

*Proceedings of the International Conference
South Slavic Languages in Digital Environment
– JuDig, Vol. I/1: Thematic Collection of
Papers; editors Jasmina Moskovljević Popović
and Ranka Stanković. Belgrade : University,
Faculty of Philology, 2025, 504 pp.*

Милица Динић Маринковић

milica.dinic.marinkovic@fil.bg.ac.rs

ORCID: 0000-0002-2641-8806

Универзитет у Београду

Филолошки факултет

Катедра за општу

лингвистику

Београд, Србија

РАД ПРИМЉЕН:

30. децембар 2025.

РАД ПРИХВАЋЕН:

31. децембар 2025.

Тематски зборник *Proceedings of the International Conference South Slavic Languages in Digital Environment – JuDig, Vol. I/1*¹ објављен је у децембру 2025. године у едицији „Научни скупови“ Филолошког факултета Универзитета у Београду. Тренутно је доступан на званичној интернет страници конференције² и трајно похрањен у DOI репозиторијуму Филолошког факултета Универзитета у Београду.³

1. Proceedings of the International Conference South Slavic Languages in the Digital Environment JuDig : Thematic Collection of Papers / [ed. Jasmina Moskovljević Popović, Ranka Stanković]. - Belgrade : University of Belgrade – Faculty of Philology, 2025. - ISBN 978-86-6153-791-2.

2. Proceedings of the International Conference South Slavic Languages in the Digital Environment JuDig

3. doiFil/Faculty of Philology, University of Belgrade. Proceedings of the International Conference South Slavic Languages in the Digital Environment JuDig

Садржи 29 одабраних прилога, првобитно изложених на међународној научној конференцији „South Slavic Languages in Digital Environment – JuDig“, одржаној 21–23. новембра 2024. године у сурорганизацији Филолошког факултета и Друштва за језичке ресурсе и технологије – ЈеРТех на Филолошком факултету Универзитета у Београду. Ова публикација представља јасно заокружену тематску целину посвећену савременим истраживањима из области језичких технологија, корпусне лингвистике, дигиталне хуманистике и обраде природног језика (NLP), у чијем су фокусу јужнословенски језици, али и са израженом отвореношћу ка ширем, међународном и вишејезичном истраживачком контексту.

Већ сама структура зборника сведочи о интердисциплинарном карактеру савремених лингвистичких истраживања у обради природних језика. Одабрани научни радови повезују теоријска лингвистичка питања са методолошким и технолошким решењима, ослањајући се на савремене приступе у рачунарству, вештачкој интелигенцији и анализи великих количина језичких података. Тематски оквир публикације обухвата кључне сегменте савремене рачунарске лингвистике: изградњу и евалуацију језичких корпуса, системе лингвистичке аотације, развој и примену језичких модела, обраду говора и писаног језика, као и питања употребе дигиталних језичких ресурса у образовању, лексикографији и очувању језичког и културног наслеђа.

Посебно значајну тематску осу зборника чине радови посвећени корпусима као основној инфраструктури савремених истраживања у области обраде природних језика. У раду *Quality Textual Corpora and New South Slavic Language Models* (Mihailo Škorić, Saša Petalinkar) разматра се однос између квалитета текстуалних корпуса и перформанси савремених језичких модела, уз посебан нагласак на изазове које доноси рад са јужнословенским језицима у контексту великих језичких модела. Рад *Sintaksičko-semantička anotacija elektronskog korpusa srpskog jezika* (Nataša Kiš) бави се питањима дизајна и имплементације сложених аотационих система, указујући на значај прецизне и теоријски утемељене аотације за даљу аутоматску обраду и лингвистичку анализу.

Тематика паралелних корпуса и њиховог истраживачког потенцијала за контрастивна истраживања, унапређење машинског превођења, теорију превођења и испитивање специјализованих жанрова такође је заступљена. У раду *Značaj paralelnih korpusa za istraživanje frazemskih konstrukcija u nemačkom i srpskom jeziku* (Kristina

Plić) паралелни корпуси се сагледавају као кључни ресурс за контрастивна истраживања, посебно у домену фразеологије и лексичке семантике. Сличан методолошки оквир, али са фокусом на изградњи специјализованих вишејезичних ресурса, развијен је у раду *Metodologija izrade višezjezičnog paralelnog korpusa na osnovu onlajn digitalnih uputstava za upotrebu: korpus Hilti uputstava* (Nikola Janковић), који детаљно приказује фазе прикупљања, обраде и упаривања података у савременим паралелним корпусима.

Другу велику тематску целину чине радови усмерени ка развоју и евалуацији алата и метода за аутоматску обраду језика. У раду *Exploring the Synergy Between LLMs and Knowledge Graphs for Advanced Abusive Speech Detection in Serbian* (Danka Jokić, Ranka Stanković) разматра се интеграција великих језичких модела и графова знања у задацима детекције увредљивог говора, чиме се савремени приступи машинског учења повезују са симболичким моделима знања. Рад *Нове унапређене верзије програма „Ка минималним паровима“* (Данило Алексић) доноси унапређења постојећих алата за фонолошку и фонетску анализу, показујући како се класична лингвистичка питања могу успешно повезати са савременим софтверским решењима.

Обрада говора и мултимодалних података заступљена је у раду *Употреба модела Whisper Large v3 Sr за транскрипцију говора на српском језику у програмском језику Пајтон на платформи Гугл колаб* (Никола Јанковић, Јована Иваниш) који се бави практичном применом савремених АСР модела у истраживачком и образовном контексту. У сродном али типолошки ширем оквиру, рад *Noun Phrase and Prepositional Phrase Chunking for the Greek Language with spaCy* (Nikitas N. Karanikolas) показује како се савремени NLP алати могу прилагодити језицима са различитим морфосинтаксичким карактеристикама, чиме се додатно наглашава међународна димензија зборника.

Посебну пажњу заслужују радови који повезују језичке технологије са лексикографијом и културним наслеђем. У раду *Korpus Srpsko kao tehnološka osnova za izradu Rečnika savremenog srpskog jezika Matice srpske* (Dušanka Vujović, Branko Milosavljević) корпус се сагледава као темељ савремене лексикографске праксе, док рад *Od rječnika kao korpusa do hiponimije i meronimije* (Маја Матијевић) истражује односе између лексикографских ресурса и семантичких релација, указујући на потенцијал корпуса као извора за теоријску и примењену лексичку семантику. Шири друштвени и образовни контекст примене језичких технологија разматра се у раду *Integration of Serbian Language Version of*

Wikipedia into Educational Systems and the Advancement of Language Technologies (Nebojša Ratković), који анализира улогу отворених дигиталних ресурса у образовању и развоју језичких технологија. Овакви прилози повезују техничке аспекте обраде језика са питањима доступности знања, стандардизације и институционалне подршке развоју језичких ресурса.

Зборник се може посматрати и као својеврсни документ тренутка у којем јужнословенски језици учвршћују своју позицију и повећавају видљивост на мапи актуелних истраживања у области језичких технологија. Прилози овог зборника не само да приказују постојећа достигнућа, већ и јасно указују на правце даљег развоја, потребу за изградњом одрживе истраживачке инфраструктуре и значај континуиране међуинституционалне и међународне сарадње. С тим у вези, *Proceedings of the International Conference South Slavic Languages in Digital Environment – JuDig, Vol. I/1* јасно се профилише као публикација која документује и истовремено гради савремену истраживачку мрежу у области језичких технологија и рачунарске лингвистике на јужнословенском простору. Посебна вредност зборника огледа се у видљивој и садржински утемељеној сарадњи истраживача из Србије, Хрватске и Словеније, који кроз заједничке пројекте, алате и ресурсе доприносе развоју интероперабилне и одрживе инфраструктуре за истраживања јужнословенских језика у дигиталном окружењу.

Овај аспект нарочито долази до изражаја у радовима посвећеним развоју заједничких ресурса и методологија за јужнословенске језике. Поред рада *Quality Textual Corpora and New South Slavic Language Models* (Mihailo Škorić, Saša Petalinkar), који се бави питањима изградње и евалуације корпуса као заједничке основе за језичко моделовање јужнословенских језика, значајан допринос регионалној сарадњи представља и рад *Models for Automatic Morphological Inflection of Serbian and Croatian Based on the srLex and hrLex Morphological Lexicons* (Јака Ћibeј). У овом раду се, полазећи од постојећих морфолошких речника, *srLex* и *hrLex*, развијају модели за аутоматску флективну морфологију српског и хрватског језика, уз ослањање на искуства и алате развијене првобитно за словеначки језик. Тиме се јасно показује како се знања и ресурси развијени за један језик могу систематски преносити и прилагођавати сродним језицима, уз отворени приступ и доступност резултата.

Још један изразит пример овакве сарадње представља рад рад *The ELEXIS-WSD Parallel Sense-Annotated Corpus and South Slavic*

Languages: Subcorpora for Croatian, Serbian, and Slovene (Jaka Čibej, Ranka Stanković, Ana Ostroški Anić, Simon Krek, Carole Tiberius), у коме се описује проширивање паралелног семантички анотираниог корпуса ELEXIS-WSD у оквиру пројекта ELEXIS и COST акције UniDive. Фокус на хрватски, српски и словеначки поткорпус јасно илуструје напоре ка изградњи висококвалитетних ручно анотираних ресурса који су истовремено методолошки усклађени и употребљиви у различитим NLP задацима, укључујући и разрешавање вишезначности. Овакви ресурси имају вишеструки значај: они повезују лексикографске традиције сродних језика, омогућавају контрастивна и вишејезична истраживања и представљају основу за даљи развој технологија.

Регионална димензија сарадње додатно је оснажена и радовима који повезују јужнословенске језике у домену фонологије и обраде језика, попут рада *Sonority Based Syllabic Hyphenation of Macedonian and Serbian* (Katerina Zdravkova, Jana Kuzmanova), у коме се заједнички методолошки оквир примењује на генеалогски блиске језике, указујући на могућности ширег регионалног повезивања.

У посебној тематској равни, рад *О фамилији корпуса савременог српског језика СрпКор* (Душко Витас, Ранка Станковић, Цветана Крстев) пружа драгоцен преглед развоја рачунарске лингвистике у Србији кроз призму еволуције једног од најважнијих језичких ресурса. Приказ развоја СрпКора од промена у алатима и димензијама корпуса, преко проширења временске и жанровске покривености, до све сложенијих нивоа анотације и повезивања са лексичким ресурсима кроз систем *Лексимирка*, не представља само историјски преглед већ и јасан увид у то како се домаћа истраживачка инфраструктура и рачунарска подршка развијала заједно са међународним научним токовима.

Изузетна вредност публикације огледа се у успостављеној равнотежи између изузетно актуелних истраживачких тема из области обраде природног језика, попут великих језичких модела, машинског учења и аутоматске обраде говора, и радова усмерених ка дугорочном развоју језичких ресурса, лексикографији и језичком наслеђу. Сагледан у целини, зборник успешно спаја „горуће“ технолошке приступе са истраживањима која се баве очувањем, документовањем и систематизацијом језичке грађе, чиме нуди истовремено теоријске, методолошке и практично-примењене увиде. Посебна вредност публикације огледа се и у повезивању истраживача из различитих дисциплина и институција, кроз које се гради мрежа стручњака посвећених развоју језичких ресурса и технологија за јужнословенске

језике. Тиме *Proceedings of the International Conference South Slavic Languages in Digital Environment – JuDig, Vol. I/1* превазилази оквире појединачног конференцијског зборника и потврђује се као релевантна референтна тачка за разумевање савременог тренутка и будућих праваца развоја језичких технологија и рачунарске лингвистике у јужнословенском и ширем европском контексту.