

eLex 2025: интелигентна лексикографија између аутоматизације и лингвистичке одговорности

САЖЕТАК: Рад доноси приказ међународне конференције “Electronic Lexicography in the 21st Century (eLex 2025)”, чија је главна тема била интелигентна лексикографија (*Intelligent Lexicography*). Конференција је одржана од 18. до 20. новембра 2025. године на Бледу, у Словенији, и била је посвећена савременим токовима у електронској лексикографији, с посебном пажњом на примени вештачке интелигенције и великих језичких модела у изради, обogaћивању и евалуацији лексикографских ресурса. У раду се разматрају доминантни истраживачки трендови, методолошки изазови и отворена питања која су обележила конференцију, као и позиционирање српске лексикографије у том међународном контексту. Посебан део посвећен је доприносима истраживача из Србије и њиховим искуствима у раду на *Речнику савременог српског језика* и у експериментима с великим језичким моделима у области фразеологије и вишечланих лексичких јединица.

КЉУЧНЕ РЕЧИ: електронска лексикографија, интелигентна лексикографија, дигитална хуманистика, велики језички модели, вишечлане лексичке јединице, “eLex 2025”, српски језик.

РАД ПРИМЉЕН: 18. децембар 2025.

РАД ПРИХВАЋЕН: 30. децембра 2025.

Ранка Станковић

ranka@rgf.rs

ORCID: 0000-0001-5123-6273

*Универзитет у Београду
Рударско-геолошки факултет
Београд, Србија*

Александра Марковић

malexa39@gmail.com

ORCID: 0000-0003-2985-3042

*Српска академија наука и
уметности
Институт за српски језик
Београд, Србија*

1. eLex као референтна тачка савремене лексикографије

Серија конференција “Electronic Lexicography in the 21st Century (eLex)” већ више од десет година представља једно од кључних међународних места за размену идеја у области лексикографије, корпусне лингвистике, језичких технологија и дигиталне хуманистике. Конференција “eLex 2025”, одржана на Бледу, тематски је експлицитно усмерена на појам интелигентне лексикографије, чиме се наглашава потреба да се савремени речници посматрају не само као дигитализовани производи традиционалне лексикографије већ као динамични, технолошки потпомогнути системи знања.

Зборник радова с конференције (Kosem et al. 2025), објављен у отвореном приступу, сведочи о изузетној тематској разноврсности и методолошкој ширини савремене лексикографске заједнице. У њему су заступљени радови који се баве теоријским питањима издвајања значења и структуре речничког описа, али и истраживања усмерена ка практичним решењима, аутоматизацији и евалуацији алата заснованих на методама вештачке интелигенције.

2. Програмска структура конференције

Програм конференције био је пажљиво структуриран тако да омогући уравнотежен дијалог између предавања по позиву, тематских секција, постер и демо сесија, као и радних дискусија. Током три конференцијска дана излагања су организована паралелно у више сала, чиме је обухваћен широк спектар тема из савремене електронске и интелигентне лексикографије. Осим тога, конференцији је претходила радионица “CLASSLA-express 2.0 workshop”, одржана 17. новембра, а након затварања, 20. новембра, одржана је постконференцијска радионица “GWLN-7 Workshop”.

Прва радионица, с поднасловом “Corpora vs. LLMS: a hands-on workshop on using CLASSLA and AI tools for phraseology and lexicography in South Slavic languages”, замишљена је као једна у низу радионица којима је циљ комбиновање могућности које пружају алати за претрагу корпуса са онима које нуде алати вештачке интелигенције, као и испитивање способности корпуса и великих језичких модела (у даљем тексту ВЈМ) у извођењу различитих задатака релевантних за лексикографију и лингвистику. На радионици се испитивала поменута

способност у вези са вишечланим лексичким јединицама (у даљем тексту ВЛЈ) у јужнословенским језицима, углавном с недовољно развијеним језичким ресурсима и технологијама.

Друга радионица, с поднасловом “Globalex Workshop on Lexicography and Neology”, део је низа GWLN радионица и на њој се полазницима представљају методе и технике за проналажење, истраживање, класификацију и опис неологизама у контексту њихове лексикографске обраде и приступа у различитим језицима.

Посебно место у програму заузимала су три предавања по позиву, која су пружила теоријски и методолошки оквир за даљу дискусију. Предавачи по позиву бавили су се улогом ВЈМ у лексикографији, институционалним искуствима у њиховој примени, као и фундаменталним питањима структуре података у савременим речницима. Ова предавања јасно су указала на померање фокуса са саме дигитализације ка дубљем разумевању односа између лингвистичког знања, података и алгоритама.

Прво предавање по позиву одржали су Карол Тибериус и Џес Дос (Carole Tiberius & Jesse de Does), с темом: “LLMs and lexicography at the Dutch language Institute” (оба аутора као афилијацију наводе управо овај институт). Аутори су представили рад у институту који има дугу традицију израде речника, али и речничких база за холандски језик (и неке друге с њим повезане језике). Лексикографски рад је рачунарски подржан, али се значајан део посла и даље обавља манулено. Тренутно се врше истраживања посвећена начинима на које нове технологије, укључујући и ВЈМ, могу бити употребљене за побољшање појединих фаза лексикографског посла, али тако да се не угрозе квалитет података и њихова поузданост. У излагању је поменуто на који начин се испитује могућност имплементације процедуре Pattern Analysis, коју је осмислио чувени лингвиста Patrick Hanks.

Марко Робник-Шикоња (са Универзитета у Љубљани) изложио је тему “Large Language Models for Lexicography”. Истакао је да ВЈМ редефинишу методолошке приступе у многим областима, па и у лингвистици и лексикографији. Модели су преобучени на великим текстуалним корпусима помоћу предвиђања следећег токена и прилагођени за интеракцију са човеком. Нису имуни на халуцинације, због чега је човек и даље неопходан у њиховој примени. У контексту лексикографије модели могу да подрже неколико задатака. Аутор је указао на начин на који се информације из језичке базе података могу употребити за побољшавање ВЈМ у обављању лексикографских

задатака. Методологија коју је представио заснована је на екстракцији графова знања, континуираном претренирању ВЈМ, дизајнирању упита и полуаутоматској евалуацији.

Последње предавање по позиву одржао је Михаел Мехура (Michal Měchura) (Lexical Computing & Dublin City University), аутор система за писање речника Lexonomy i DMLex-a, модерног модела података за лексикографију. У предавању “We need to talk about data structures in Lexicography”, аутор је рекао да је време за критички осврт на моделе података коришћене за представљање речничке структуре у системима за писање речника и другим софтверима. Укратко је представио историју моделовања лексикографских података и истакао да је без обзира на коришћену нотацију (XML, JSON i др.), у основи образац са структуром стабла, у ком су различите структурне јединице (одреднице, значења, дефиниције) уређене у виду хијерархије родитељ–дете. Аутор сматра да образац у виду стабла није довољно изражајан за све појаве у речницима (кросреференцијално упућивање, смештање вишечланих пододредница, сложена хијерархија подзначења) и да би се то лакше представило на основу структуре података засноване на графовима (као што су релационе базе података или граф знања у стилу семантичког веба). Мехура сматра да речнички пројекти који инсистирају само на моделу података заснованом на стаблу не користе у пуној мери могућности дигиталног медија, као и да прелазак на модел заснован на графовима није једноставан јер је у главама лексикографа и корисника речника укорењено размишљање које се ослања на хијерархију стабла. Аутор на крају представља недавно стандардизован модел података за лексикографију, DMLex, који, будући хибридни модел, има циљ да олакша поменути прелазак тако што комбинује структуре у виду стабла са структуром графа кад је неопходно.

Тематске секције биле су организоване око кључних области: аутоматизације лексикографских процеса, примене вештачке интелигенције и корпусних метода, обраде ВЛЈ и фразеологизама, неологизама, као и корисничких аспеката електронских речника. Оваква организација омогућила је да се сродни радови посматрају у заједничком методолошком оквиру, али и да се јасно уоче разлике у приступима и истраживачким стратегијама.

Значајан део програма чиниле су постер и демо сесије, које су пружиле увид у конкретне алате, ресурсе и експерименталне системе. Оне су омогућиле непосредну размену искустава и дискусију о практичним проблемима имплементације интелигентних

лексикографских решења. Поред тога, групне дискусије и радионице додатно су допринеле интерактивности конференције и размени ставова о отвореним питањима и будућим правцима развоја.

Укупна организација програма “eLex 2025” показала је настојање да се конференција не сведе на низ појединачних излагања, већ да функционише као интегрисани форум у којем се теоријска разматрања, емпиријска истраживања и практичне демонстрације међусобно допуњују.

3. Интелигентна лексикографија као истраживачки оквир

Једно од централних питања које се провлачи кроз велики број радова јесте: шта заправо чини лексикографију “интелигентном”? У дискусијама на конференцији овај појам не своди се на пуко коришћење вештачке интелигенције, већ подразумева промишљену интеграцију алгоритамских метода у лексикографске процесе, уз очување теоријске утемељености и лингвистичке одговорности.

ВЈМ појављују се као моћан али амбивалентан алат. Са једне стране, омогућавају аутоматизовано генерисање дефиниција, примера употребе, па чак и предлога речничке структуре. Са друге стране, бројни радови показују да овакви модели често производе садржај који је формално прихватљив, али семантички непрецизан, недовољно диференциран или неосетљив на културолошке разлике. Управо та граница између ефикасности и поузданости чини један од главних тематских токова конференције “eLex 2025”.

3.1 Вишечлане лексичке јединице као “критична тачка” аутоматизације

Посебно место у зборнику и конференцијским дискусијама заузимају радови посвећени ВЛЈ (полилексемским јединицама), идиомима и фраземима. Ове језичке јединице доследно се појављују као један од највећих изазова за аутоматизовану обраду јер захтевају истовремено разумевање синтаксичке структуре, семантичке некомпозиционалности и прагматичког контекста.

У том смислу, “eLex 2025” јасно показује да су ВЛЈ својеврсни “*laxmus-parig*” за интелигентну лексикографију: управо на њима се

најјасније уочавају границе статистичких и генеративних приступа. Више радова наглашава да ВЈМ могу бити корисни у фазама предселекције или асистенције, али да њихови резултати захтевају пажљиву стручну евалуацију и корекцију.

3.2 Аутоматизација, евалуација и питање контроле

Други важан тематски правац односи се на евалуацију лексикографских алата заснованих на вештачкој интелигенцији. У више радова истакнута је потреба за јасним евалуационим метрикама, златним стандардима и поновљивошћу резултата. Варијабилност излаза ВЈМ, зависност од формулације упита и недостатак транспарентности у погледу извора знања представљају озбиљан изазов за њихову примену у научно утемељеним лексикографским пројектима.

У том контексту, конференција “eLex 2025” јасно сугерише померање фокуса са питања “Да ли користити вештачку интелигенцију?” ка питању како је користити одговорно, контролисано и у складу с лексикографским принципима.

4. COST акције

У раду (Makhachashvili et al. 2025) у оквиру COST акције CA21167 “Universality, diversity and idiosyncrasy in language technology” (UniDive)¹ разматрају се кључни проблеми везани за представљање морфосинтаксичких својстава ВЛЈ у савременим лексикографским ресурсима намењеним примени у обради природног језика. Полазећи од прегледне анализе постојећих лексикона који садрже синтаксичке описе ВЛЈ, рад систематизује актуелне приступе кодирању њихове унутрашње структуре, аргументске структуре, реда речи, дисконтинуираности и глаголских алтернација. Посебан акценат стављен је на значај прецизног синтаксичког описа за различите NLP задатке, чиме се указује на јаз између лексикографске теорије и потреба савремених језичких технологија. Аутори закључују да унапређење и стандардизација синтаксичког описа ВЛЈ представљају кључни предуслов за ефикаснију интеграцију лексикографских ресурса у будуће NLP системе.

У раду (Kallas et al. 2025) представљени су резултати опсежног истраживања спроведеног у оквиру COST акције “European Network

1. COST Action CA21167 UNIDIVE

on Lexical Innovation” (ENEOLI),² које се бави савременим праксама у проучавању и документовању лексичких иновација. На основу анкете спроведене крајем 2024. године, у којој је учествовало 249 испитаника из 50 земаља, рад анализира приступе идентификацији и обради неологизама и нових значења из перспективе лексикографије и терминологије. Посебан акценат стављен је на састав истраживачких тимова, употребу корпуса и дигиталних алата, као и на образовне и професионалне потребе у овој области. Резултати указују на промену профила знања и вештина потребних за рад са лексичким иновацијама, наглашавајући значај интердисциплинарности и континуиране обуке у савременој лексикографској и терминолошкој пракси.

У раду (Lindemann and Salgado 2025) представљене су концептуалне основе и практична реализација платформе Wikibase развијене у оквиру COST акције ENEOLI, намењене заједничком раду на ресурсима из области неологије и лексичких иновација. Платформа интегрише три међусобно повезана ресурса у јединствени граф знања: библиографски и текстуални репозиторијум за истраживања неологије (NeoCorpus), вишејезични метатерминолошки речник за моделовање и усклађивање појмовног апарата (NeoVoc) и колекцију неологизама која документује нове лексичке јединице у различитим европским језицима. Рад показује како се кроз Викибасе омогућава интероперабилно повезивање ових ресурса и њихово објављивање као ФАИР повезаних података, уз опис радних токова и првих резултата. Аутори закључују да овакав приступ може послужити као референтни модел за сличне колаборативне пројекте засноване на графовима знања у области дигиталне лексикографије.

Рад (Beliga and Filipović Petrović 2025) представља хибридни методолошки оквир за идентификацију фразеолошких конструкција, заснован на комбиновању ВЈМ, корпусних метода и стручне лингвистичке евалуације. Полазећи од пилот-студије о хрватским репетитивним конструкцијама, истраживање показује како се вештачка интелигенција може користити за генерисање кандидата, парафраза и упита за корпусну претрагу, док се корпусни подаци користе за емпиријску потврду фреквенције и употребе, уз стални надзор људског експерта. Методологија је илустрована поређењем немачке конструкције *X über X* и њеног хрватског еквивалента *X za X*-ом, при чему се указује на разлике у осетљивости различитих језичких модела на идиомагичност и формалну структуру. Рад наглашава

2. COST Action CA22126 ENEOLI

потенцијал оваквог слојевитог приступа за полуаутоматизовану обраду фразеологије и развој вишејезичних фразеолошких ресурса, али и јасно указује на ограничења и потребу за пажљивим дизајном упита и стручном контролом резултата. Истраживање је везано за резултате COST акције “PhraConRep: A Multilingual Repository of Phraseme Constructions in Central and Eastern European Languages” (CA22115).³

5. Српска лексикографија у међународном контексту

Део програма конференције обележили су и доприноси истраживача из Србије, који су представили резултате рада на савременим лексикографским ресурсима и експериментима са применом ВЈМ. Рад на Речнику савременог српског језика (РССЈ) представљен је као пример систематског лексикографског пројекта који се суочава са изазовима аутоматизације и дигиталне трансформације (Stanković et al. 2025). Дигитални лексикографски ресурс РССЈ намењен је и људским корисницима, путем веб и мобилних апликација, као и машинама, кроз програмске интерфејсе (API-је). Пројекат, који координирају удружење из дијаспоре “Окупљени око језика” и Друштво за језичке ресурсе и технологије (JePTex), има за циљ израду речника од око 50.000 фреквентних лексичких јединица које одражавају употребу језика у последњих педесетак година у различитим функционалним стилевима. Листа одредница (лема) аутоматски је екстрахована из корпуса српског језика СрпКор и затим је ручно коригована. Костур речничких чланака је аутоматски креиран и обогаћен подацима из лексичке базе LeXimírka. Посебна пажња посвећена је аутоматизацији лексикографског процеса, укључујући примену статичких векторских репрезентација и великих језичких модела за идентификацију синонима и израду почетних верзија дефиниција, као и аутоматску екстракцију честих спојева речи, синтаксичких образаца и кратких илустративних примера. Рад показује да овакви приступи могу убрзати поједине фазе израде речника, али да истовремено намећу потребу за прилагођавањем лексикографских радних токова и редефинисањем традиционалне лексикографске праксе.

У раду (Marković and Stanković 2025) се разматрају могућности и ограничења примене ВЈМ у савременој описној лексикографији,

3. COST Action CA22115, PhraConRep

полазећи од концепта лексикографије накнадног редиговања (engl. *post-editing lexicography*), у коме лексикограф преузима улогу уредника аутоматски генерисаних речничких нацрта. На основу експеримента спроведеног над скупом од готово 400 ВЛЈ са идиоматизованим значењем, рад испитује употребљивост различитих ВЈМ у задацима идентификације идиома, издвајања значења, писања дефиниција и генерисања илустративних примера за српски језик. Посебан акценат стављен је на разлику између примера који илуструју идиоматизовано значење израза од оних у којима је дато значење парафразирано, као и на поређење перформанси више савремених модела у зависности од формулације упита и начина приступа (интерактивни чет наспрам API-ја). Резултати указују да, упркос значајном напретку, примена великих језичких модела у лексикографији и даље захтева пажљиву евалуацију и снажну стручну интервенцију, нарочито у обради семантички и прагматички сложених језичких јединица.

Представљени радови показали су да се српска лексикографија активно укључује у савремене међународне токове, али и да задржава критички однос према непоузданој примени нових технологија.

6. Закључна разматрања

Конференција “eLex 2025” јасно показује да се савремена лексикографија налази на прекретници. Вештачка интелигенција и велики језички модели нуде снажне алате за убрзавање и обогаћивање лексикографског рада, али истовремено отварају низ теоријских, методолошких и етичких питања. “Паметни речници” не настају аутоматски применом интелигентних алгоритама; они захтевају промишљен дизајн, јасне стандарде и стално присуство људске експертизе.

У том смислу, интелигентна лексикографија, како је представљена на конференцији “eLex 2025”, не означава крај традиционалне лексикографије већ њену трансформацију у оквиру дигиталне хуманистике, дисциплине која повезује технологију, језик и хуманистичко знање на нов начин. У том процесу, улога лексикографа остаје кључна и незаменљива: без њиховог теоријског знања, интерпретативног знања и одговорности у доношењу одлука, лексикографија се своди на техничку обраду језика, а не на његово систематско и поуздано описивање. Управо људска експертиза

обезбеђује да аутоматизација служи лексикографији, а не да је замењује.

Захвалност

Овај рад је финансијски подржало Министарство науке Републике Србије (уговор бр. 451-03-136/2025-03/200174) и Фонд за науку Републике Србије (пројекат бр. 7276, Text Embeddings – Serbian Language Applications, TESLA), као и COST акција CA21167 – Universality, Diversity, and Idiosyncrasy in Language Technology (UniDive).

Литература

- Beliga, Slobodan, and Ivana Filipović Petrović. 2025. “AI-and Corpus-Based Strategies for Identifying Phraseme Constructions: A Pilot Study on Croatian Repetitive Constructions.” In *Electronic lexicography in the 21st century (eLex 2025) Intelligent Lexicography*, 95–115. Brno: Lexical Computing CZ sro.
- Kallas, Jelena, Kristina Koppel, Kris Heylen, Ilan Kernerman, Ana Ostroški Anić, Federica Vezzani, and Špela Arhar Holdt. 2025. “Neology in Practice: Lexicographic and Terminological Approaches to Lexical Innovation.” In *Electronic lexicography in the 21st century (eLex 2025) Intelligent Lexicography*, 652–672. Brno: Lexical Computing CZ sro.
- Kosem, Iztok, Miloš Jakubiček, Marek Medveď, Karolina Zgaga, Špela Arhar Holdt, Tina Munda, and Ana Salgado, eds. 2025. *Electronic lexicography in the 21st century (eLex 2025): Intelligent lexicography. Proceedings of the eLex 2025 conference. Bled, 18–20 November 2025*. Bled: Lexical Computing CZ s.r.o. <https://elex.link/elex2025/proceedings/>.
- Lindemann, David, and Ana Salgado. 2025. “ENEOLI Wikibase: A collaborative working platform for the European Network on Lexical Innovation.” In *Electronic lexicography in the 21st century (eLex 2023): Intelligent lexicography. Proceedings of the eLex 2025 conference*, 285–296. Bled: Lexical Computing CZ s.r.o. <https://doi.org/10.5281/zenodo.17583262>.

- Makhachashvili, Rusudan, Verginica Mititelu, Voula Giouli, Gražina Korvel, Chaya Liebeskind, Irina Lobzhanidze, Stella Markantonatou, Alexandra Markovic, and Ivelina Stoyanova. 2025. “The Challenges of Syntactic Descriptions of Multiword Expressions in Electronic Lexicography.” In *Electronic lexicography in the 21st century (eLex 2025) Intelligent Lexicography*, 252–269. Brno: Lexical Computing CZ sro.
- Marković, Aleksandra, and Ranka Stanković. 2025. “So Close but Still Far: Case Study on Application of LLMs in Idioms Identification, Definition, and Generation of Illustrative Examples.” In *Electronic lexicography in the 21st century (eLex 2025) Intelligent Lexicography*, 79–94. Brno: Lexical Computing CZ sro.
- Stanković, Ranka, Rada Stijović, Mihailo Škorić, and Cvetana Krstev. 2025. “The Dictionary of Contemporary Serbian Language (RSSJ): Advanced Automation and Other Challenges.” In *Electronic lexicography in the 21st century (eLex 2025) Intelligent Lexicography*, 62–78. Brno: Lexical Computing CZ sro.