



**Odbor za leksikografiju Razreda za filološke znanosti Hrvatske akademije znanosti i umjetnosti**

**Projekt *Digitalni hrvatski frazeološki rječnik***

**Zavod za lingvistička istraživanja Hrvatske akademije znanosti i umjetnosti**

## **Okrugli stol**

### **Frazeologija i elektronička leksikografija: aktualna istraživanja i metode**

(7. listopada 2022., Knjižnica Hrvatske akademije znanosti i umjetnosti)

#### **Program**

Pozdravne riječi

11:00 – 11:15

Akademik Ranko Matasović, tajnik Razreda za filološke znanosti Hrvatske akademije znanosti i umjetnosti

Akademik Milan Mihaljević, voditelj Zavoda za lingvistička istraživanja Hrvatske akademije znanosti i umjetnosti

Dr. sc. Amir Kapetanović, predsjednik Odbora za leksikografiju Razreda za filološke znanosti Hrvatske akademije znanosti i umjetnosti

Izlaganja

11:15 – 11:45 Iztok Kosem: *Povezovanje in souporaba slovarskih podatkov: novi izzivi za digitalno leksikografijo*

11:45 – 12:15 Ivana Filipović Petrović: *Leksikografsko-lingvistički projekt u elektroničko doba*

12:15 – 12:45 Stanka za kavu

12:45 – 13:15 Polona Gantar: *Frazeologija in digitalna leksikografija*

13:15 – 13:45 Jelena Parizoska: *Mrežni frazeološki rječnik hrvatskoga jezika: metode izrade i leksikografska obrada*

## Sažeci izlaganja

Iztok Kosem

Institut Jožef Stefan i Filozofski fakultet Sveučilišta u Ljubljani

### **Povezovanje in souporaba slovarskih podatkov: novi izzivi za digitalno leksikografijo**

Leksikografija oz. izdelava slovarjev je v zadnjih desetletjih doživela marsikatero spremembo in izzive. V zadnjem času so izzivi vezani predvsem na uporabo novih metod (besedne vložitve, nevronske mreže, množičenje) in novih načinov organizacije podatkov, kjer je vse večji poudarek na vpeljavi relacijskih baz ter povezovanju in souporabi različnih slovarskih in drugih virov. To terja ponovne razmisleke o pristopih k načrtovanju celotnega leksikografskega procesa, saj posamezni slovarski projekti (in koraki njihove izdelave) nikakor niso več neodvisni. To po eni strani prinaša številne prednosti, kot je npr. hitrejša posodobljanje podatkov v različnih virih, po drugi pa tudi preizkušnje za leksikografe, ki se morajo med drugim navaditi na vse večjo raznolikost in razdrobljenost njihovega dela.

V svoji predstavitvi bom predstavil izdelavo slovarjev in ostalih jezikovnih virov na Centru za jezikovne vire in tehnologije Univerze v Ljubljani, ter na posameznih primerih izpostavil konkretne izzive in rešitve pri leksikografskem delu. V zaključku bom podal predloge za posodobitev sheme procesa izdelave spletnih slovarjev, kot ga najdemo v Klosa (2013).

### **Linking and sharing dictionary data: new challenges for digital lexicography**

Lexicography or dictionary-making has experienced numerous changes and challenges in recent decades. Lately, these challenges are mainly related to the introduction of new methods (word embeddings, neural networks, crowdsourcing) and new ways of data organization, with more and more emphasis on the introduction of relational databases, linking different dictionaries, and sharing data across several dictionaries and related language resources. This requires a rethink of approaches to planning the entire lexicographic process, given that individual dictionary projects (and steps of their compilation) are no longer independent. On the one hand, this brings several advantages, such as quicker updating of data in different resources, and on the other hand, it introduces new demands for lexicographers who have to get used to increasing diversity and fragmentation of their work.

In my presentation, I will present the compilation of dictionaries and other language resources at the Centre for Language Resources and Technologies University of Ljubljana. I will give different examples of concrete challenges and solutions used in our lexicographic work. I will conclude by offering suggestions on how Klosa's (2013) diagram depicting a lexicographical process for online dictionaries could be improved.

Ivana Filipović Petrović

Zavod za lingvistička istraživanja Hrvatske akademije znanosti i umjetnosti, Zagreb

### **Leksikografsko-lingvistički projekt u elektroničko doba**

Projekti izrade rječnika često su u povijesti leksikografije nadživjeli svoje leksikografe. Digitalne mogućnosti ubrzale su sve: prikupljanje jezičnih podataka za rječnik, izradu rječnika i njegovo objavljivanje. Brzina i elektronički medij ipak nisu izbrisali neke trajne okupacije leksikografa, poput statusa višerječnica s figurativnim značenjem u rječniku. Štoviše, donijeli su nove izazove: traže inovativna rješenja i drukčije leksikografske vještine.

U ovome izlaganju, predstavlja se nastajanje i rad jednog suvremenog leksikografskog projekta u sklopu kojeg se izrađuje izvorno elektronički frazeološki rječnik u otvorenom pristupu. Posebna se pažnja posvećuje digitalnim alatima i resursima koji se koriste u radu, zatim metodama obrade korpusnih podataka i oblikovanju tehnički naprednijih rješenja za *stare* frazeografske probleme, poput izravnog pristupa cijelome frazemu već na razini makrostrukture, neograničenih prostornih rješenja i brzog pretraživanja prema svim sastavnicama. Konačno, predstaviti će se projektne aktivnosti te budućnost projekta.

Polona Gantar

Filozofski fakultet Sveučilišta u Ljubljani

### **Frazeologija in digitalna leksikografija**

V svoji predstavitvi obravnavam povezavo med frazeologijo kot jezikoslovno vedo na eni strani in digitalno leksikografijo, ki predstavlja združitev jezikovnotehnoloških in leksikografskih pristopov, na drugi. Teme, ki povezujejo vse tri vidike: jezikoslovnega, računalniškega in leksikografskega, se nanašajo na definicijo frazeoloških enot in njenih tipov ter na vprašanja leksikalne semantike in pragmatike. Računalniška potreba po razumevanju jezikovnih podatkov potrebuje zapis frazeološke enote v obliki, ki jo je mogoče računalniško obdelati: med drugim formalizacijo zgradbe frazeološke enote in določitev kanonične oblike. Zadnji, vendar ne najmanj pomemben, je leksikografski vidik, ki zahteva ureditev podatkov o frazeološki enoti v okvir slovarskega gesla. Ta naloga sama po sebi odpira številna vprašanja leksikografskega urejanja digitalnega gradiva, ki ga leksikografi (še vedno) razumemo predvsem hierarhično.

V prispevku se bom na primeru luščenja frazeoloških enot iz korpusa za namene izdelave Leksikona FE osredotočila na zapis frazeološke enote v računalniško berljivi obliki, zlasti na pravila zapisa v kanonični obliki in v razmerju do variant in skladijskih pretvorb.

### **Phraseology and digital lexicography**

In my presentation, I discuss the link between phraseology as a linguistic discipline on the one hand and digital lexicography, which is a fusion of language-technological and lexicographical approaches, on the other. The themes that link all three aspects: linguistic, computational, and lexicographic, concern the definition of the phraseological unit and its types, as well as issues of lexical semantics and pragmatics. The computational need to understand linguistic data requires a record of the phraseological unit in a form that can be processed by a computer: among other things, the formalisation of the structure of the phraseological unit and the definition of the canonical form. Lastly, the lexicographic aspect requires organising the linguistic data into a dictionary entry. This task in itself raises several issues of lexicographic organisation of digital data, which is (still) understood by lexicographers mainly in hierarchical terms.

In this presentation, I will focus on the case of the extraction of phraseological units from a corpus to create the PU Lexicon, particularly on the rules of recording a phraseological unit in a canonical form concerning variants and syntactic transformations.

Jelena Parizoska

Učiteljski fakultet Sveučilišta u Zagrebu

### **Mrežni *Frazeološki rječnik hrvatskoga jezika*: metode izrade i leksikografska obrada**

U izlaganju će se prikazati metode izrade mrežnog *Frazeološkog rječnika hrvatskoga jezika* (Filipović Petrović i Parizoska 2022) te leksikografska obrada triju elemenata, a to su ustaljene varijante, kreativna upotreba frazema i konceptualna (tematska) organizacija.

Bilježenje frazeoloških varijanata temelji se na teorijskim postavkama kognitivne gramatike. U ovome frazeološkom e-rječniku varijantni se oblici daju u istoj natuknici, a njihov redosljed i način bilježenja zasnivaju se na frekvencijskim podacima iz korpusa hrWaC. Informacije o kreativnoj upotrebi daju se za one frazeme čiji leksički sastav, strukturu i(li) značenje govornici često mijenjaju po određenom obrascu prema podacima iz hrWaC-a. U rječniku se kreativna upotreba bilježi u posebnim odjeljcima na kraju natuknice te se daju primjeri. Na kraju nekih natuknica upućuje se na druge frazeme koji imaju slično i(li) suprotno značenje, a korisnik im može pristupiti klikom na poveznice. Frazemi se u tematske skupine grupiraju na temelju triju kriterija – značenja, strukture i upotrebe. Konceptualna (tematska) organizacija rječničke građe korisniku omogućuje da izrazi svoje ideje upotrebom figurativnog jezika.

Na općenitoj razini, pokazat ćemo da su za korisnike najveće prednosti elektroničkog frazeološkog rječnika lako pretraživ sadržaj i bolja preglednost natuknice kod navođenja varijantnih oblika. Također, nova rješenja e-leksikografije omogućuju uključivanje informacija koje nadilaze one uobičajene o ustaljenim oblicima, značenju i upotrebi, a to su česta ili specifična upotreba, podrijetlo frazema, kreativna upotreba te značenjski povezani frazemi.