



Centar za jezična istraživanja
Center for Language Research
Sveučilište u Rijeci
Filozofski fakultet
University of Rijeka
Faculty of Humanities
and Social Sciences



Hrvatsko društvo
za primijenjenu lingvistiku
Croatian Applied
Linguistics Society
Kroatische Gesellschaft
für Angewandte Linguistik
Association croate de linguistique appliquée



Laboratorij za istraživanje
kulture složenosti
Cultural Complexity Lab
Sveučilište u Rijeci
Filozofski fakultet
University of Rijeka
Faculty of Humanities
and Social Sciences

4. simpozij SCIMETH

Digitalni alati i resursi u jezikoslovlju

15. svibnja 2019.

Filozofski fakultet u Rijeci, Sveučilišna avenija 4, 51000 Rijeka

PROGRAM

- 9:15 – 9:45 Registracija sudionika
- 9:45 – 10:00 Otvaranje simpozija
- 10:00 – 11:00 **Tomaž Erjavec**, Institut Jožef Stefan, Slovenija
Introduction to the research infrastructure CLARIN
- 11:00 – 11:30 Stanka za kavu
- 11:30 – 12:30 **Nikola Ljubešić**, Institut Jožef Stefan, Slovenija
Repozitorij i mrežni servisi infrastrukture CLARIN.SI
- 12:30 – 13:30 Stanka za zajednički ručak svih sudionika
- 13:30 – 14:15 **Ivana Lalli Pačelat**, Sveučilište u Puli, Fakultet
za interdisciplinarnu, talijansku i kulturološku studije
O specijaliziranim i prijevodnim korpusima i njihovoj uporabi
- 14:15 – 15:00 **Ivana Filipović Petrović**, Zavod za lingvistička istraživanja HAZU i
Jelena Parizoska, Učiteljski fakultet u Zagrebu
*Sketch Engine i Lexonomy: digitalni alati za izradu frazeološkoga
rječnika hrvatskoga jezika*
- 15:00 – 15:30 Stanka za kavu
- 15:30 – 16:15 **Benedikt Perak**, Sveučilište u Rijeci, Filozofski fakultet
Stvaranje i analiza korpusa parlamentarnih debata
- 16:15 – 18:00 Okrugli stol: *Računalni resursi i alati u lingvističkim istraživanjima*
moderator: Benedikt Perak

Simpozij se održava na I. katu Filozofskoga fakulteta, Sveučilišna avenija 4:

- registracija sudionika: pokraj predavaonice F-139

- predavanja i okrugli stol: predavaonica F-139

ČLANOVI ORGANIZACIJSKOGA ODBORA

Mihaela Matešić, predsjednica Hrvatskoga društva za primijenjenu lingvistiku

Tihana Kraš, voditeljica Centra za jezična istraživanja

Anita Memišević, tajnica Hrvatskoga društva za primijenjenu lingvistiku

Anastazija Vlastelić, Regionalni aktiv HDPL-a Rijeka, predsjednica Organizacijskoga odbora

Benedikt Perak, Laboratorij za istraživanje kulturne složenosti, potpredsjednik Organizacijskoga odbora

Martina Blečić, Centar za jezična istraživanja

Cecilija Jurčić Katunar, Centar za jezična istraživanja

Borana Morić-Mohorovičić, Regionalni aktiv HDPL-a Rijeka

Jakob Patekar, Regionalni aktiv HDPL-a Rijeka

Petra Radošević, Regionalni aktiv HDPL-a Rijeka

Diana Stolac, Regionalni aktiv HDPL-a Rijeka

Iris Vidmar, Centar za jezična istraživanja

Zvezdana Vrzić, Centar za jezična istraživanja

Okrugli stol: *Računalni resursi i alati u lingvističkim istraživanjima*

moderator: Benedikt Perak

16.20 – 16.35

SANDRA LJUBAS

mag. philol. germ. et suet./ doktorandica

Poslijediplomski doktorski studij Humanističke znanosti, smjer Filologija – jezikoslovlje

Sveučilište u Zadru

sandra.ljubas1@gmail.com

Izazovi rada s digitalnim alatima i resursima u kontekstu istraživanja strojnog prevođenja

Digitalni alati i resursi osim instrumenta, mogu biti i glavni fokus jezikoslovnih istraživanja. Prvi se izazov istraživanja kvalitete strojnih prijevoda krije u tome što su rezultati uvijek specifični za određen jezični par, karakteristike tih jezika i dostupne digitalne resurse, pa tako katkad radimo s gotovo neupotrebljivim izlaznim podacima, a katkad s prilično dobrima. Drugi je izazov brzo zastarijevanje podataka: zbog čestih ažuriranja strojnoprevoditeljskih sustava (primjerice Google Prevoditelja) rezultati istraživanja do trenutka objavljivanja već mogu prestati vrijediti, a to sa sobom povlači nereplikabilnost istraživanja. Nadalje, za svako je istraživanje vrednovanja strojnih prijevoda prvenstveno važno odabrati dobar paralelan korpus. Repozitoriji poput CLARIN-a uvelike olakšavaju i ubrzavaju proces pripreme istraživanja u usporedbi sa situacijama kada se ručno prikupljaju tekstovi i slažu korpusi samo za potrebe istraživanja. Ipak, valja zapamtiti da se i gotovi, preuzeti paralelni tekstovi smiju i trebaju prilagoditi potrebama pojedinog istraživanja. U preliminarnom istraživanju utjecaja višejezičnosti na ljudsku procjenu kvalitete strojnih prijevoda pokazalo se da odabrani referentni prijevod nije u potpunosti adekvatan i da su odabrani dijelovi teksta pisani prestručnim jezikom, pa je tekst prvo morao biti preuređen i lektoriran. Zaključno se može reći da u Hrvatskoj još uvijek vlada određeno nepovjerenje prema razvijanju jezičnih tehnologija, a tom se problemu može doskočiti samo daljnjim naporima na usavršavanju digitalnih resursa i ohrabivanju i ostalih aktera, poput prevoditelja, da se i njih uključe.

16.40 – 16.55

KATJA NADINE PASSERI

univ. bacc. philol. germ. et hist.

Sveučilište u Rijeci, Filozofski fakultet u Rijeci

katja.passeri@hotmail.com

„Ted Talks“ – nova generacija digitalizacije unutar jezikoslovlja?

Ovo je stoljeće obilježeno je visokom zastupljenošću tehnologije i digitalizacije. To nam omogućava lakši pristup bilo kakvim podacima, pa samim time i podacima vezanim uz jezikoslovlje. Mnogi informatičari i jezikoslovci zajedničkim snagama osmišljavaju raznovrsne programe i sustave, koji

pridonose razvitku jezikoslovlja kao znanosti. „Ted Talks“, odnosno „Ted Conferences“ je organizacija koja na internetskim stranicama (uglavnom na svojoj stranici www.ted.com) objavljuje snimke sa svojih konferencija, koje se održavaju diljem svijeta. Na konferenciji nastupaju vrlo stručni, pomno izabrani znanstvenici, političari, utjecajni pojedinci i umjetnici, koji u 18 minuta prezentiraju svoju određenu tematiku.

Možemo se pitati sljedeće: u kojoj mjeri „Ted Talks“ mogu pridonijeti razvitku jezikoslovlju kao znanosti? Jesu li snimke sa konferencija i stručnih okupljanja nova vrsta digitalizacije znanstvenih otkrića i stručnih radova? Može li snimka sa Tedovih konferencija biti jednako vrijedna kao i stručni ili znanstveni članak digitaliziran u PDF formatu? Istraživanje kojem se bavim pokušava pronaći odgovore na postavljena pitanja. Zaključci će biti prezentirani na konferenciji u Rijeci, koju organizira Hrvatsko društvo za primijenjenu lingvistiku.

16.55 – 17.10

DARIA LAZIĆ

mag. educ. philol. croat. et mag. philol. suet.

Institut za hrvatski jezik i jezikoslovlje

dlazic@ihjj.hr

Jezične tehnologije u leksikografiji na primjeru Hrvatskoga mrežnog rječnika – Mrežnika

U Institutu za hrvatski jezik i jezikoslovlje u okviru projekta Hrvatski mrežni rječnik – Mrežnik, koji financira Hrvatska zaklada za znanost, izrađuje se jednojezični mrežni rječnik hrvatskoga standardnog jezika. Riječ je o rječniku koji izvorno nastaje kao e-rječnik koji će po završetku projekta biti objavljen i slobodno dostupan na mreži, te se stoga sadržajno i strukturno uvelike razlikuje od tiskanih rječnika. Rječnik se piše u programu TLex, a osnovni izvor rječničke građe čine tekstni korpusi hrvatskoga jezika dostupni i pretraživi u programu Sketch Engine. U izlaganju će se ukratko predstaviti projekt Mrežnik te neki problemi povezani s uporabom jezičnih tehnologija u leksikografskome radu, kao što su pitanje homonimije, polisemije te odnosa standardnih i nestandardnih oblika u korpusu. Osim toga, spomenut će se i neke prednosti uporabe suvremenih jezičnih izvora i alata, primjerice uvid u nova značenja. Jezične tehnologije za hrvatski jezik koje se rabe pri izradi hrvatskoga mrežnog rječnika usporedit će se s izvorima i alatima za skandinavske jezike, s naglaskom na mrežnim izvorima za švedski jezik Sveučilišta u Göteborgu.

17.10 – 17.25

JOSIP MIHALJEVIĆ

doktorand Instituta za hrvatski jezik i jezikoslovlje na projektu Mrežnik

Institut za hrvatski jezik i jezikoslovlje

jmihalj@ihjj.hr

Uporaba računalne tehnologije za izradu jezičnih igara (na primjeru igrifikacije Hrvatskoga mrežnog rječnika - Mrežnika)

U radu će se predstaviti računalne jezične igre izrađene na temelju sadržaja hrvatskoga mrežnog rječnika Mrežnik. Izlaganje će se u prvome redu temeljiti na predstavljanju tehnologije i digitalnih resursa koji se koriste za izradu tih igara. Mrežnik će imati tri modula: opći modul, modul za djecu i modul za strance. Mrežnik se izrađuje u programu TLex s pomoću Sketch Enginea. Igre moraju svojim sadržajem biti prilagođene ciljnoj skupini. Objasnit će se koji će se tipovi igara koristiti za usvajanje različitih jezičnih sadržaja poput gramatike, značenja riječi i pravopisa. Prikazat će se i implementacija suvremenih elemenata igrifikacije unutar igara rječnika poput bodovanja, mrežnoga natjecanja i tablica s rezultatima. Ti elementi igrifikacije koji nagrađuju točne odgovore mogu uključiti više igrača u igru te ih potaknuti da ostvare bolje rezultate u igri ili da češće igraju određenu igru. Tu je važno da igra bude dizajnirana tako da igrač više uči time što više igra igru, pa mehanike igranja moraju biti spojene s obrazovnim sadržajima. Također, će se prikazati primjeri jezičnih igara i igrifikacije koji se primjenjuju pri igrifikaciji stranih leksikografskih izdanja poput igra na stranici Merriam-Webster i Vocabulary.com.

RASPRAVA