

Jezički raskol u ‘stvarnom’ i digitalnom okruženju: slučaj srpskohrvatskog

Par Marija RUNIĆ

Publication en ligne le 26 avril 2025

Mots-Clés

srpskohrvatski, jezičke politike, digitalno okruženje, jezički sukob, ideološke pristrasnosti.

Table des matières

Introduction

Jezički raskol u ‘stvarnom’ okruženju i slučaj Bosne i Hercegovine

Digitalna vitalnost i jednakost: jezici bivšeg srpskohrvatskog prostora

Jezička situacija u Bosni i Hercegovini i digitalno okruženje

Zaključak

Texte intégral

This paper explores the consequences of the fragmentation of the once-unified Serbo-Croatian language into separate languages, each with distinct language policies, by analyzing how these divisions manifest in digital spaces and language technologies. It

investigates the concept of digital inequality, particularly the uneven availability and quality of linguistic data across these languages. A key focus is the ongoing language conflict in Bosnia and Herzegovina (BiH), where competing language policies—especially in education—reflect struggles over linguistic dominance and identity. This conflict is most visible in disputes over the language's name and the use of scripts (Latin vs. Cyrillic). In the digital sphere, these tensions result in inadequate representation of linguistic varieties and scripts, ultimately diminishing the performance of language technologies. By examining these dynamics, the paper demonstrates how the digital environment mirrors broader debates about linguistic boundaries following the dissolution of Serbo-Croatian. It also highlights persistent ideological biases embedded in linguistic data.

Introduction

Raspadom socijalističke Jugoslavije 90-tih godina prošlog stoljeća, vijek i po dugi procesi zajedničkih promišljanja, standardizacije, planiranja i politika policentričnog jezika kojim se govori(lo) na prostoru današnje Hrvatske, Bosne i Hercegovine (BiH), Srbije i Crne Gore su, čini se, nepovratno zaustavljeni. Iako je još od samih početaka polovinom XIX vijeka dolazilo do otpora kojim se utirao put budućim odvojenim standardizacijama, koje su se potom odvijale u različitom obimu i različitim tempom unutar pojedinačnih jezika nasljednika (Alexander 2013), političko-društveni kontekst jedinstvene države je ipak donekle omogućavao da ovi procesi budu uslovljeni međusobnom saradnjom. Sa promjenom društveno-političkog okruženja stvoreni su preduslovi da se i procesi standardizacije odigravaju potpuno zasebno, iako se to ispostavilo nemogućim u slučaju BiH i Crne Gore (Šipka 1999; Halilović 2014; Glušica 2020).

Razvoj jezičkih tehnologija i sve veće prisustvo jezika nastalih na temelju nekadašnjeg srpskohrvatskog/hrvatskosrpskog u digitalnom okruženju nisu uspjeli da prevaziđu stare izazove i neriješena pitanja. I pored toga što su razmjena i primjena različitih rješenja u razvoju resursa za pojedinačne jezike sastavni dio njihovog razvoja (up. Clarin.SI), među otvorenim, a naslijeđenim problemima, i dalje ostaju nejednaka zastupljenost različitih jezika, odnosno varijeteta unutar jednog jezika, definisanje i razumijevanje jezičkih granica uslijed neriješenih etničkih odnosa, ponajviše u BiH, kao i održavanje određenih ideoloških pristrasnosti u široko dostupnim tehnologijama.

Namjera nam je da u radu predstavimo neke od izazova koji nosi postojanje zasebnih jezika u digitalnom okruženju, posebno u vezi sa BiH. Rad počinje sa kratkim osvrtom na ono što je jezički raskol donio u periodu nakon 'raspada' srpskohrvatskog, sa posebnim pogledom na aktuelnu bosanskohercegovačku situaciju. Izdvojili smo samo nekoliko najvažnijih

karakteristika s obzirom na iscrpnu literaturu o ovim pitanjima, i u domaćoj, i u međunarodnoj sociolingvistici (među najuticajnijima radovima su Bugarski 1994, 1997, 2002, Bugarski & Hawkesworth 2004, Mønnesland 2005, Greenberg 2008, Vajzović 2008; Kordić 2010, Alexander 2013, Požgaj Hadži 2013; Langston & Peti Stantić 2014). U središnjem dijelu rada predstavljen je odnos pojedinačnih jezika u okviru pojma 'digitalne jednakosti' i raspoloživosti jezičkih resursa, kao i odnosa ćirilice i latinice. Treći dio rada detaljnije opisuje na koji se način trenutna jezička situacija u BiH odražava na prisustvo jezika kojim govore stanovnici BiH u digitalnoj stvarnosti.

Jezički raskol u 'stvarnom' okruženju i slučaj Bosne i Hercegovine

U svakoj od novonastalih država nakon raspada Jugoslavije jezička pitanja zauzimala su središnje mjesto u razumijevanju nacionalnog pitanja, te bila i ostala predmetom obimnih zakonodavnih intervencija. Neizostavan dio konstituisanja posebnih službenih nacionalnih jezika – bosanskog, hrvatskog, crnogorskog i srpskog – činio je i odnos prema nasljeđu nekadašnjeg zajedničkog jezika. Na ideološkom planu, ovaj odnos je odredio standardizaciju hrvatskog u pravcu udaljavanja od srpskog, odnosno od istočne varijante nekadašnjeg zajedničkog jezika, te u pravcu jezičkog purizma (Kordić 2010, Kapović 2011, Starčević, Kapović & Sarić 2018). Kod nosilaca standardizacije srpskog jezika vremenom je dolazilo do sve većeg isticanja ćiriličnog pisma spram rasprostranjene digrafije, koja je kulminirala donošenjem Zakona o upotrebi srpskog jezika u javnom životu i zaštiti i očuvanju ćiriličkog pisma u Srbiji (Sl. glasnik RS", br. 89/2021) i Zakona o zaštiti, očuvanju i upotrebi jezika srpskog naroda i ćiriličkog pisma u entitetu Republika Srpska u BiH (Službeni glasnik Republike Srpske broj: 63/22). U standardizaciji bosanskog i crnogorskog odvijali su se, pak, uporedivi procesi koji su imali za cilj da promovišu u standard posebnosti sopstvenih varijanti u odnosu na sve ostale varijante srpskohrvatskog, prvenstveno kroz isticanje posebnih leksičkih i fonetskih/fonoloških obilježja (npr. Jahić 1999, Čirgić 2007, 2009).

Svi ovi procesi nisu prolazili bez kritika i žustrih debata, ne samo između jezičkih stručnjaka, već su uključivali i širu javnost (Bugarski 2018; Bugarski 2021). Zamjerke su se uglavnom sastojale iz ukazivanja na pogubnost nacionalno usmjerenih standardizacija i preskriptivizma i praćene su zahtjevima za većom tolerancijom u jezičkim pitanjima. I u okviru samih nacionalnih procesa standardizacije dolazilo je promjene stava prema zajedničkom nasljeđu. Najbolji primjer je izrada dvije varijante pravopisa bosanskog jezika istog autora u razmaku od 20-tak godina (Halilović 1996, Halilović 2018). U prvoj verziji

pravopisa, prema riječima autora, insistiralo se na smanjenu višestrukih rješenja i crpljenju primjera iz „bošnjačke jezičke baštine“ (Halilović 1996: 7). Druga verzija pravopisa, insistira pak na čuvanju dvostrukosti kao temelja bosanskohercegovačke jezičke stvarnosti (Halilović 2018).

Iako je postojanje jednog jezičkog i kulturnog prostora na temelju nekadašnjeg srpskohrvatskog neupitno, zvanične jezičke politike, uz nekoliko rijetkih izuzetaka, i dalje se zasnivaju na principu isključivosti i privilegovanja sopstvenih varijanti na uštrb drugih, što u stvarnom okruženju dovodi do jezičke, a i drugih oblika diskriminacije (Vajzović 2008, OSCE 2018). P zasebnih jezika i razumijevanja šta su njihove granice, koliko god teško odgovoriti u ovim okolnostima (Greenberg 2008), ima važne posljedice u stvarnom životu. One se tiču obrazovnih jezičkih politika, definisanja jezičkih manjina i jezičkih prava, upotrebe pisma/pisama u zvaničnoj, ali i neformalnoj komunikaciji, nastave stranog jezika/jezika u zemlji i inostranstvu, kao i jezičkih politika u izdavačkim i medijskim kućama, jezičke industrije, te konačno i samih naučnih paradigmi, odnosno pristupa u proćavanju ovih jezika i akademske karijere istraživača unutar pojedinačnih zemalja (up. Runić & Božović, u štampi). U kojoj mjeri sva ova pitanja utiču nepovoljno na samu vitalnost navedenih jezika, na njihovu atraktivnost za strane učenike i da li bi situacija bila bolja ukoliko bi postojao samo jedan jezik ostaje pitanje koje u današnjoj postavci nema jednoznačan odgovor budući na značaj nacionalnog jezika za razumijevanje nacionalnog identiteta (up. Buden 2017).

Kada se pak posmatraju sami jezički podaci, odnosno arealna distribucija jezičkih crta koja se vezuje za pojedinačne standardne jezike u novonastalim okolnostima, jedinu studiju koja se bavi razlikama između pisanih standarda bosanskog, hrvatskog, crnogorskog i srpskog jezika sproveli su Ljubešić, Milićević & Samardžić (2018) na osnovu podataka sa Twittera (današnjeg X-a). Prema rezultatima do kojih su došli, najmanja jezička udaljenost uočava se između BiH i Crne Gore, zatim između BiH i Srbije, dok slijede Crna Gora i Srbija. Srbija i Hrvatska su se pokazale kao jezički najudaljenije, što se donekle oslanja na tradicionalno priznate dvije osnovne varijante, istočnu (srpski) i zapadnu (hrvatski), koji su tokom zajedničke jezičke politike regulisani u dva glavna centra za jezičku politiku, u Srbiji i Hrvatskoj. Od svih zemalja, Hrvatska se ističe kao jezički najposebnija u odnosu na ostale zemlje, dok je Srbija najkonzistentnija u raspodjeli analiziranih jezičkih odlika, čineći najhomogeniji jezički prostor. Prema raspodjeli analiziranih jezičkih odlika od svih zemalja najraznolikijom se ispostavlja BiH.

Ovakvi nalazi u vezi sa BiH direktna su potvrda onog što se prema Davies & Dubinsky (2018: 310) karakteriše kao takmičenje za jezičku dominaciju, u kome „dvije [tri] grupe imaju prevlast u određenim regionima zemlje, boreći se za dominaciju ili nezavisnost, pri čemu

prava na jezik igraju ulogu u artikulaciji te borbe“. Naime, pitanje raspada Jugoslavije i jezičkog nasljeđa u BiH iznjedrio je tri službena jezika, bosanski, hrvatski i srpski, shodno težnjama njenih konstitutivnih naroda da ostvare pravo na sopstveni jezik (Šipka 1999). Takav pristup sa sobom je podrazumijevao niz teško rješivih problema, usko povezanih i sa administrativno-političkim ustrojem zemlje. Podijeljenost zemlje na dva entiteta shodno etničkom principu – Federaciju BiH, sa većinskim bošnjačkim i hrvatskim stanovništvom, razučene administrativno-politički organizacije od 10 kantona, i Republiku Srpsku, sa većinskim srpskim stanovništvom i izrazito centralizovanim administrativno-političkim sistemom, te Brčko Distrikt sa nezavisnom upravom i multietničkim sastavom – doveo je do mogućnosti da se jezičke politike kreiraju nezavisno, da često budu u suprotnosti jedna sa drugom i da dovode do stvarnih posljedica po živote njenih stanovnika, kroz različite oblike diskriminacije i jezičkog profilisanja. Za Srbe i Hrvate centri jezičkih politika izmješteni su van zemlje, u Beograd i Zagreb, dok je Sarajevo, i pored nastojanja grupe lingvista da ostane referentan centar za jezičku politiku na teritoriji cijele BiH, što jeste bio slučaj od 70-ih godina prošlog vijeka do početka sukoba, *de facto* ostao centar jezičkih politika samo za Bošnjake. Jezičke institucije, odsjeci za nacionalne jezike, izdavačka djelatnost rijetko kada uključuju lingviste iz cijele zemlje, dijelom i zbog toga što ovakvi oblici saradnje na jezičkim pitanjima može ugroziti karijeru i pristup resursima lingvistima koji na saradnju pristanu (Runić & Božović, u štampi).

Udaljavanje na planu osmišljavanja i sprovođenja jezičkih politika se u praksi nastavlja i pored početnih napora za harmonizacijom i međusobnim priznavanjem, proisteklih prije svega kroz inicijative međunarodne zajednice za pomirenjem, tokom prve decenije nakon završetka oružanog sukoba. Zbog svoje izrazito fragmentirane političko-administrativne strukture, BiH ostaje podložna međudržavnom nacionalizmu. Najnoviji potez u ovom smjeru je donošenje Deklaracije Svesrpskog sabora 2024. godine, u kojoj pitanja jezika zauzimaju važno mjesto [\[1\]](#). Ističe se posebno “ujedinjujući značaj srpskog jezika i ćirilskog pisma”, pravo na “istu srpsku standardizaciju/i ćirlicu”, predlažu mjere za “organizovanje i sprovođenje jedinstvenih, posebno standardizovanih programa učenja srpskog jezika, ćirilskog pisma”.

Sukob na planu jezika, čiji korijeni sežu do prvih pokušaja definisanja i sprovođenja jezičkih politika u vrijeme austro-ugarske okupacije (Šator 2005), i dalje se prvenstveno ispoljava kroz neslaganje oko imenovanja jezika, upotrebe pisma, prije svega ćirilice, te kroz obrazovne jezičke politike, kao najistaknutijeg domena borbe za jezičkom dominacijom. U imenovanju problema najviše nesuglasica izaziva naziv „bosanski“, koga osporavaju, pa čak i zabranjuju, veći dio srpskih i hrvatskih lingvista, zbog navodnog straha da bi ovo ime moglo zamijeniti srpski i hrvatski jezik u BiH i koristiti se kao jedinstven naziv za jezik cijele BiH (up. Odbor za standardizaciju srpskog jezika 1998). Treba istaći kako alternativni naziv

za jezik koga isti lingvisti predlažu, "bošnjački jezik", zapravo nije prihvatljiv za bosnistiku, proglašen je neustavnim još 1997. godine (Halilović 2014), a često ga i sami govornici bosanskog jezika smatraju pogrdnim. Stoga naziv 'bošnjački' opstaje uglavnom u konzervativnom, ali politički uticajnom dijelu srbistike i kroatistike.

Kada se u obzir uzmu sve ove okolnosti, jezička raznolikost koju su Ljubešić *et al.* (2018) uočili u jezičkim podacima nastalim u BiH zapravo su, pored dijalektalnih razlika koje postoje nezavisno od jezičkih politika i uopšte varijacije koja karakteriše BiH jezički prostor (Mønnesland 2005, Halilović 2014), i posljedica razlaza u normativnim trendovima, funkcionalne aproprijacije/odbacivanja ćirilicnog pisma i djelovanja različito usmjerenih jezičkih tijela unutar same zemlje.

Digitalna vitalnost i jednakost: jezici bivšeg srpskohrvatskog prostora

U savremenom svijetu položaj jednog jezika, kroz razumijevanja njegovog značaja i vitalnosti nije moguće razumjeti ako se ne uzme u obzir i prisustvo jezika u *online* prostoru i na digitalnim uređajima (Kornai 2013, Soria *et al.* 2017, Simons *et al.* 2022). Digitalna vitalnost stoga sve više postaje jedna od temeljnih pretpostavki za razumijevanje stvarne vitalnosti jezika. Soria *et al.* (2017: 6) izdvajaju sljedeće indikatore za mjerenje digitalne vitalnosti: digitalne kapacitete (konekciju, digitalnu pismenost, pokrivenost internetom, kodirano pismo, dostupnost jezičkih resursa); digitalno prisustvo i upotrebu (upotrebu u elektronskoj komunikaciji i na društvenim mrežama, dostupnost internetskih medija, Vikipediju); digitalnu efikasnost (dostupne usluge na internetu, lokalizovane društvene mreže, lokalizovane softvere, usluge i alate mašinskog prevođenja, poseban internet domen).

Sve zemlje u kojim se govori nekadašnji srpskohrvatski jezik su u dobrom položaju kada su u pitanju digitalni kapaciteti. [2] S obzirom na veliki broj korisnika društvenih mreža u svim zemljama u kojima se govori nekadašnji srpsko-hrvatski (npr. prema izvještaju Datereportala za 2023, oko 70 posto stanovništva koristi društvene mreže u Srbiji i Hrvatskoj), upotreba svih jezika proizašlih iz srpskohrvatskog na društvenim mrežama je velika, kao i prisustvo internetskih medija, posjeduju Internet domen, itd. Svi jezici sem crnogorskog imaju lokalizovanu verziju društvenih mreža. Među dostupnim jezičkim paketima za Windows postoje i hrvatski i srpski (i na latinici i na ćirilici), ali ne i za bosanski i crnogorski, dok jezički interfejs paket uključuje i ćirilicu verziju srpskog koji se koristi u Bosni i Hercegovini. [3] Kada je u pitanju mašinsko prevođenje, kao najrasprostranjenija

usluga mašinskog prevođenja, barem prije pojave ChatGPT-a, Google Translate (GT) nudi među ponuđenim jezicima sve jezike sem crnogorskog, s tim što srpski prepoznaje kao ekavski izgovor i ćirilčno pismo.

Slučaj Vikipedije je posebno zanimljiv. Prva verzija stranice pokrenuta je na srpskohrvatskom 2002. godine, da bi za njom uslijedile verzija na bosanskom (2002) i srpskom i hrvatskom (2003). Ova verzija sada sadrži preko 450 hiljada unosa (članaka), što je čini drugom najvećom Vikipedijom na južnoslovenskim jezicima i 32. po veličini na svijetu. Sadrži članke na latinici sa konvertorom za preslovljavanje na ćirilicu. [4] Trenutno ne postoji verzija Vikipedije na crnogorskom jeziku. Kada se uporede brojevi članaka na svakoj pojedinačnoj Vikipediji, prednjači broj članaka na srpskoj verziji (skoro 700 hiljada, na 22. mjestu po broju članaka), za kojom idu srpskohrvatska, zatim hrvatska (oko 220 hiljada, na 51. mjestu po brojnosti članaka) i bosanska (oko blizu 94 hiljade, što je stavlja na 76. mjesto). U poređenju sa nekim drugim evropskim jezicima, bosanska Vikipedija po količini sadržaja se može uporediti sa oksitanskom, bjeloruskom i albanskom, hrvatska sa litvanskom i galicijskom, srpskohrvatska sa rumunskom i baskijskom, a srpska sa norveškom i katalanskom. [5]

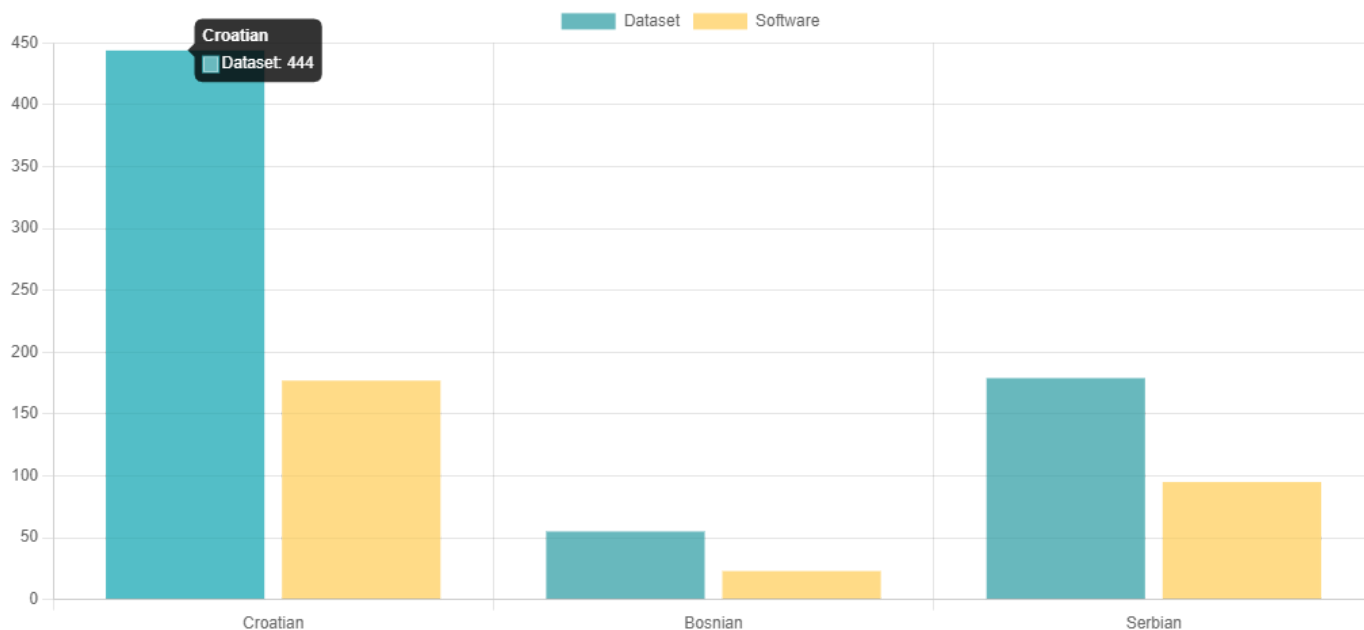
Kvalitet jezičkih podataka je također bitan element u procjeni jezičke vitalnosti. On često utiče na to da li će korisnici uopšte koristiti određenu uslugu na svom jeziku, iako i neki drugi faktori mogu imati uticaja (v. niže). Kvalitet je umnogome određen postojanjem adekvatnih jezičkih resursa, najčešće javno dostupnih, kao što su rječnici, tezaursi, anotirani korpusi, uporedni korpusi, posebno za kvalitet mašinskog prevođenja, gramatički opisi, na kojima se određene jezičke tehnologije treniraju. Prema izvještajima za pojedinačne jezike u posljednjoj deceniji postoje značajni pomaci u razvoju jezičkih resursa (Tadić 2023, Krsteva & Stanković 2023, JePtex [6]), posebno za hrvatski, što je podstaknuto institucionalnom podrškom nakon pridruživanja EU, a hrvatsko predsjedavanje Vijećem EU je dodatno ubrzalo i unaprijedilo položaj hrvatskog u oblasti mašinskog prevođenja. U Hrvatskoj je najviše urađeno na polju razvoja istraživanja u oblasti jezičkih tehnologija, dok jezička industrija u Srbiji pravi pomake na razvoju govornih tehnologija (Krsteva & Stanković 2023: 205), iako je zaključak autorki da jezičke tehnologije i istraživanja na ovom polju za srpski jezik znatno nedostaju kada se uporede sa hrvatskim. Dosta lošija situacija je za bosanski (Bulić *et al.* 2023), koji pored dosta manjeg broja resursa gotovo da nema nikakva istraživanja na ovom planu a državna strategija po pitanju razvoja jezičnih tehnologija za bosanski ne postoji, između ostalog i zato što je pitanje jezika osjetljivo pitanje u BIH (Ćušić 2022: 102). [7]

Nesrazmjer između jezika koji su vitalni u stvarnom svijetu i jezika koji su ugroženi ili im je neophodna podrška da bi opstali daleko je izraženiji kad se razmatra digitalna sfera.

Napredak tehnologije i prirodne obrade jezika zapravo se tiče vrlo malog broja jezika (Biasi, Anastasopoulos & Neubig 2022). Prema podacima preuzetim sa Statiste, za kreiranje sadržaja na internetu najviše se koristi engleski (52.1%), a prvi nakon njega je španski, sa 5.5%. Kada se uporede podaci iz 2012. i 2022. godine, jaz između engleskog i ostalih jezika zapravo raste (Giagkou 2022). Svi evropski jezici, uz izuzetak španskog, njemačkog i francuskog, nemaju dovoljno resursa, tehnologija i podataka, koji bi im osigurali izvjestan opstanak u digitalnoj upotrebi (*ibidem*). Razlozi za zapostavljenost treba tražiti ne samo u nedostatku i nerazvijenosti resursa već i u nedovoljnom zalaganju institucija, donosioca odluka i kreatora jezičkih politika, kao i u ograničenom komercijalnom interesu (Rehms & Way 2023: 6).

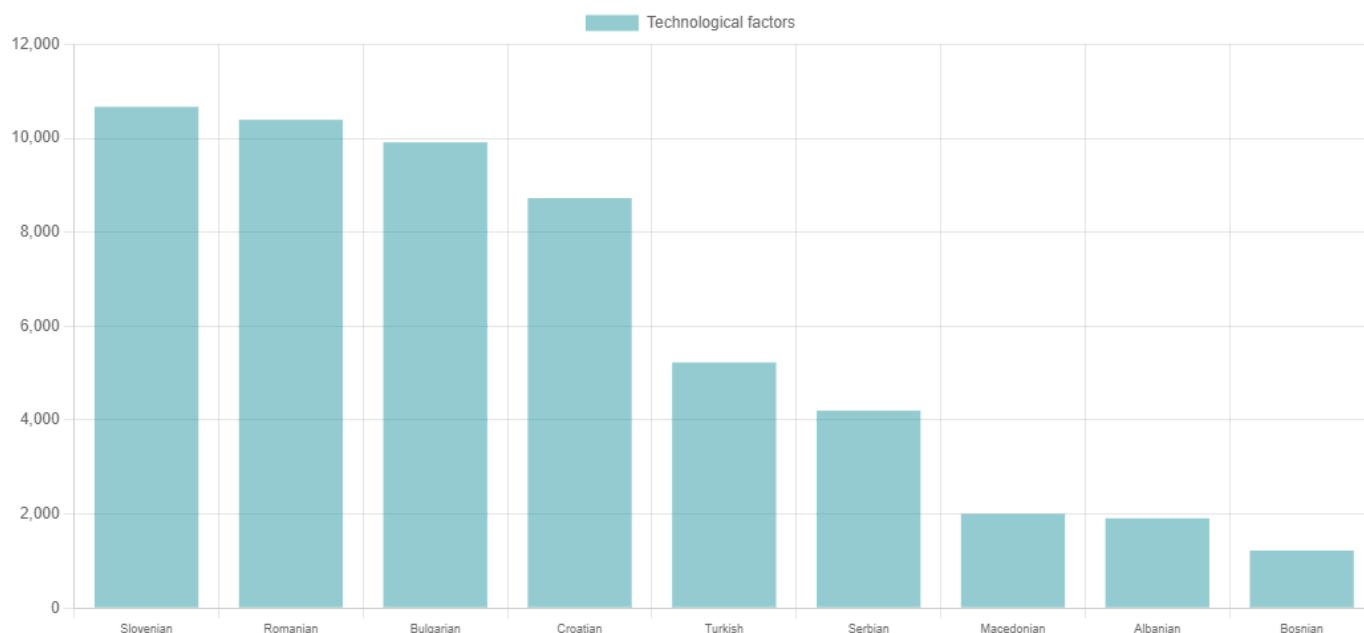
S obzirom na ovakvu situaciju, koja prijeteći da ugrozi višejezičnost kao temeljnu odliku evropskog jezičkog prostora, Evropa je posvećena postizanju digitalne jednakosti za sve jezike, ne samo za zvanične jezike Evropske Unije, već i nezvanične i regionalne/manjinske jezike. [8] U tu svrhu definisan je koncept digitalne jezičke jednakosti, kao „stanje u kojem svi jezici imaju neophodnu tehnološku podršku i kontekst da bi nastavili postojati i ostvarivati napredak kao živi jezici u digitalnom dobu.“ (*ibidem*: 43). Digitalna jezička jednakost je osmišljena kao mjerljiv koncept posredstvom preciznih parametara, a nastala je iz potrebe da se na odgovarajući način pristupi ostvarivanju digitalne jezičke jednakosti u Evropi do 2030 godine. Za mjerenje su ključne dvije grupe činilaca, od kojih su prvi tehnološke, a drugi socioekonomske, odnosno kontekstualne prirode. Prva grupa se tiče vrste, broja i kvaliteta resursa i podataka za mogućnost dalje obrade (kao recimo raznih tipova korpusa i njihove anotacije, usluga i alata za prirodnu obradu jezika), te njihove dostupnosti, odnosno mogućnosti slobodnog pristupa. U drugu grupu činilaca spadaju ekonomska snaga određene jezičke zajednice, digitalno obrazovanje, ulaganje u istraživanje, razvoj jezičke industrije, dostupnost *online* medija, tehnološku infrastrukturu, jezičke politike, pravni okvir i administraciju do stavova govornika prema upotrebi svog jezika u digitalnom svijetu. Na sljedećim dijagramima, preuzetim sa stranice projekta Evropske jezičke jednakosti (<https://european-language-equality.eu/>), moguće je uporediti bosanski, hrvatski i srpski prema rastu broja resursa (Dijagram 1) kao i dostupnim podacima i softverima (Dijagram 2)

Dijagram 1. Rast broja resursa za hrvatski, bosanski i srpski za period 2020-24, preuzeto sa <https://live.european-language-grid.eu/catalogue/dashboard> (Evolution over time)

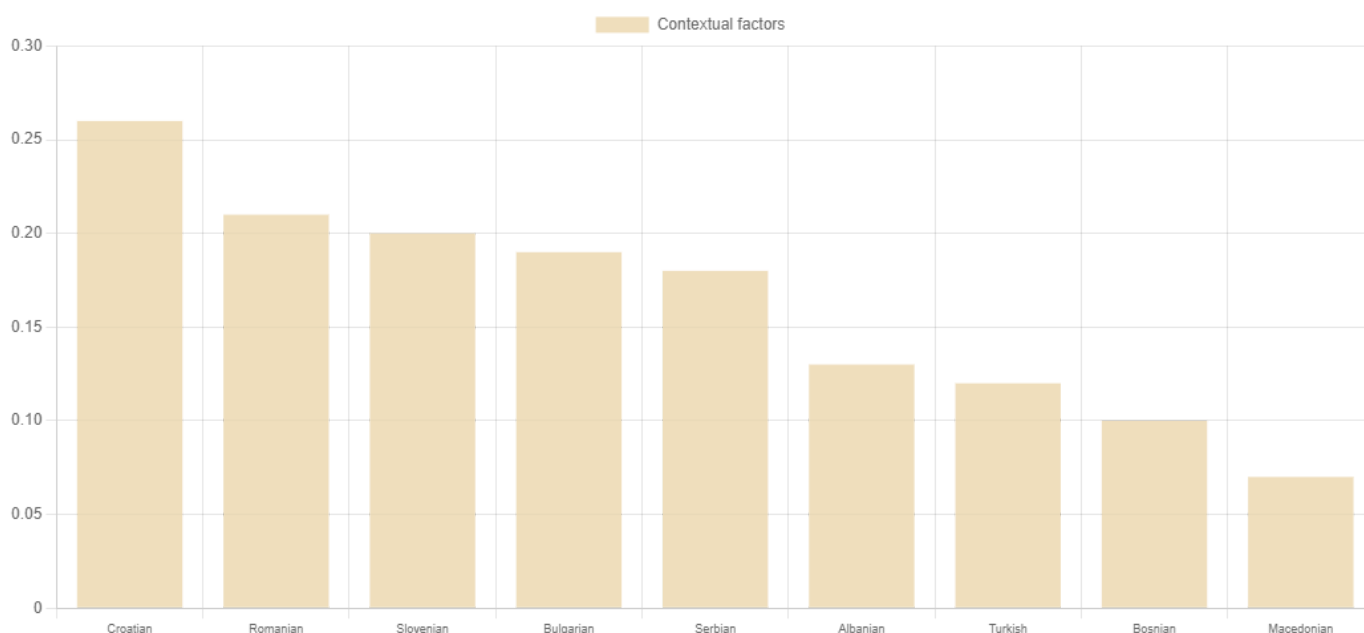


Dijagram 2. Opšti pregled dostupnih baza podataka i resursa za hrvatski, bosanski i srpski jezik za period 2020-24, preuzeto sa <https://live.european-language-grid.eu/catalogue/dashboard> (Cross-language comparison)

Kada se uzmu u obzir tehnološki i kontekstualni faktori digitalne jezičke jednakosti, postoji nesrazmjer između evropskih jezika koji su zvanični jezici EU i onih koji to nisu. Na dijagramima 3 i 4, može se vidjeti kakav je položaj bosanskog, hrvatskog i srpskog u odnosu na druge jezike bivše Jugoslavije, slovenački i makedonski (za crnogorski ne postoje podaci), kao i u odnosu na neke balkanske jezike, poput bugarskog, rumunskog i turskog, u vezi sa tehnološkim (Dijagram 3) i kontekstualnim faktorima (Dijagram 4). Jezici čije su zemlje ranije pristupile Evropskoj Uniji posjeduju veći broj resursa, bez obzira na broj govornika. Kada su pak poređenja između bosanskog, hrvatskog i srpskog u pitanju, po kriterijumu tehnoloških faktora hrvatski je u dvostruko boljoj poziciji od srpskog i višestruko boljoj poziciji od bosanskog, iako su prema kontekstualnim faktorima razlike nešto blaže. [9]



Dijagram 3. Opšti pregled rezultata evropske digitalne jednakosti za tehnološke faktore odabranih evropskih jezika, preuzeto sa <https://live.european-language-grid.eu/catalogue/dashboard> (Technological factors)



Dijagram 4. Opšti pregled rezultata evropske digitalne jednakosti za kontekstualne faktore odabranih evropskih jezika, preuzeto sa <https://live.european-language-grid.eu/catalogue/dashboard> (Contextual factors)

Drugo važno pitanje srpskohrvatskog jezičkog prostora tiče se nesrazmjera između latiničnog i ćiriličnog pisma u digitalnom okruženju. Internet je izrazito latiničnocentričan, iako se čine značajni pomaci da se ovaj nesrazmjer prevaziđe (up. Registar nacionalnog Internet domena Srbije). Postoji veliki nesklad u upotrebi pisma u javnim ustanovama i

javnim medijima u Srbiji i Republici Srpskoj, koji uglavnom posjeduju sajtove na oba pisma ili koriste samo ćirilično pismo, te komercijalnog sektora i privatnih medija, koji uglavnom koriste samo latinicu. Pored toga, postojeći ćirilični domen .cp6 koristi izrazito mali broj komercijalnih korisnika, dok ćirilični .ba domen ne postoji, iako su i ćirilica i latinica zvanična pisma u BiH. Postoji nekoliko razloga za manju upotrebu ćirilice. S jedne strane, prisutna su tehnička ograničenja koja zahtijevaju neznatno veća ulaganja i uloženo vrijeme, kao i ograničenja tržišta, s obzirom da većina govornika nekadašnjeg srpskohrvatskog jezika koristi latinično pismo (RNIDS 2021). S druge strane, govornici su naviknuti da koriste latinično pismo u digitalnom okruženju upravo zbog toga što je ona lakša za upotrebu, barem u ovom trenutku. Većina govornika koja koristi aktivno ćirilicu, piše njom kada zapisuju rukom, dok u digitalnom okruženju, uključujući i privatnu prepisku, pretežno upotrebljava latinicu. No nije isključeno da će se ova situacija u budućnosti promijeniti, kao što je i 'očišćana' latinica (bez dijakritičkih znakova) doživjela značajan pad u upotrebi. Za veću potporu ćirilici potreban je rad na izgradnji infrastrukture, ali i rad na promjeni navika, pri čemu se ne može isključiti scenario da će upotreba ćirilice na mrežama postati i dio trendova. Veliki jezički modeli, poput ChatGPT-a, donekle prevazilaze ovaj problem s obzirom na mogućnost postavljanja upita i generisanja odgovora na ćirilici. Međutim, vidnoj promjeni navika, barem kod jednog dijela govornika, ne idu u prilog ni nacionalistički narativi o ugroženosti ćirilice koji je stavljaju u domen izrazito konzervativnih i identitetskih okvira, ne ističući značaj jezičkog i grafičkog diverziteta, očuvanja jezičke ekologije, te važnosti digrafije u kognitivnom i kulturološkom smislu [10].

Jezička situacija u Bosni i Hercegovini i digitalno okruženje

I kada je u pitanju digitalno okruženje, slučaj BiH se izdvaja svojom složenošću. Naime, pitanje nedostatka jedinstvene državne strategije u razvoju jezičkih resursa i tehnologija, kako ističe Čušić (2022), za posljedice ima lošiji kvalitet i manju dostupnost jezičkih podataka za bosanski, u odnosu na srpski i hrvatski, kao i nedovoljnu zastupljenost jezičkih varijanti srpskog i hrvatskog u Bosni. Na primjer, najšire dostupan prevodilački alat Google prevodilac pod srpskim jezikom podrazumijeva ekavski izgovor i ćirilično pismo a većina dostupnih tehnologija koje se izrađuju u Srbiji, obično samo nudi samo ijekavski izgovor i nisu trenirana direktno na podacima iz BiH. [11] S druge strane, Internet stranice institucija koje su u administrativnim jedinicama sa srpskom ili hrvatskom većinom nerijetko izbjegavaju upotrebu državnog Internet domena .ba. Stoga prvi bosanski internetski korpus BosWac, koji je trebao biti reprezentativan za cijeli državni prostor, povlačeći podatke samo

sa .ba domena, što je uobičajena praksa prilikom kreiranja web korpusa, zapravo nije predstavljao na odgovarajući način jezičku upotrebu na cijeloj državnoj teritoriji. Sa druge strane web korpus za hrvatski i srpski nisu povlačili podatke sa web sajtova kreiranih na teritoriji Bosne i Hercegovine (npr. srWac ili PDRS 1.0), utoliko manje ako su u pitanju specijalizovani korpusi (Corpus of Legislation texts of Republic of Serbia 1.0), uz rijetke izuzetke (npr. Zečević & Vujičić Stanković 2015 su za kreiranje korpusa koristili tekstove Glasa Srpske i neke kolumne BiH verzije Dojče Velea). Međutim, identifikacija ijekavice kao bosanskog ili hrvatskog a ne srpskog ostaje problem, koji se uočava i prilikom jezičkog outputa na Google prevodiocu, s tim da je ovaj problem donekle prevaziđen kod velikih jezičkih modela, poput ChatGPT-a.

Iako su kasnije ovi nedostaci u nekim novijim verzijama korpusa ispravljani, zahvaljavajući saradnji između lokalnih lingvista i stručnjaka za prirodnu obradu jezika, problem nepostojanja BiH kao jedinstvenog jezičkog prostora, i samim tim, jedinstvenog jezičkog tržišta i dalje ostaje aktuelan problem, koji u ovom trenutku, s obzirom na suprotstavljene jezičke politike i razlike u jezičkim ideologijama nije moguće prevazići (up. Runić & Božović, u štampi). Njegove neposredne posljedice osjećaju se ne samo u domenu razvoja digitalnih jezičkih rješenja već i u istraživanjima jezika u BiH, nastave jezika BiH, kao prvog i stranog jezika i mobilnosti učenika, studenata i nastavnog kadra unutar zemlje.

Razlike u kvalitetu i dostupnosti podataka utiču i na izbor određenih tehnologija kod samih korisnika. U idealnom scenariju, govornici bilo kog od navedenih jezika mogu izabrati bilo koju varijantu i mogu prebacivati iz jedne u drugu varijantu, odnosno lokalizovati je shodno sopstvenim potrebama. Iako je ova vještina donekle svojstvena govornicima koji žive u BiH, zbog višestrukosti formi koji odlikuju ovaj prostor (Halilović 2014) i većoj izloženosti različitim varijantama u odnosu na, recimo, srpski i hrvatski kojim se govori u Srbiji i Hrvatskoj, problem nezastupljenosti sopstvene varijante otvara staro pitanje nedovoljne zastupljenosti BiH na srpskohrvatskom jezičkom prostoru, kao i pitanje jezičkih prava u kontekstu jezika u digitalnoj sferi. [12] U istraživanju koje smo sproveli 2023. godine o upotrebi Google Prevodioca među studentima italijanskog jezika i književnosti na Univerzitetu u Banjoj Luci, kao daleko najčešći par se javio italijanski – srpski (i srpsko-italijanski), u 18 odgovora od ukupno 20 ispitanika, nakon njega italijansko-hrvatski (i hrvatsko-italijanski) u 4 odgovora, dok je par koji je uključivao bosanski imao samo jedan odgovor, iako je ta varijanta najbliža onom što je varijanta koju koriste ispitanici studenti. Iako identitetska pitanja igraju odgovarajuću ulogu prilikom odabira jezika, nije zanemarljiv ni kriterijum dostupnosti i kvaliteta jezičkih podataka (Runić 2023b, Runić & Božović, u štampi). S druge strane, rijetki pokušaji izrade jezičkih rješenja za srpski koji se koristi u BiH – poput interaktivne aplikacije Jatolog za konverziju ekavice u ijekavicu u više od četiri hiljade riječi, trenirane na podacima prikupljenim sa lokalnog portala mondo.ba – nije

zaživjela. Aplikacija pravopisa bosanskog jezika (Halilović 2018), koja uzima u obzir rješenja koja su prihvatljiva na cijeloj teritoriji BiH u današnjim uslovima, nije pak promovisana niti prihvaćena od jezičkih autoriteta pod hrvatskom i srpskom jurisdikcijom.

Pored pitanja međusobnog odnosa u smislu broja, dostupnosti i kvaliteta određenih resursa, zanimljivo je i pitanje granica pojedinačnih jezika, njihove uporedivosti i nekih ideoloških pristrasnosti koje se mogu uočiti u jezičkim podacima na internetu. Tehnički aspekti razlikovanja sličnih jezika – bosanskog, hrvatskog i srpskog - predmet su većeg broja istraživanja, te sljedstveno tome i praktičnih rješenja (up. Ljubešić *et al.* 2007; Ljubešić & Klubička 2014; Zečević & Vujić Stanković 2015; Ljubešić & Lauc 2021; Ljubešić & Rupnik 2022). Iz korisničke perspektive, pak, svakako je zanimljivo u kojoj mjeri tehnologije mogu da razlikuju pojedinačne jezike, prebacuju iz jedne u drugu varijantu i imaju metalingvističko znanje o granicama različitih jezika. Trenutno najšire korištena jezička tehnologija zasnovana na konverzacijskom modelu koji oponaša ljudsku komunikaciju je ChatGPT-4, sa kojim je moguće komunicirati na sva tri jezika i oba pisma. Zbog njegove velike dostupnosti, kvaliteta podataka, posebno u odnosu na neke alternativne lokalne modele (npr. YugoGPT, koji na postavljena jezička pitanja daje dosta više 'halucinacija' u odnosu na ChatGPT-4), te mogućnosti se koristi i kao prevodilačka alatka, ovdje ćemo se osvrnuti samo na neke od uočenih nedostataka, prije svega na planu pristrasnosti (takozvanih *biasa*).

OpenAI, kompanija koja je stvorila ovaj model, nije otkrila podatke koji bi mogli biti od interesa za naučnu zajednicu, uključujući i podatke na kojima je model istreniran (OpenAI 2023). Međutim, s obzirom da radi o podacima koji su dostupni na internetu, moguće je izvesti pretpostavke. U tom smislu, zanimljivo je pitanje 'bosanskog jezika' i načina na koji je ono obrađeno u različitim vikipedijama. Srpska i hrvatska Vikipedija koriste inače neprihvatljiv naziv „bošnjački jezik“ kao glavnu unosnicu. Hrvatska Vikipedija konzistentno koristi pridjev „bošnjački“ kada god je u pitanju jezik. I sam ChatGPT, koji je, po našoj pretpostavci treniran i na podacima sa Vikipedije, koristi ravnopravno termin bosanski i bošnjački kada mu se postavi upit na ćirilici, odnosno nije ideološki osviješten, dok je u latiničnoj verziji u mogućnosti da bude ideološki neutralan i da ponudi i samo obrazloženje o nazivu, kao što je slučaj sa odgovorom u (1). U odgovoru međutim ne postoje naznake o negativnim konotacijama koje ovaj naziv ima za same govornike bosanskog.

1. Bošnjački jezik nije zasebno priznat jezik, već je to termin koji se ponekad koristi u neformalnim kontekstima kao sinonim za bosanski jezik. Zvanično, prema ustavima Bosne i Hercegovine, te prema lingvističkoj klasifikaciji, u Bosni i Hercegovini postoje tri službena jezika: bosanski, srpski i hrvatski. [...] Korištenje termina "bošnjački jezik"

najčešće je politički ili identitetski motivisano, a u formalnom i akademskom kontekstu se uglavnom koristi termin "bosanski jezik."

(OpenAI. (2024). ChatGPT (GPT-4), 28.9.2024, upit 'Šta je bošnjački jezik?')

Još jedno kontroverzno pitanje u srpskom jeziku – pitanje upotrebe rodno osjetljivog jezika, čemu se odlučno protivi veliki dio srbistike, u Srbiji i BiH (Bjelaković & Savić 2021) reflektuje se u jezičkim podacima generisanim na velikim jezičkim modelima. U politički uticajnom dijelu srbistike najveći otpor je prema upotrebi novih socijalnih femininativa (npr. *psihološkinja*, *pilotkinja* ili *pedijatrica*) ili upotrebi femininativa u nereferencijalnoj upotrebi (kao u primjeru *Tijana je doktorka/doktorica nauka, profesorka/profesorica univerziteta*). Kada se uporede odgovori ChatGPT-a na ćirilici u vezi sa upotrebom rodno osjetljivog jezika, broj femininativa je manji i nije dosljedan, kao npr. u latiničnoj verziji, ukoliko se upiti tiču ličnosti iz srpske istorije i umjetnosti. U prevodima ChatGPT ipak dosljedno koristi rodno osjetljiv jezik.

Nova verzija ChatGPT-ja je znatno bolja u smislu rječničkog fonda i daje kvalitetniji output kada prebacuje iz jednog jezika u drugi, iako su neke jezičke crte generisane bespotrebno različito ili čak agramatične (najviše one sintaksičke prirode, gdje su inače najmanje razlike među jezicima). [13] Međutim, u ovom trenutku nije moguće dati neki opšti sud o kvalitetu odgovora s obzirom da ChatGPT svaki put generiše drugačiji output i da smo sada u procesu finog podešavanja (*fine-tuning*). Vrijeme će pokazati da li će i u kojoj mjeri neki od uočenih nedostataka i pristrasnosti biti prevaziđeni, s obzirom na to da su oni inherentni samim podacima. U krajnjem slučaju, i pitanje granica između različitih varijanti jezika, posebno u BiH je takođe teško uhvatljiv koncept, često podložen proizvoljnim sudovima onih koji te granice crtaju.

Zaključak

Razvoj jezičnih tehnologija za jezike koji su nastali na temeljima nekadašnjeg srpskohrvatskog jezika od početka je pratio visok stepen saradnje i razmjene rješenja (Ljubešić *et al.* 2022). Međutim, u radu smo se prvenstveno bavili negativnim posljedicama političko-ekonomskog i jezičkog raskola na postojanje ovih jezika u digitalnom dobu. Među izazovima koje smo istakli su neujednačen razvoj resursa za pojedinačne jezike, različite količine i kvalitet podataka za svaki od jezika, nedovoljno prisustvo svih varijanti uzrokovano političkim podjelama i jezičkim ideologijama, navike korisnika i različite pristrasnosti koje se javljaju u trenutno dominantnim jezičkim tehnologijama.

Polje prirodne obrade jezika je izuzetno dinamično polje. Nova rješenja se iznalaze svakog dana, a često su ona usmjerena na prevazilaženje i unaprijeđenje nedostataka. Povratne informacije koje daje jezička struka i šire, zajednica govornika, od presudnog je značaja za kvalitet i uspjeh određenih tehnologija. Bez obzira što je naše tržište malo, razvoj prirodne obrade jezika je u zamahu. Učešće jezičke struke u razumijevanju nedostataka i saradnja sa stručnjacima iz oblasti prirodne obrade jezika je ključna za uspjeh jezičkih tehnologija, čemu, nadamo se, doprinose i argumenti izneseni u ovom radu.

Bibliografija :

Alexander, Ronelle (2013). Language and Identity: The Fate of Serbo-Croatian. In R. Daskalov & T. Marinov (Eds). *Entangled Balkans*, v. 1. Leiden, Brill Publishers, pp. 341-417.

Bjelaković, Isidora & Savić, Viktor (Eds) (2021). Položaj srpskog jezika u savremenom društvu. Izazovi, problemi, rešenja. Zbornik saopštenja s okruglog stola održanog u Matici Srpskoj 3. jula 2021. godine Novi Sad-Beograd. Matica Srpska, Odbor za standardizaciju srpskog jezika.

Blasi, Damian; Anastasopoulos, Antonios & Neubig, Graham (2022). Systematic Inequalities in Language Technology Performance across the World's Languages. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics Volume 1: Long Papers*, pp. 5486- 5505.

Bogdanović, Miloš & Tošić, Jelena (2022). Corpus of Legislation texts of Republic of Serbia 1.0, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1754>.

Bogdanović, Miloš, Kocić, Jelena, Stoimenov, Leonid (2024). SRBerta—A

Transformer Language Model for Serbian Cyrillic Legal Texts. *Information*. Vol. 15, no 2, pp. 1-12. <https://doi.org/10.3390/info15020074>

Buden, Boris (2017). Padaj (jezična) silo i nepravdo! Produktivni paradoks

Deklaracije o zajedničkom jeziku. *Slobodni Filozofski*, 28. prosinca 2017. <https://slobodnifilozofski.com/2017/12/padaj-jezicna-silo-nepravdo-produktivni-paradoks-deklaracije-zajednickom-jeziku.html>

Bugarski, Ranko (1994). *Jezik od mir do rata*. Beograd, Beogradski krug.

Bugarski, Ranko (1997). *Jezik u društvenoj krizi*. Beograd, Čigoja štampa.

Bugarski, Ranko (2002). Nova lica jezika: sociolingvističke teme. Beograd, Čigoja štampa.

Bugarski, Ranko (2018). Govorite li zajednički. Beograd, XX vek.

Bugarski, Ranko (2021). Saga o ćirilici. Beograd, XX vek.

Bugarski, Ranko, Hawkesworth (2004). Language in the former Yugoslav lands. Bloomington, Slavica.

Bulić, Halid, Durmišević, Elma, Hodžić, Čavkić, Azra, Barjaktarević, Enisa,

Ahmetspahić-Peljto, Azra, Šabić, Belmin (2023). Sarajevski korpus SMS poruka na bosanskom jeziku. Sarajevo, Filozofski fakultet.

Čirgić, Adnan (2007). Jezički neprebol. Cetinje: Institut za crnogorski jezik i jezikoslovlje „Vojislav P. Nikčević“.

Čirgić, Adnan (2009). Rječnik njeguškoga govora. Podgorica – Cetinje: Matica crnogorska.

Ćušić, Tarik (2023). Language Report Bosnian. In: G. Rehm & A. Way, A. (Eds)

European Language Equality. A Strategic Agenda for Digital Language Equality. Cham, Springer, pp. 203-206.

Dawies, William D, Dubinsky, Stanley (2018). Language Conflict and Language Rights. Cambridge, Cambridge University Press.

Giagkou, Maria (2022). Digital Language Equality in Europe: How are our languages doing? Conference presentation at „Workshop Towards full digital language equality in a multilingual European Union“ held at Workshop Towards full digital language equality in a multilingual European Union, November 8 2022, European Parliament, Strasbourg.

Glušica, Rajka (2020). Crnogorski jezik i nacionalizam. Beograd, XX vek.

Greenberg, Robert (2008). Language and Identity in the Balkans: Serbo-Croatian and Its Disintegration. Oxford, Oxford University Press.

Halilović, Senahid (1996). Pravopis bosanskoga jezika. Sarajevo, Preporod.

Halilović, Senahid (2014). Jezička stvarnost u Bosni i Hercegovini. Socjolingwistyka XXVIII, pp.121-135.

Halilović, Senahid (2018). Pravopis bosanskoga jezika. Sarajevo, Slavistički komitet.

Jahić, Dževad (1999). Školski rječnik bosanskog jezika. Sarajevo, Ljiljan.

Kapović, Mate (2011). Čiji je jezik? Zagreb, Algoritam.

Kordić, Snježana (2010). Jezik i nacionalizam. Zagreb, Durieux.

Kornai, András (2013). Digital Language Death. PLOS ONE 8(10): e77056.

<https://doi.org/10.1371/journal.pone.0077056>

Krstev, Cvetana & Stanković, Ranka. (2023). Language Report Serbian.

A. Way, A. (Eds) European Language Equality. A Strategic Agenda for Digital Language Equality. Cham, Springer, pp. 203-206.

Ljubešić, Nikola, Mikelić, Nives & Damir, Boras (2007) Language identification: How to distinguish similar languages? 2007 29th International Conference on Information Technology Interfaces, Cavtat, Croatia, 2007, pp. 541-546, doi: 10.1109/ITI.2007.4283829.

Ljubešić, Nikola & Klubička, Filip (2016). Serbian web corpus srWaC 1.1, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1063>.

Ljubešić, Nikolić; Petrović Miličević, Maja & Samardžić, Tanja. (2018). Borders and boundaries in Bosnian, Croatian, Montenegrin and Serbian: Twitter data to the rescue, Journal of Linguistic Geography, Volume 6, Issue 2, pp. 100-124.

Langston, Keith & Peti-Stantić, Anita (2014). Language Planning and National Identity in Croatia. London, Palgrave Macmillan.

Ljubešić, Nikola & Rupnik, Peter (2022). The Twitter user dataset for discriminating between Bosnian, Croatian, Montenegrin and Serbian Twitter-HBS 1.0. Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1482>.

Ljubešić, Nikola; Erjavec, Tomaž; Miličević Petrović, Maja; Samardžić, Tanja (2022). Together We Are Stronger: Bootstrapping Language Technology Infrastructure for South Slavic Languages with CLARIN.SI. Berlin, De Gruyter. pp. 429-456

Mønnesland, Svein (ed) (2005). Jezik u Bosni i Hercegovini. Sarajevo, Institut za jezik u Sarajevu-Institut za istočnoevropske i orijentalne studije (Oslo).

Organization for Security and Cooperation in Europe (OSCE) (2018). "Dvije škole pod jednim krovom". Najvidljiviji primjer diskriminacije u oblasti obrazovanja u Bosni i Hercegovini. Izvještaj OSCE-a iz novembra 2018. godine.

Požgaj Hadži, Vesna (2013). Jezik između lingvistike i politike. Beograd, XX vek.

Registar nacionalnog internet domena (RNIDS) (2021). Ćirilica u digitalnom svijetu.

You tube video, uploaded January 28 2021 by Registar nacionalnog internet domena.

Rehm, Georg, Way, Andy (Eds) (2023). European Language Equality: A Strategic

Agenda for Digital Language Equality. Cognitive Technologies. Cham, Springer.

Runić, Marija (2023a) It's High Time: Linguistics in High Schools in Bosnia and Herzegovina.

In Pinneda, A. & Corr A. (eds) Theoretical Linguistics in Pre-University Classroom.

Proceedings of the British Academy. Oxford, Oxford University Press, pp. 68-89.

Runić, Marija (2023b). Per uno più consapevole dei traduttori automatici nell'insegnamento di una LS [U prilog savjesnijoj upotrebi alata za mašinsko prevođenje u nastavi stranog jezika], In Baselica, G., Di Giacom Lazarević P. & Perišić, O. (eds.), I gigli nel campo. Studi in onore di Ljiljana Banjanin. Edizioni dell'Orso, Alessandria, pp. 521-530.

Runić, Marija & Božović, Jelena (u štampi). Integrating an Ethnographic Perspective into Researching Languages in Conflict: Insights from Bosnia and Herzegovina. In S. Dubinsky, M. Gavin & H. Starr (Eds), Handbook of Language and Political Conflict. Cambridge, Cambridge University Press.

Simons, F. Garry; Thomas, L. Abbey & White, K. Chad (2022). Assessing Digital Language Support on a Global Scale. In Proceedings of the 29th International Conference on Computational Linguistics October 12–17, 2022, pp. 4299–4305.

Soria, Claudia; Quochi, Valeria; Russo, Irene; Gurrutxaga, Antton; Ceberio, Klara (2017). A Digital Vitality Scales and Indicators. Manuscript.

Starčević, Anđel; Kapović, Mate & Sarić, Daliborka (2019). Jeziku je svejedno. Zagreb, Sandorf.

Šator, Muhamed (2005). Jezička politika u vrijeme Austro-Ugarske. In: S. Mønnesland (ed.) Jezik u Bosni i Hercegovini. Sarajevo, Institut za jezik u Sarajevu-Institut za istočnoevropske i orijentalne studije (Oslo), pp. 321-344.

Šipka, Milan (1999). Standardni jezik u Bosni i Hercegovini u dokumentima jezičke politike. Open Society Institute.

Tadić, Marko (2023). Language Report Croatian. In: G. Rehm & A. Way (Eds)

European Language Equality. A Strategic Agenda for Digital Language Equality. Cham, Springer, pp. 111-114.

Vajzović, Hanka (2008). Jezik i nacionalni identiteti: sociolingvističke teme. Sarajevo, Fakultet političkih nauka.

Wasserscheidt, Philipp (2023). Serbian Web Corpus PDRS 1.0. Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1752>

Zečević, Anđelka & Vujičić Stanković, Saša (2015). Language identification-the case of Serbian. In G. Pavlović-Lažetić, C. Krstev, D. Vitas, I. Obradović (Ed.) Natural Language Processing for Serbian - Resources and Applications, Faculty of Mathematics. University of Belgrade, Belgrade pp. 101–112. Faculty of Mathematics. University of Belgrade, Belgrade.

Notes

[1] Tekst deklaracije dostupan je na <https://www.predsednik.rs/pres-centar/saopstenja/deklaracija-o-zastiti-nacionalnih-i-politickih-prava-i-zajednickoj-buducnosti-srpskog-naroda> [17.11.2024.]

[2] Prema izvještaju za januar 2024. godinu pokrivenost interneta u Bosni i Hercegovini je 84.81%, Hrvatskoj 81.89%, Srbiji 82.95%, Crnoj Gori 90.34%. U Njemačkoj je u poređenju taj procenat 85.56%, u Francuskoj 85.3%, a u Švedskoj 90.44%. Svi podaci preuzeti su sa Statiste (<https://www.statista.com/markets/424/internet/>) [27.9.2024.]

[3] <https://learn.microsoft.com/en-us/windows-hardware/manufacture/desktop/available-language-packs-for-windows?view=windows-11> [27.9.2024.]

[4] Podaci o Vikipediji na srpskohrvatskom preuzeti su sa stranice Vikipedije

[5] Podaci su preuzeti sa https://meta.wikimedia.org/wiki/List_of_Wikipedias [28.9.2024]

[6] Akronim JePTex (jezički resursi i tehnologije) koristi Društvo za jezičke resurse i tehnologije (<https://jerteh.rs/>), sa sjedištem u Beogradu.

[7] S obzirom da se za industrijski i komercijalni sektor radi o malim jezicima i malim tržištima, gotovo da ne postoji nikakav komercijalni interes za razvoj tehnologija za prepoznavanje govora, koje i jesu najskuplje. U ovom trenutku, Office 365 ima verziju za diktiranje i transkripciju samo za hrvatski (doduše, kao verziju u pripremi). Nijedan od vokalnih asistenata (Siri i Google Assistant) ne podržava nijedan od navedenih jezika i

uglavnom je ograničen na jezike sa većim brojem govornika, uz izuzetak švedskog. Takođe, nijedan od jezika ne postoji na platformi za učenje stranih jezika Duolingo.

[8] Opis projekta digitalne jednakosti, zajedno sa pratećom publikacijom dostupan je na stranici <https://european-language-equality.eu>.

[9] Veliki podsticaj za razvoj resursa za jezike na prostoru bivše Jugoslavije i uopšte za južnoslovenske jezike načinjeni su u okviru infrastrukture Clarin.SI, kao i regionalne istraživačke mreže ReLDI (Regional Linguistic Data Initiative, <https://reldi.spur.uzh.ch/>). O značaju pristupa odozdo na gore u razvijanju resursa za srpski i hrvatski, te uloge koju je Slovenija imala kao infrastrukturno razvijenija zemlja, vidjeti Ljubešić *et al.* (2022).

[10] U svakodnevnom jeziku upotreba ćirilice ili insistiranje na ćirilici se obično vezuje za konzervativne, regresivne i desne ideje, a latinica za moderne i progresivne. 'Ćirilica' je tako zadobila i negativne konotacije pa se u neformalnom jeziku može sresti izraz 'ćiriličari' u značenju 'desničari i nacionalisti' dok je 90-tih godina je bio u potrebi izraz 'ćirilica' za novokomponovanu narodnu muziku, koja se obično vezivala za lošiji ukus. U Srbiji se često uspostavlja ideološka dihotomija između pobornika rodno osjetljivog jezika (lijevi i liberalni) i pobornika ćirilice i protivnika rodno osjetljivog jezika (desni i konzervativni).

[11] Googlov pristup srpskom kao ekavskom izgovorom i ćiriličnom pismu ima prednost da omogućava pretragu na ćirilici.

[12] U nekim od glavnih medijskih kuća na prostoru BiH, uobičajeno je da se izjave čuvaju u originalnom obliku (npr. Buka, Nezavisne, ali ne i Klix) ili se ne lektorišu jezički izrazi autora novinara, dok je u banjalučkim kinima moguće gledati distribucije filmova sa prevodom titlova na ekavski i ćirilicu, hrvatski i latinicu, te jezičkom varijantom kojom se govori u BiH. S druge strane, u medijima objavljenim na hrvatskoj teritoriji, sve izjave se prilagođavaju na hrvatski, kao i članci napisani nekom drugom varijantom (kao npr. Portal Novosti, koji angažuje veći broj autora iz regiona). U medijima na teritoriji Srbije mogu se naći oba principa (Danas, recimo, čuva izjave u originalu kao i RTS u nekim slučajevima), iako većina najčitanijih medija prebacuje druge varijante na ekavicu i varijantu srpskog u Srbiji. Međutim, Savet za srpski jezik predlaže da se lektorišu izjave date na hrvatskom ([Briga o ćirilici: Stati na put kroatizaciji srpskog jezika - Vreme](#)) i zahtijeva donošenje novog Zakona o upotrebi jezika i pisama.

[13] ChatGPT 4-o daje tako vrlo detaljno objašnjenje u vezi sa značenjem riječi 'ibretiti se', što u prethodnoj verziji nije bio slučaj, dok je za određene regionalizme u upotrebi u Dalmaciji bio kadar dati vrlo detaljno objašnjenje. I pored toga, pitanje da li ChatGPT ima svijest kojom varijantom zapravo govori je upitno jer ga je lako zbuniti, a i za mnoge upite

koji se tiču arealnih distribucija određenih jezičkih osobina daje ili pretjerano uopštavane odgovore ili suviše neodređene, iz kojih nije moguće izvući posebne informacije.

Pour citer ce document

Par Marija RUNIĆ, «Jezički raskol u ‘stvarnom’ i digitalnom okruženju: slučaj srpskohrvatskog», *Revue du Centre Européen d'Etudes Slaves* [En ligne], Environnement linguistique, Numéro 9, La revue, mis à jour le : 14/04/2025, URL : <https://etudesslaves.edel.univ-poitiers.fr/index.php?id=1999>.

Quelques mots à propos de : Marija RUNIĆ

Marija Runić is a linguist from Bosnia and Herzegovina who works at The University of Banja Luka. She earned her doctorate in linguistics from The University of Padua in Italy. Her research focuses on formal linguistics, specifically in the area of syntax and syntax-semantics interface, sociolinguistics and linguistic and education. In 2019-2020, she was a Fulbright Scholar at the Western Washington University with a research project on linguistics and education in post-conflict societies. Runić ...

Droits d'auteur



This is an Open Access article distributed under the terms of the Creative Commons Attribution License CC BY-NC 3.0 (<https://creativecommons.org/licenses/by-nc/3.0/fr/>) / Article distribué selon les termes de la licence Creative Commons CC BY-NC.3.0 (<https://creativecommons.org/licenses/by-nc/3.0/fr/>)