



eLex 2025 – Pametni rečnici, tvrdoglavi jezici: Leksikografija u doba velikih jezičkih modela



RANKA STANKOVIĆ, ALEKSANDRA MARKOVIĆ
Seminar Jerteha, 25. 12. 2025.



Electronic Lexicography in the 21st Century (eLex 2025)

- ▶ eLex svake druge godine, počev od 2009
- ▶ jedno od ključnih međunarodnih mesta za razmenu ideja u oblasti leksikografije, korpusne lingvistike, JT i DH
- ▶ eLex2025 **inteligentna leksikografija** – savremeni rečnici ne samo kao digitalizovani proizvodi tradicionalne leksikografije već kao dinamični, tehnološki potpomognuti sistemi znanja



Electronic lexicography in the 21st century (eLex 2025)

Intelligent Lexicography

Proceedings of the eLex 2025 conference

edited by Iztok Kosem
Miloš Jakubiček
Marek Medved
Karolina Zgaga
Spela Arhar Holdt
Tina Munda
Ana Salgado

Programska struktura konferencije

- ▶ predavanja po pozivu
- ▶ tematske sekcije
- ▶ poster i demo sesije
- ▶ radne diskusije
- ▶ 2 radionice

Predavanja po pozivu

- ▶ Carole Tiberius & Jesse de Does: ***LLMs and lexicography at the Dutch language Institute***
 - kako nove tehnologije, uključujući i VJM, mogu biti upotrebljene za poboljšanje pojedinih faza leksikografskog posla, ali tako da se ne ugroze kvalitet podataka i njihova pouzdanost
- ▶ Marko Robnik-Šikonja: ***Large Language Models for Lexicography***
 - VJM redefinišu metodološke pristupe i u lingvistici i leksikografiji; zbog halucinacija čovek i dalje nezaobilazan; informacije iz jezičke baze podataka mogu se upotrebiti za poboljšavanje VJM u obavljanju leksikografskih zadataka; metodologija zasnovana na ekstrakciji grafova znanja, kontinuiranom pretreniranju VJM, dizajniranju upita i poluatوماتskoj evaluaciji.

Predavanja po pozivu

- ▶ Michal Měchura: ***We need to talk about data structures in Lexcography*** (Lexonomy, DMLex)
 - rečnički projekti koji insistiraju samo na modelu podataka zasnovanom na stablu ne koriste u punoj meri mogućnosti digitalnog medija, a prelazak na model zasnovan na grafovima nije jednostavan – ukorenjeno razmišljanje koje se oslanja na hijerarhiju stabla.
 - **DMLex kao hibridni model**, koji ima za cilj da olakša pomenuti prelazak tako što kombinuje strukture u vidu stabla sa strukturom grafa kad je neophodno.

Predavanja po pozivu

- ❑ uloga VJM u leksikografiji
- ❑ institucionalna iskustva u njihovoj primeni
- ❑ fundamentaln pitanja strukture podataka u savremenim rečnicima

predavanja su jasno ukazala na **pomeranje fokusa** sa same digitalizacije ka dubljem razumevanju odnosa između lingvističkog znanja, podataka i algoritama

Tematske sekcije

- ▶ automatizacija leksikografskih procesa
- ▶ (kombinovana) primena VI i korpusnih metoda
- ▶ obrada VLJ i frazeologizama, neologizama
- ▶ korisnički aspekti elektronskih rečnika

Radionice

- ▶ **CLASSLA-express 2.0 workshop – Corpora vs. LLMs: a hands-on workshop on using CLASSLA and AI tools for phraseology and lexicography in South Slavic languages (17. 11. 2025)**

(kombinovanje mogućnosti koje nude alati za pretragu korpusa s mogućnostima alata VI u vezi s pitanjima relevantnim za leksikografiju, MWE u južnoslovenskim jezicima)

- ▶ **GWLN-7 Workshop – Globalex Workshop on Lexicography and Neology (20. 11. 2025)**

metode i tehnike za pronalaženje, istraživanje, klasifikaciju i opis neologizama u kontekstu njihove leksikografske obrade i pristupa u različitim jezicima

Poster i demo sesije, radne diskusije

- ▶ uvid u konkretne alate, resurse i eksperimentalne sisteme; neposredna razmena iskustava i diskusija o praktičnim problemima implementacije inteligentnih leksikografskih rešenja
- ▶ Program eLex 2025 odraz nastojanja da se konferencija ne svede na niz pojedinačnih izlaganja, već da funkcioniše kao **integrisani forum** u kojem se teorijska razmatranja, empirijska istraživanja i praktične demonstracije međusobno dopunjuju.

Šta leksikografiju čini inteligentnom?

- ▶ Ne samo upotreba VI već promišljena **integracija** algoritamskih metoda u leksikografske procese, uz očuvanje **teorijske utemeljenosti i lingvističke odgovornosti**
- ▶ VJM omogućavaju automatizaciju različitih delova leksikografskog procesa, ali je **sadržaj često samo formalno prihvatljiv** (semantički neprecizan, neosetljiv na kulturološke razlike).
- ▶ **granica između efikasnosti i pouzdanosti** čini jedan od glavnih tematskih tokova konferencije eLex 2025

VLJ kao kritična tačka automatizacije

- ▶ Jedan od **najvećih izazova za automatizovanu obradu** (moraju se razumeti semantička nekompozicionalnost, sintaksička struktura, diskontinuiranost, pragmatički kontekst)
- ▶ VLJ su **lakmus papir** za inteligentnu leksikografiju – upravo na njima se najjasnije uočavaju granice statističkih i generativnih pristupa

Automatizacija, evaluacija i pitanje kontrole

- ▶ evaluacija leksikografskih alata zasnovanih na VI – potreba za jasnim evaluacionim metrikama, zlatnim standardima i ponovljivošću rezultata
- ▶ varijabilnost izlaza VJM, zavisnost od formulacije upita i nedostatak transparentnosti u pogledu izvora znanja predstavljaju ozbiljan izazov za njihovu primenu u naučno utemeljenim leksikografskim projektima
- ▶ fokus se pomera s pitanja: **Da li koristiti veštačku inteligenciju?** na pitanje: **Kako koristiti VI odgovorno, kontrolisano i u skladu s leksikografskim principima?**

Akcije COST – Universality, diversity and idiosyncrasy in language technology (UniDive)

- ▶ Barbu Mititelu et al. (2025) *The Challenges of Syntactic Descriptions of Multiword Expressions in Electronic Lexicography*
 - ključni problemi vezani za predstavljanje morfosintaksičkih svojstava VLJ u savremenim leksikografskim resursima namenjenim primeni u NLP
 - unapređenje i standardizacija sintaksičkog opisa VLJ predstavljaju **ključni preduslov** za efikasniju integraciju leksikografskih resursa u buduće NLP sisteme

Akcije COST – European Network on Lexical Innovation (ENEOLI)

- ▶ Kallas et al. (2025) *Neology in Practice: Lexicographic and Terminological Approaches to Lexical Innovation*
- ▶ Na osnovu ankete sprovedene krajem 2024. godine, u kojoj je učestvovalo 249 ispitanika iz 50 zemalja, rad analizira pristupe identifikaciji i obradi neologizama i novih značenja iz perspektive leksikografije i terminologije.
- ▶ promena profila znanja i veština potrebnih za rad sa leksičkim inovacijama; značaj interdisciplinarnosti i kontinuirane obuke u savremenoj leksikografskoj i terminološkoj praksi

Akcije COST – European Network on Lexical Innovation (ENEOLI)

- Lindemann & Salgado (2025) *ENEOLI Wikibase: A collaborative working platform for the European Network on Lexical Innovation*

Wikibase – platforma integriše tri međupovezana resursa u jedinstveni graf znanja – interoperabilno povezivanje resursa i njihovo objavljivanje kao FAIR povezanih podataka.

ovakav pristup može poslužiti kao referentni model za slične kolaborativne projekte zasnovane na grafovima znanja u oblasti digitalne leksikografije

Akcije COST – PhraConRep: A Multilingual Repository of Phraseme Constructions in Central and Eastern European Languages



- ▶ Beliga i Filipović-Petrović (2025) *AI- and Corpus-Based Strategies for Identifying Phraseme Constructions: A Pilot Study on Croatian Repetitive Constructions*
- **hibridni** metodološki okvir za identifikaciju frazeoloških konstrukcija, zasnovan na kombinovanju VJM, korpusnih metoda i stručne lingvističke evaluacije
- polazeći od pilot-studije o repetitivnim konstrukcijama (poređenje nemačke konstrukcije *X über X* i hrvatskog ekvivalenta *X za X-om*) autori pokazali kako se VI mogu generisati kandidati, parafraza i upiti za pretragu korpusa, dok se korpusni podaci koriste za empirijsku potvrdu frekvencije i upotrebe, uz stalni nadzor eksperta

Srpska leksikografija u međunarodnom kontekstu

Stanković et al. (2025) *The Dictionary of Contemporary Serbian Language (RSSJ): Advanced Automation and Other Challenges*

- ▶ *Okupljeni oko jezika i Društvo za jezičke resurse i tehnologije (JeRTeh)*
- ▶ Rečnik savremenog srpskog jezika (RSSJ) – digitalni leksikografski resurs namenjen i ljudskim korisnicima, putem veb i mobilnih aplikacija, kao i mašinama, kroz programske interfejsse (API-je)
- ▶ 50.000 frekventnih leksičkih jedinica, upotreba jezika u poslednjih pedesetak godina u različitim funkcionalnim stilovima

Automatizacija

- ▶ lista lema automatski ekstrahovana iz SrpKor-a (ručno korigovana)
- ▶ kostur rečničkih članaka automatski kreiran i obogaćen podacima iz LeXimirke
- ▶ automatizacija leksikografskog procesa (statičke vektorske reprezentacije i VJM za identifikaciju sinonima, izrada početnih verzija definicija, kao i automatska ekstrakcija čestih spojeva reči, sintaksičkih obrazaca i kratkih ilustrativnih primera
- ▶ ovakvi pristupi mogu ubrzati pojedine faze, ali nameću potrebu za **prilagođavanjem** leksikografskih radnih tokova i **redefinisanjem** tradicionalne leksikografske prakse

Srpska leksikografija u međunarodnom kontekstu

- ▶ Marković & Stanković (2025) *So Close but Still Far: Case Study on Application of LLMs in Idioms Identification, Definition, and Generation of Illustrative Examples*
- mogućnosti i ograničenja primene VJM u savremenoj opisnoj leksikografiji

VJM – plusevi i minusi

- ▶ Minusi: lažna polisemija (Jakubiček & Rundell, 2023); kvalitet ilustrativnih primera
- ▶ Plusevi: pisanje definicija(ChatGPT (Plus) generisao definicije u vidu punih rečenica (COBUILD) (Lew, 2023), dostupnost, poboljšavanje upita, stalni razvoj

Mali rečnik srpskih idioma

- ▶ 400 idioma
- ▶ izdvajanje značenja
- ▶ definicije na standardnom srpskom jeziku
- ▶ dve vrste ilustrativnih primera:
 - 1) sa idiomom
 - 2) s njegovom **parafrazom**



parafraza pokazuje na koji način VJM razumeju idiome (Gantar, 2024):

- ❑ ChatGPT često daje doslovnu parafrazu za idiome koje ne razume (ili razume pogrešno)

Zaključak

- ▶ Iako su VJM već “vrlo blizu” pružanju značajne podrške za frazeološku leksikografiju u srpskom jeziku, moraju se rešiti značajni problemi
- ▶ buduća istraživanja treba da se usmere na izdvajanje značenja, smanjivanje halucinacija i razvijanje ciljanih evaluacionih metrika za idiomatske izraze u srpskom jeziku

I na kraju

- ▶ savremena leksikografija na **prekretnici** – VI i VJM nude moćne alate za ubrzavanje i obogaćivanje leksikografskog rada, ali istovremeno otvaraju niz teorijskih, metodoloških i etičkih pitanja.
- ▶ „pametni rečnici“ ne nastaju automatski primenom inteligentnih algoritama; oni zahtevaju promišljen dizajn, jasne standarde i stalno prisustvo ljudske ekspertize
- ▶ **inteligentna leksikografija** ne označava kraj tradicionalne leksikografije, već njenu **transformaciju** u okviru DH
- ▶ **uloga leksikografa** ostaje ključna i nezamenljiva – automatizacija u službi leksikografije





Šrećna nova 2026. godina!

