрастом између позитивних језичких израза и

негативних ситуација или стања, чиме се потврђује веза сарказма и сентимента.

За разлику од сарказма, категорија ироније показала је готово равномерну расподелу између негативних (31), неутралних (31) и позитивних твитова (33), што указује да иронија није изразито повезана са одређеном поларношћу сентимента.

4.4. Општа запажања везана за коришћење ироније и сарказма у твитовима

За препознавање или отклањање сумње у присуство ироније и сарказма у твитовима често је неопходно узети у обзир додатни контекст. У анализи су разматрани следећи извори контекстуалних информација:

1. Структурни делови твита у форми @mention

Пример:

@FilipCukanovich @b92vesti to se zove pesnicka sloboda.

Тумачење: Пошто се у @mention наводе медији, јасно је да је реч о вестима које су медији пренели и односу према њима (слободно тумачење стварности).

Пример:

@b92vesti Bravo Nole

Тумачење: Овде не сумњамо у могућност ироније, јер се наводе медији који преносе вести, те у одсуству другачије интерпункције сматрамо да је похвала на рачун познатог спортисте.

2. Емоји/емотикони који допуњују значење, потврђују или дефинишу супротан тон исказа.

Примери:

@BalkanGoran @Bliconline Ama jel mogu'tje?...vazdan vas ima AV virus ovde :))))))

@IvanGrujic ti su ga odsluzili još 80-ih 70-ih :) Ponovo ista starosna kategorija učestvovala u analizi. cc @Bliconline

@blablaTomiccc @nszorans @Bliconline @stjepa Divni ste. ❤️
Bravo

3. Псовке, пежоративи у тексту које дефинишу значење претходног исказа

Примери:

A nadam sevi zelim od sveg srca da cela redakcija @Blic_online ostane bez zivota. Jebem li vam sve mrtvo i živo.
<https://t.co/falJzDu9nD>

@Pescanik Mozda bi najbolje bilo da radnici rade besplatno, sram vas bilo!

4. Интерпункцијски знаци који могу довести у питање дословно значење исказа

Примери:

@b92vesti Nadam se da Bog gleda kako radite, pa će da Vas "nagradi"

@Blic_online a vi ste list sa "evropskim" standardima? Nepriistrasan? Profesionalan!? A zatooo podrzavate Vucica!! <https://t.co/8VvU6hf96I>

Тумачење: Употреба наводника у тексту.

Пример:

Ako je vucic protiv @N1info onda ja smatram da su oni najbolja TV. Kako vucic a ja kontra i nema pogreske!"

Тумачење: Употреба узвичника у тексту.

Примери:

@b92vesti Opet ste stavili prikladnu fotografiju uz vest?

@Blic_online opet briljira! Ja sam mislila da je ovo koala ali je izgleda izabranik ???????? <https://t.co/UhbZbf1t9J>

@BalkanGoran @Blic_online A zar je moguće da Milanka Mara bude 2x na listi SNS sa sve djeverom Dragomircetom???

Тумачење: Употреба упитника у тексту.

5. Лексички маркери на основу којих се могу препознати иронија и сарказам у исказима

Пример:

@BirdLifeSerbia Баш ме занима да ли ћете нешто да напишете о овоме... @Blic_online @ALOnovine @RatkoDmitrovic @InformerNovine @KurirVesti

Тумачење: Уобичајени интензификатора као што су „баш“ „стварно“, „невероватно“.

Примери:

Kako mali Đokica zamišlja #inostranstvo?! #novinari
@Bliconline <https://t.co/PloeO0M5k8>

@Blic_online spaskovski.gde ga samo nađoše majke ti.zlo .

Тумачење: Устаљене конструкције „Како мали Ђокица (Перица) замишља...“ „Чуди ме да...“, „Има или ико нормалан да...“ „где га нађоше“ и сл. У првом примеру је препозната иронија јер се ситуација предочава позитивно а мисли се негативно („замишља“ уместо „фантазира или сагледава искривљено“).

6. Коришћење оксиморона (неспојивих, контрадикторних речи у низу) или неспојивих делова исказа

Примери:

@DCrnjakovic odlicno sam, 37,2 ako te zanima, ali bice bolje, pozdrav, hvala za klik :-)

Bravo za @N1info . Jedini koju su objavili vest o velikom gradonacelniku- kriminalcu!

@b92vesti iznenadice se verovatno i odusevljeni pripadnici saobraćajne policije..A da skratite put pitajte u Dir. za mere i dragocene metale

Тумачење: У првом примеру, аутор је болестан, те се не може осећати добро са високом температуром. Припадници саобраћајне полиције не исказују одушевљење на послу, како је наведено у последњем примеру.

7. Опонашање публицистичког стила у вестима

Примери:

@CEASSerbia saopstenje "Zajednicki zataskavajući poduhvat u slucajevima RTS, Topcider i Leskovac"<https://t.co/HiaD2Kjo9G> @N1infoBG @N1info

@Blic_online Vučić je mudrom politikom osujetio još jedan pokvaren pokušaj Amerike da isprovocira sukobe na severu Kosova i uvuče Srbiju.

8. Употреба закључних реченица у коментару

Примери:

@Blic_online a vi ste list sa "evrop-skim" standardima? Nepriistrasan? Profesionalan!? A zatoooo podrzavate Vucica!! <https://t.co/8VvU6hf96I>"

@DLebedi ...jer ga mrzimo verovatno, nemamo preca posla

@PetarPe22192741 nas zakon je van zakona jer zakon to je ON :/ @DLebedi @N1info @JugoslavCosicN1 @tatjanaaleksic @b92vesti @KurirVesti

Ako je vucic protiv @N1info onda ja smatram da su oni najbolja TV. Kako vucic a ja kontra i nema pogreske!"

@N1info @JugoslavCosicN1 @tatjanaaleksic @b92vesti @avucic količina Kenine gluposti je obrnuto proporcionalna kolicini kose na njegovoj glavi-zaključak posle emisije @N1info . @demokrate.

Тумачење: Изводи се закључак користећи закључне везнике или опонашајући научни стил или афоризме у извођењу закључака. У појединим случајевима аутор се надовезује на саму вест.

9. Коришћење прилога или именских синтагми – „понекад“, „ретко“, „икада“, „најбољи икада“, именских синтагми „свих времена“, „у историји“

Примери:

@b92vesti postoji u teoriji teror vecine, a ovo je ocito nov teorijski doprinos M.O. najveći u istoriji Srbije sto bi rekao njezin šef

Primer da i u @Blic_online ponekada izadje lep i pristojan tekst. Retko ali ipak se desi. <https://t.co/87V6n8iyMx>

Тумачење: У другом примеру, пошто је у питању познат портал вести, коментар да се понекад објави леп и пристојан текст сведочи о иронији.

10. Коришћење суперлатива придева

Пример:

Sutra na @N1info nova zvezda tvitera, heroj opozicije i student sa najboljim prosekom na Magistarskom fakultetu u Čačku Ilic Velimir.

11. Коришћење парафраза у исказу

Примери:

A nadam sevi zelim od sveg srca da cela redakcija @Blic_online ostane bez zivota. Jebem li vam sve mrtvo i zivo. <https://t.co/falJzDu9nD>

@FilipCukanovich @b92vesti to se zove pesnicka sloboda.

Тумачење: Користи се израз „остати без живота“ уместо „умрети“. У другом примеру се „песничка слобода“ употребљава у контексту слободног тумачења чињеница.

12. Ономатопеје којима се дефинише значење исказа

Пример:

@IvanGrujic ti su ga odslužili još 80-ih 70-ih :) Ponovo ista starosna kategorija učestvovala u analizi. cc @Blic_online

Тумачење: Цоктање у знак негодовања.

5. Закључак

У овом раду представљен је СентИроСа корпус, први јавно доступан ручно анотирани корпус за препознавање ироније и сарказма на српском језику. Корпус садржи 1117 твитова који су, поред постојећих ознака сентимента, додатно анотирани ознакама за иронију и сарказам од стране три независна анотатора.

Спроведена квалитативна и квантитативна анализа показала је да препознавање ироније и сарказма представља изазован задатак, што потврђују и резултати анализе слагања анотатора. Анализа образаца неслагања указала је на посебне потешкоће у идентификацији ироније и њеном разграничењу од сарказма, нејасних случајева и одсуства фигуративног језика. Истовремено, идентификовани су лингвистички маркери који могу бити корисни за развој система за аутоматску детекцију ових феномена.

Резултати су показали статистички значајну повезаност између сентимента и расподеле категорија анотације у коначном скупу података. Посматрано појединачно, сарказам је чешће заступљен у текстовима са негативним сентиментом, док је иронија готово равномерно распоређена међу различитим сентименталним поларностима. Ови резултати указују да сентимент може

бити користан додатни показатељ за аутоматско препознавање ових феномена, али сам по себи није довољан за њихову детекцију и разграничење.

Очекујемо да Сентироса унапреди истраживања у области обраде природног језика за српски језик. Једно од ограничења овог ресурса јесте релативно мала величина корпуса, због чега је планирано његово проширивање новим текстовима. У будућем раду планирана је и евалуација корпуса применом модела за аутоматско препознавање ироније и сарказма, укључујући велике језичке моделе. Корпус ће бити коришћен и за истраживање улоге ироније и сарказма у анализи сентимента и детекцији имплицитног увредљивог говора. Будући да је анализа спроведена на текстовима са платформе Твитер/Х, добијени налази и идентификовани лингвистички маркери првенствено одражавају специфичности комуникације на друштвеним мрежама. Стога је потребно испитати њихову применљивост и на другим типовима текстова и у различитим доменима.

СентИроСа је јавно доступан на [ДОИ]: [10.5281/zenodo.21069061](https://doi.org/10.5281/zenodo.21069061).

Захвалност

Захваљујемо се студентима основних и мастер студија Филолошког факултета, који су учествовали у анотацији СентИроСа скупа.

Литература

- Ibrahim Abu Farha, Silviu Vlad Oprea, Steven Wilson, and Walid Magdy. 2022. [SemEval-2022 Task 6: iSarcasmEval, Intended Sarcasm Detection in English and Arabic](#). In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 802–814. Association for Computational Linguistics.
- Vuk Batanović. 2020. *Metodologija rešavanja semantičkih problema u obradi kratkih tekstova napisanih na prirodnim jezicima sa ograničenim resursima*. Ph.D. thesis, Elektrotehnički fakultet u Beogradu.
- Marius Cristian Buzea, Stefan Trausan-Matu, and Traian Rebedea. 2022. [Automatic Fake News Detection for Romanian Online News](#).

- Information*, 13(3):151.
- Angelly Cabrera, Linus Lei, and Antonio Ortega. 2025. [Transfer Learning via Lexical Relatedness: A Sarcasm and Hate Speech Case Study](#). *arXiv preprint arXiv:2508.16555*.
- Paula Carvalho, Luís Sarmiento, Mário J. Silva, and Eugénio De Oliveira. 2009. [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, Hong Kong China. ACM.
- Silvia Casola, Simona Frenda, Soda Marem Lo, Erhan Sezerer, Antonio Uva, Valerio Basile, Cristina Bosco, Alessandro Pedrani, Chiara Rubagotti, Viviana Patti, and Davide Bernardi. 2024. [MultiPICO: Multilingual Perspectivist Irony Corpus](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16008–16021. Association for Computational Linguistics.
- Elena Filatova. 2012. [Irony and sarcasm: Corpus generation and analysis using crowdsourcing](#). In *Proceedings of the eighth international conference on language resources and evaluation (LREC'12)*, pages 392–398, Istanbul, Turkey. European Language Resources Association (ELRA).
- Simona Frenda, Alessandra Teresa Cignarella, Valerio Basile, Cristina Bosco, Viviana Patti, and Paolo Rosso. 2022. [The Unbearable Hurtfulness of Sarcasm](#). *Expert Systems with Applications*, 193:116398.
- Olivera Grljević. 2016. [Sentiment u sadržajima sa društvenih mreža kao instrument unapređenja poslovanja visokoškolskih institucija](#). Ph.D. thesis, University of Novi Sad (Serbia).
- Aditya Joshi, Pushpak Bhattacharyya, and Mark J. Carman. 2018. [Automatic Sarcasm Detection: A Survey](#). *ACM Computing Surveys*, 50(5):1–22.
- Ramona Kühn, Jelena Mitrović, and Michael Granitzer. 2026. [Computational Approaches to the Detection of Lesser-Known Rhetorical Figures: A Systematic Survey and Research Challenges](#). *ACM Computing Surveys*, 58(2):1–36.
- Adela Ljajić and Ulfeta Marovac. 2019. [Improving Sentiment Analysis for Twitter Data by Handling Negation Rules in the Serbian](#)

- Language. *Computer Science and Information Systems*, 16(1):289–311.
- Miljana Mladenović, Cvetana Krstev, Jelena Mitrović, and Ranka Stanković. 2017. [Using Lexical Resources for Irony and Sarcasm Classification](#). In *Proceedings of the 8th Balkan Conference in Informatics*, pages 1–8. ACM.
- Alice R. Myers. 1977. Toward a definition of irony. In Ralph W. Fasold and Roger W. Shuy, editors, *Studies in language variation: semantics, syntax, phonology, pragmatics, social situations, ethnographic approaches*, pages 171–183. Georgetown University Press.
- Silviu Oprea and Walid Magdy. 2020. [iSarcasm: A Dataset of Intended Sarcasm](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1279–1289, Online. Association for Computational Linguistics.
- Matic Rašl, Mitja Žalik, and Vid Keršič. 2021. [Transformer-based Sarcasm Detection in English and Slovene Language](#). In *Proceedings of the 2021 7th Student Computer Science Research Conference (StuCoSReC)*, pages 43–46. University of Maribor Press.
- Antonio Reyes and Paolo Rosso. 2014. [On the Difficulty of Automatically Detecting Irony: Beyond a Simple Case of Negation](#). *Knowledge and Information Systems*, 40(3):595–614.
- Antonio Reyes, Paolo Rosso, and Tony Veale. 2013. [A multidimensional approach for detecting irony in Twitter](#). *Language Resources and Evaluation*, 47(1):239–268.
- Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert, and Ruihong Huang. 2013. [Sarcasm as Contrast between a Positive Sentiment and Negative Situation](#). In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 704–714. Association for Computational Linguistics.
- Victoria L Rubin, Niall Conroy, Yimin Chen, and Sarah Cornwell. 2016. [Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News](#). In *Proceedings of the second workshop on computational approaches to deception detection*, pages 7–17.
- Svetlana Slijepčević Bjelivuk. 2023. [The Ukrainian Crisis on Twitter: Sentiment Analysis – Possible Interpretations](#). *Media Studies and Applied Ethics*, 4(2):43–53.

- Charlotte Taylor. 2017. [The Relationship Between Irony and Sarcasm: Insights from a First-order Metalanguage Investigation](#). *Journal of Politeness Research*, 13(2):209–241.
- Cynthia Van Hee, Els Lefever, and Veronique Hoste. 2018. [SemEval-2018 Task 3: Irony Detection in English Tweets](#). In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 39–50. Association for Computational Linguistics.
- Byron C. Wallace. 2015. [Computational Irony: A Survey and New Perspectives](#). *Artificial Intelligence Review*, 43(4):467–483.

SentIroSa: A Corpus for Irony and Sarcasm Detection in Short Serbian Texts

Danka Jokić, Milica Andrić

Abstract

The detection of irony and sarcasm in text can improve the performance of many Natural Language Processing (NLP) systems, including text summarization, sentiment analysis, and abusive language detection. The development of automatic irony and sarcasm detection systems requires manually annotated datasets for model training and evaluation. However, until now, no publicly available manually annotated corpus for automatic irony and sarcasm detection in Serbian has been available. In this paper, we present SentIroSa, the first publicly available manually annotated corpus for irony and sarcasm detection in Serbian. The corpus was created by extending an existing collection of short texts from Twitter/X, previously annotated for sentiment, with new manual annotations for irony and sarcasm. It consists of 1,117 tweets annotated by three independent annotators. We describe the annotation procedure and present the results of qualitative and quantitative analyses of the corpus, including inter- and intra-annotator agreement, linguistic markers of irony and sarcasm, and the relationship between these phenomena and sentiment. SentIroSa provides a new language resource for research on irony, sarcasm, and other related phenomena in Serbian NLP.

Key words: irony, sarcasm, corpus

Works are licensed under the terms and conditions of the Creative Commons Attribution 4.0 International License.

<https://creativecommons.org/licenses/by/4.0/>

