

Ocene in poročila – Reviews and Reports

Konferenca Jezikovne tehnologije in digitalna humanistika 2022

15. in 16. septembra je potekala že trinajsta konferenca Jezikovne tehnologije in digitalna humanistika, ki jo vsaki dve leti organizira Slovensko društvo za jezikovne tehnologije (SDJT) v sodelovanju z Inštitutom za novejšo zgodovino, Centrom za jezikovne vire in tehnologije Univerze v Ljubljani (CJVT), ter raziskovalni infrastrukturi CLARIN.SI in DARIAH-SI. Konferenca, ki ima že več kot dvajsetletno tradicijo, je postala pomembna vez med področjem jezikovnih tehnologij in digitalno humanistiko ter je tudi letos predstavljala multidisciplinarni dogodek od razširitve programa konference na področje digitalne humanistike leta 2016.

Poleg osrednjega dela konference sta v sredo, 14. septembra, v okviru konference JTDH 2022 na Inštitutu za novejšo zgodovino potekala tudi predkonferenčna seminarja. Prvo delavnico o tematskem modeliranju parlamentarnih razprav je izvedla Ajda Pretnar Žagar, drugi seminar o raziskovalni infrastrukturi CLARIN.SI pa sta prevzela Kristina Pahor de Maiti in Jakob Lenardič.

Letošnja konferenca se je odvijala na Fakulteti za družbene vede Univerze v Ljubljani. V dveh dneh so svoje prispevke predstavili vabljeni predavatelji in avtorji sprejetih prispevkov, ker pa je bila konferenca mednarodna, je bil program razdeljen na sekcije v slovenskem in angleškem jeziku. Zvrstile so se slovenska in angleška študentska sekcija, dve slovenski in tri angleške redne sekcije ter angleška in slovenska poster sekcija, tako za redne kot za študentske prispevke.

Prvi dan se je konferenca pričela z uradno otvoritvijo, kmalu za tem pa sta na študentski sekciji svoje prispevke predstavila David Bordon, ki je govoril o tem, kako ljudje razumemo jezik sodobnih strojnih prevajalnikov, ter Špela Antloga s prispevkom *Korpusni pristopi za identifikacijo metafore in metonimije: primer metonimije v korpusu g-KOMET*. Sledilo je predavanje vabljenega predavatelja Eetuja Mäkeläja, ki je predstavil, kako najbolje oblikovati računalniške sisteme za podporo humanistiki in družboslovju v namen znanstvenih raziskav. Po vabljenem predavanju sta potekali vzporedni sekciji v slovenskem in angleškem jeziku. Na slovenski sekciji je bilo predstavljenih šest prispevkov. O metodah, vsebini in kategorizaciji besedil v spremljevalnem korpusu Trendi so govorili Iztok Kosem, Jaka Čibej, Kaja Dobrovoljc in Nikola

Ljubešić. Eva Pori, Jaka Čibej, Tina Munda, Luka Terčon in Špela Arhar Holdt so se predstavili s prispevkom o lematizaciji in oblikoskladenjskem označevanju korpusa SentiCoref, o nadgradnji smernic, učnih podatkov in razčlenjevalnega modela na primeru Universal Dependencies za slovenščino pa so govorili Kaja Dobrovoljc, Luka Terčon in Nikola Ljubešić. Na sekciji so sodelovali tudi Darinka Verdonik, Andreja Bizjak, Andrej Žgank in Simon Dobrišek s prispevkom o metapodatkih pri posnetkih in govorcih v govornih virih na primeru baze Artur. S prispevkom *Primerjava načinov razcepljanja besed v strojnem prevajanju slovenščina-angleščina* sta se predstavila Gregor Donaj in Mirjam Sepesy Maučec, sekcija pa se je zaključila s predstavitvijo raziskovalne infrastrukture CLARIN.SI avtorjev Tomaža Erjavca, Kaje Dobrovoljc, Darje Fišer, Jana Jone Javorška, Simona Kreka, Taje Kuzman, Cypriana Laskowskega, Nikole Ljubešića in Katje Meden.

Vzporedno sekcijo v angleškem jeziku sta otvorila Jakob Lenardič in Kristina Pahor de Maiti s prispevkom *Slovenian Epistemic and Deontic Modals in Socially Unacceptable Discourse Online*. Jure Skubic in Darja Fišer sta predstavila prispevek na temo zgodovinopisja, govorila sta o metodah tekstualnih analiz in primerjala skupne značilnosti ter razlike v pristopih, uveljavljenih v zgodovinopisju, s pristopi, ki temeljijo na korpusu. To naj bi omogočalo boljše razumevanje zgodovinske analize parlamentarnega diskurza ter poudarilo pomen projekta ParlaMint in integracije parlamentarnih korpusov v zgodovinske raziskave. O parlamentarnem diskurzu sta spregovorila tudi Nikola Ljubešić in Peter Rupnik. Pri angleški sekciji so s svojim prispevkom o obogatitvi dokumentov kot orodju za avtomatizirano kodiranje intervjujev sodelovali tudi Ajda Pretnar Žagar, Nikola Đukić in Rajko Muršič, sekcijo pa so zaključile Marta Petrak, Mia Uremović in Bogdanka Pavelin Lešić s prispevkom *Fine-grained human evaluation of NMT applied to literary text: case study of a French-to-Croatian translation*.

Po rednih sekcijah je potekala predstavitev plakatov ob kavi. Predstavljene so bile štiri vsebine na temo jezikovnih tehnologij in digitalne humanistike. Jasna Cindrič, Lara Kuhelj, Sara Sever, Živa Simonišek in Miha Šemen so tako govorili o zbiranju podatkov in definicijskih opombah za ekstrakcijo semantičnih relacij, Branko Žitko, Lucija Bročić, Angelina Gašpar, Ani Grubišić, Daniel Vasić in Ines Šarić-Grgić pa so predstavili prispevek z naslovom *Automatic Predicate Sense Disambiguation Using Syntactic and Semantic Features*. Katja Meden je govorila o analiziranju parlamentarnih razprav s pristopi uporabe leksikonov, Petra Matović in Katarina Radić pa sta na sekciji predstavili poster na temo digitalne filologije in poučevanja klasičnih jezikov na Hrvaškem.

Tudi zadnja sekcija prvega dneva konference je bila razdeljena na dva dela: slovenskega in angleškega. Slovenska sekcija, kjer je bilo predstavljenih šest prispevkov, se je začela s predstavitvijo prispevka o sovražnem govoru in grobem besedišču v odzivnem Slovarju sopomenk sodobne slovenščine, ki so ga pripravili Špela Arhar Holdt, Polona Gantar, Iztok Kosema, Eva Pori, Nataša Logar Berginc, Vojko Gorjanc in Simon Krek. Na temo slovarjev je bil predstavljen še prispevek z naslovom *Raba kolokacijskega slovarja sodobne slovenščine pri prevajanju kolokacij* avtorjev Martina Antona Grada in Nataše Hirci. O gradnji Korpusa študentskih besedil KOŠ sta na sekciji

govorili Tadeja Rozman in Špela Arhar Holdt; kako v praksi uporabiti Europeanin podatkovni model (EDM), pa sta na primeru digitalizacije Skuškovske zbirke iz Slovenskega etnografskega muzeja v okviru projekta PAGODE-Europeana China predstavili Maja Veselič in Dunja Zorman. Matija Marolt, Mark Žakelj, Alenka Kavčič in Matevž Pesek so predstavili proces poravnave zvočnih posnetkov s transkripcijami narečnega govora in petja, sekcija pa se je zaključila s predstavitvijo zadnjih napredkov pri samodejni slovenski grafemsko-fonemski pretvorbi, ki so jo pripravili Janez Križaj, Simon Dobrišek, Aleš Mihelič in Jerneja Žganec Gros.

Na vzporedni angleški sekciji je bilo prav tako predstavljenih šest prispevkov. Thi Hong Hanh Tran, Matej Martinc, Andraz Repar, Antoine Doucet in Senja Pollak so v prispevku *A Transformer-based Sequence-labeling Approach to the Slovenian Cross-domain Automatic Term Extraction* govorili o pristopu k označevanju zaporedja. Michal Mochtak, Peter Rupnik in Nikola Ljubešić so predstavili nabor podatkov ParlaSent-BCS o parlamentarnih razpravah, o avtomatični analizi besedila pri ocenjevanju jezika na primeru spletne aplikacije MultiDis pa so govorili Sara Košutar, Dario Karl, Matea Kramarić in Gordana Hržica. Primerjalno študijo o različnih sistemih za pridobivanje ključnih besed so predstavili Boshko Koloski, Senja Pollak in Matej Martinc, na zadnji predstavitvi prvega dne konference pa sta Andrejka Žejn in Mojca Šorli govorili o anotacijah entitet na primeru korpusa May68.

Drugi dan konference se je začel s študentsko sekcijo, ki je potekala v angleškem jeziku. Sekcijo sta otvorili Ruzica Farmakovski in Natalija Tomic s prispevkom na temo srbske in hrvaške spletne enciklopedije – Wikipedie. Meta Jazbinšek, Teja Hadalin, Sara Sever, Erika Stanković in Eva Boneš so predstavile nevronske prevajalske modele, specializirane za prevajanje angleških govorov TED (angl. TED Talks) v slovenščino, o pripravi korpusa in sistema odgovorov na vprašanja za slovenščino pa so govorili Uroš Šmajdek, Maj Zirkelbach, Matjaž Zupanič in Meta Jazbinšek. Zadnji prispevek na študentski sekciji je predstavil Tvrto Balič, ki je govoril o CCRU (Cybernetic Culture Research Unit) kot poskusu delovanja filozofije v digitalnem svetu.

Sledilo je predavanje vabljenega predavatelja Benoita Sagota z naslovom *Large-scale language models: challenges and perspective*. Predavanje je potekalo na temo razvoja večjezičnega korpusa OSCAR in znanja, pridobljenega med izpopolnjevanjem francoskega jezikovnega modela CamemBERT, prvega obsežnega enojezičnega modela za jezik, ki ni angleščina. Po vabljenem predavanju je potekala redna sekcija v angleškem jeziku s petimi prispevki. Taja Kuzman, Nikola Ljubešić in Senja Pollak so predstavili prispevek z naslovom *Assessing Comparability of Genre Datasets via Cross-Lingual and Cross-Dataset Experiments*, o vrednotenju strojnih prevodov sta govorila Špela Vintar in Andraz Repar. Aleksandar Petrovski je predstavil dvojezični angleško-ukrajinski leksikon poimenovanih entitet, izvlečen iz Wikipedije, o populističnem in nepopulističnem diskurzu v slovenskem parlamentu od leta 1992 do leta 2018 pa so govorili Darja Fišer, Tjaša Konovšek in Andrej Pančur. Sekcija je zaključila Petra Bago s predstavitvijo prispevka *Progress of the RETROGRAM Project: Developing a TEI-like Model for Pre-standard Croatian Grammars*.

Sledila je še zadnja sekcija predstavitev – predstavitev plakatov. Predstavili so se Tina Mozetič, Miha Sever, Martin Justin in Jasmina Pegan s temo evalvacijske kategorizacije strojno izluščenih protipomenskih parov; Nina Sangawa Hmeljak, Anna Sangawa Hmeljak in Jan Hrastnik so predstavili aplikacijo Ilukana za učenje japonskih zlogovnih pisav hiragana in katakana s pomočjo asociacij; o šahovski terminološki bazi so govorili Vili Grdič, Kaja Perme, Lea Turšič in Alja Križanec. Lucija Gril, Simon Dobrišek in Andrej Žgank so predstavili plakat z naslovom *Akustično modeliranje z različnimi osnovnimi enotami za avtomatsko razpoznavanje slovenskega govora*, Saša Babič in Tomaž Erjavec pa sta predstavila izdelavo in analizo digitalizirane zbirke paremioloških enot. Na sekciji je sodelovala tudi Magdalena Gapsa, ki je govorila o ocenjevanju uporabniško dodanih sopomenk v Slovarju sopomenk sodobne slovenščine.

Sledila sta podelitev nagrade najboljšemu študentskemu prispevku, ki jo je prejel David Bordon za svoj prispevek *Govoriš nevronske? Kako ljudje razumemo jezik sodobnih strojnih prevajalnikov*, in zaključek konference. Po uradnem zaključku je sledila še posebna sekcija, kjer so bili predstavljeni dosedanja rezultati projekta *Razvoj slovenščine v digitalnem okolju*, konferenca JTDH 2022 pa se je uspešno zaključila z rednim letnim občnim zborom Slovenskega društva za jezikovne tehnologije.

Ana Cvek

Žarko Lazarevič, Delo in zemlja.

Ljubljana: Inštitut za novejšo zgodovino, 2022, 255 str.



Monografija Žarka Lazareviča *Delo in zemlja*, ki je izšla kot 47. delo v zbirki *Razpoznavanja*, je plod avtorjevega dolgoletnega srečevanja s kmečko problematiko, vse od priprave doktorske disertacije o kmečkih dolgovih pred mnogimi leti. To srečevanje je bilo včasih bolj, včasih manj intenzivno, a popolnoma iz raziskovalnega fokusa kmečka zgodovina avtorju nikoli ni ušla. Kot uvodoma pojasni tudi Lazarevič sam, so se njegova razmišljanja z leti kalila preko sodelovanja v različnih domačih in mednarodnih projektih, pa tudi v številnih strokovnih (in hkrati prijateljskih) pogovorih in debatah ter izmenjavah pogledov in mnenj. Monografija je tako rezultat dolgoletnega znanstvenega zorenja in zato tehtno in zrelo, pravzaprav življenjsko delo s področja kmečke in kmetijske zgodovine. Kot

sam poudari, je bila v času, ko se je tematike v doktorski disertaciji začel lotevati sam, to obrobna tema, raziskave pa so se je dotikale zgolj z zornega kota politične zgodovi-