

SLOVENŠČINA 2.0: LANGUAGE TECHNOLOGIES AND DIGITAL HUMANITIES

Darja FIŠER

Faculty of Arts, University of Ljubljana; Institute of Contemporary History;
Jožef Stefan Institute

Tomaž ERJAVEC

Jožef Stefan Institute

Ajda PRETNAR

Institute of Contemporary History

Fišer, D., Erjavec, T., Pretnar, A. (2021): Slovenščina 2.0: Language technologies and digital humanities. Slovenščina 2.0, 9(1): i–vi.

DOI: <https://doi.org/10.4312/slo2.0.2021.1.i-vi>

The current special issue of the journal *Slovenščina 2.0* revisits a topic that has been one of the major focal points of its editorial tradition from the start. In fact, an entire issue was devoted to Language Technologies already in the first year of the journal's existence. With this collection of papers, which arrives nearly a decade later, we take stock of the current state of affairs in the field of development of resources, tools and methods for analyzing written, spoken and multimodal communication as well as their application in Digital Humanities, which has recently become a growing area of research in Slovenia.

The special issue presents eight extended papers from Slovenian as well as international authors that were originally presented at the 2020 *Language technologies and digital humanities conference* as well as a short student paper. They comprise work in language and speech technologies, language resources, digital linguistics, and digital humanities for Slovenian as well as several other languages. The special issue was reviewed by: Špela Arhar, Marko Bajec, Václav Cvrček, Simon Dobrišek, Helena Dobrovoljc, Polona Gantar, Vojko Gorjanc, Jurij Hadalin, Mateja Jemec Tomazin, Iztok Kosem, Cvetana Krstev, Nikola Ljubešić, Nataša Logar, Maja Miličević Petrović, Igor Mozetič,

Tanja Samardžić, Miha Seručnik, Mojca Stritar Kučuk, Janez Štebe, Simon Šuster, Darinka Verdonik, Špela Vintar, Jerneja Žganec Gros and Slavko Žitnik. The editors of the special issue would like to thank the authors and the reviewers for their dedicated work.

On the topic of language and speech technologies, **Marko Robnik-Šikonja**, **Kristjan Reba** and **Igor Mozetič** use cross-lingual word embeddings to transfer classification models for a Twitter sentiment classifier between 13 languages. **Matej Ulčar**, **Anka Supej**, **Marko Robnik-Šikonja** and **Senja Pollak** evaluate Slovenian and Croatian word embeddings in terms of gender bias using word analogy calculations. **Lucija Gril**, **Mirjam Sepesy Maučec**, **Gregor Donaj** and **Andrej Žgank** present the development of an automatic recognizer of Slovenian speech for the domain of daily news broadcasts using the the UBM BNSI Broadcast News and IETK-TV databases to train the speech recognizer using deep neural networks.

With a focus on language resources and digital linguistics, **Lucia Vlášková** and **Hana Strachoňová** present the challenges and solutions for creating an online dictionary of the Czech sign language. **Dolores Lemmenmeier-Batinić** gives an account of building a corpus of spoken Serbian and discusses current challenges in the processing of spoken data, and the implications of data re-use regarding transcriptions of speech. **Jakob Lenardič** and **Darja Fišer** perform a comparative corpus analysis of modal adverbs in Slovenian academic texts from different disciplines and study levels. **Darinka Verdonik**, **Simona Majhenič**, **Špela Antloga**, **Sandi Majninger**, **Marko Ferme**, **Kaja Dobrovoljc**, **Simona Pulko**, **Mira Krajnc Ivič** and **Natalija Ulčnik** present the development of an e-learning environment for improving writing and communication skills of Slovenian pupils.

From the digital humanities perspective, **Katja Meden** and **Ana Cvek** give an account of a major overhaul of the Historiography Citation Index that will improve the indexing of citations of scientific publications for historiographers. **Rok Mrvič** and **Špela Zupančič** survey and demonstrate the functionality of Slovenian online dialectological resources and tools.

Compared to the first special issue on Language Technologies published in this journal in 2013 where the focus of research was on the development of

basic resources and tools for Slovenian and related languages, we can observe a shift to the implementation of state-of-the-art machine learning methods, multilingual approaches, critical evaluation of technologies, and development of services for the end user. This, along with a much longer list of co-authors who come from many more institutions and countries, and work on many more languages than in the original special issue, suggests that the field has advanced significantly in the past decade and will continue to thrive, so we are already looking forward to the next special issue with a similar focus in the future.

SLOVENŠČINA 2.0: JEZIKOVNE TEHNOLOGIJE IN DIGITALNA HUMANISTIKA

Pričujoča posebna številka revije Slovenščina 2.0 se vrača k temi, ki je bila ena od osrednjih uredniških izhodišč vse od nastanka revije, saj je bila jezikovnim tehnologijam posvečena že njena prva tematska številka. Skoraj desetletje kasneje z naborom prispevkov predstavimo trenutno stanje razvoja virov, orodij in metod za analizo pisne, govorne in multimodalne komunikacije, hkrati pa se posvetimo tudi njihovi praktični uporabi v digitalni humanistiki, ki postaja vse bolj razširjeno raziskovalno področje tudi v Sloveniji.

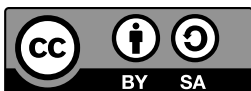
Posebna številka predstavlja osem razširjenih prispevkov slovenskih in tujih avtorjev, ki so bili izvorno predstavljeni na konferenci *Jezikovne tehnologije in digitalna humanistika* leta 2020. Prispevki vključujejo raziskave in nadgradnje jezikovnih in govornih tehnologij, jezikovnih virov, digitalnega jezikoslovja ter digitalnohumanistične raziskave tako za slovenščino kot za nekatere druge jezike. Posebno številko so recenzirali Špela Arhar, Marko Bajec, Václav Cvrček, Simon Dobrišek, Helena Dobrovoljc, Polona Gantar, Vojko Gorjanc, Jurij Hadalin, Mateja Jemec Tomazin, Iztok Kosem, Cvetana Krstev, Nikola Ljubešić, Nataša Logar, Maja Miličević Petrović, Igor Mozetič, Tanja Samarždić, Miha Seručnik, Mojca Stritar Kučuk, Janez Štebe, Simon Šuster, Darinka Verdonik, Špela Vintar, Jerneja Žganec Gros in Slavko Žitnik. Uredniki posebne številke se iskreno zahvaljujemo avtorjem in recenzentom za njihovo predano delo.

Na področju jezikovnih in govornih tehnologij **Marko Robnik-Šikonja**, **Kristjan Reba** in **Igor Mozetič** predstavijo uporabo medjezikovnih vložitev besed za prenos napovednih modelov strojnega učenja za klasifikacijo sentimenta na Twitterju med trinajstimi jeziki. **Matej Ulčar**, **Anka Supej**, **Marko Robnik-Šikonja** in **Senja Pollak** poročajo o evalvaciji spolne pristranskosti slovenskih in hrvaških besednih vložitev s pomočjo besednih analogij. **Lucija Gril**, **Mirjam Sepesy Maučec**, **Gregor Donaj** in **Andrej Žgank** pa predstavijo razvoj avtomatskega razpoznavalnika slovenskega govora za dnevna poročila, pri čemer razpoznavalnik govora z globokimi nevronskimi mrežami naučijo na podatkih UBM BNSI Broadcast News in IETK-TV.

Na področju jezikovnih virov in digitalnega jezikoslovja **Lucia Vlášková** in **Hana Strachoňová** obravnavata izzive in rešitve pri snovanju spletnega slovarja češkega znakovnega jezika. **Dolores Lemmenmeier-Batinić** opiše postopek oblikovanja korpusa govornjene srbščine in obravnava posledice ponovne rabe transkripcij govora. **Jakob Lenardič** in **Darja Fišer** izvedeta primerjalno analizo rabe modalnih prislovov v slovenskih akademskih besedilih med različnimi področji in ravnmi izobrazbe. **Darinka Verdonik**, **Simona Majhenič**, **Špela Antloga**, **Sandi Majninger**, **Marko Ferme**, **Kaja Dobrovoljc**, **Simona Pulko**, **Mira Krajnc Ivič** in **Natalija Ulčnik** predstavijo razvoj učnega spletnega okolja za razvoj pisnih in govornih veščin slovenskih učencev.

Z vidika digitalne humanistike **Katja Meden** in **Ana Cvek** opišeta pomembno prenovitev Zgodovinarskega indeksa citiranosti, ki bo zgodovinarjem v pomoč pri indeksiranju znanstvenih objav. **Rok Mrvič** in **Špela Zupančič** pregledata in prikažeta uporabnost spletnih orodij in virov za slovenska narečja.

V primerjavi s prvo tematsko številko na temo jezikovnih tehnologij iz leta 2013, kjer je bil poudarek na razvoju osnovnih virov in orodij za slovenščino in sorodne jezike, je v tokratni izdaji opazen premik k uvajanju naprednih tehnik in metod strojnega učenja, večjezikovnim pristopom, kritičnemu ocenjevanju obstoječih tehnologij ter razvoju storitev za končnega uporabnika. Ta premik, hkrati z daljšim seznamom avtorjev z različnih institucij in držav, ki se ukvarjajo z veliko širšim naborom jezikov kot v prvi številki, nakazuje na izjemen razmah področja v zadnjem desetletju. Digitalna humanistika in jezikoslovne tehnologije se bodo očitno uspešno razvijale še naprej, zato se že veselimo prihodnje številke na podobno temo.



To delo je ponujeno pod licenco Creative Commons: Priznanje avtorstva-Deljenje pod enakimi pogoji 4.0 Mednarodna. / This work is licensed under the Creative Commons Attribution-Share-Alike 4.0 International.

<https://creativecommons.org/licenses/by-sa/4.0/>