PREGLEDNI RAD

DOI: 10.5937/reci2316231M

UDC: 81'255

811.111:004.738.5

81'27

NEVENA N. MANIĆ[*]

Univerzitet u Beogradu

Filološki fakultet

JEZIČKE IDEOLOGIJE U DIGITALNIM PROSTRANSTVIMA: SLUČAJ SUPREMACIJE ENGLESKOG JEZIKA U PREVOĐENJU SADRŽAJA NA INTERNETU

Sažetak: Ovaj rad istražuje uticaj jezičkih ideologija na društvenu hijerarhiju u digitalnim prostorima, sa posebnim fokusom na dominaciju engleskog jezika u onlajn prevođenju. Oslanjajući se na postojeću literaturu i studije slučaja, ovaj rad tvrdi da jezičke ideologije igraju značajnu ulogu u oblikovanju odnosa moći i jezičke raznolikosti u digitalnim prostorima. Autor uvodi koncept sistema mašinskog prevođenja (eng. *Machine Translation System*) kao potencijalnog rešenja za etničke manjine povodom zaštite jezika u digitalnom prostoru. U radu se takođe ističe važnost jezičke raznolikosti u onlajn istraživanju i potreba da istraživači usvoje inkluzivniji pristup upotrebi jezika. Nalazi ovog rada imaju važne implikacije za političke aktere, edukatore i istraživače koji su zainteresovani za promovisanje jezičke raznolikosti i socijalne pravde na internetu.

Ključne reči: jezičke ideologije, digitalna prostranstva, internet, supremacija engleskog jezika, mašinsko prevođenje (MT), očuvanje jezika

[*] manic.nevena@gmail.com

UVOD

Verujući u ključnu ulogu jezika na internetu kao najšireg i najrasprostranjenijeg komunikativnog domena, istraživanje koje sledi ima za cilj da istraži stepen diversifikacije lokalnih (nekolonijalnih[1]) jezika na mreži i dominacije jednog konkretnog kolonijalnog jezika, na štetu reprezentativnosti svih drugih. Ova tema se smatra ključnom za poboljšanje njihove vidljivosti i ostvarivanje veće demokratičnosti u jezičkom prostoru, što doprinosi i kulturnom napretku jezičkih zajednica. Rad ima preglednu prirodu, budući da se istraživački postupak ne konkretizuje u analizi pod kontrolisanim uslovima. Umesto toga, korišćeni su hronološki predočeni i naratizovani širokodostupni podaci, uzimajući u obzir statističke informacije u poredbene svrhe, kako bi bio ponuđen akademski komentar o svekolikom stanju jezika na internetu. Preliminarna teza rada se ogleda u činjenici da, uprkos savremenom kulturološkom poretku koji propagira diversifikaciju internet sadržaja i jezičku toleranciju, postoji aktivna monolingvistička ideologija koja ometa ostvarivanje jezičke demokratičnosti kojoj teže današnja društva, akademske zajednice i korisnici interneta. Upravo navedeni problem manifestuje se kroz nadmoćan uticaj engleskog jezika, koji zauzima apsolutni virtuelni prostor na internetu, otežavajući govornicima kojima engleski nije maternji jezik pristup kvalitetnim i nepristrasnim informacijama. Osnovna pretpostavka je da nedostatak kvalitetnih sistema mašinskog prevođenja (eng. *Machine Translation Systems)* doprinosi ugroženosti lokalnih jezika, o čemu će detaljnije biti reči u narednim poglavljima, s posebnim fokusom na kardinalnog nosioca sistema odgovornog za kvalitet mašinskog prevoda, tehniku obrade prirodnih jezika (eng. *Natural Language Processing* – NLP). Biće izloženi loši primeri rada na NLP tehnici, kao i sugestije za bolji tehnički performans. Dalje, centralna poglavlja rada pružaju statističke podatke koji podržavaju pretpostavku da nekolonijalni jezici nisu dovoljno zastupljeni na internetu. U zaključku će biti prikazan primer dobre prakse, u vidu inicijative *Language Justice* i njihovih rezultata, zahvaljujući čijem aktivnom radu se dostiže veća vidljivost i novi mehanizmi sakupljanja naučnog znanja na različitim jezicima na internetu.

[1] Termin se odnosi na sve jezike koji ne pripadaju takozvanoj grupi kolonijalnih jezika, odnosno jezika dominantnih evropskih sila u vreme perioda svetske kolonijalizacije, prevalentno: engleski, španski, francuski i portugalski (v. Riney, 1998).

JEZIČKE IDEOLOGIJE I TRŽIŠTE RADA

Na pomen jezika i sveta, neminovno i nesvesno javlja se mentalna slika globalne distribucije jezika koja ne ide u podjednaku korist svim akterima na jezičkom tržištu.[2] Socijalna i ekonomska moć postižu se, kako neposrednim delovanjem, tako i posrednim angažmanom jezičkog aparata ljudi, radi ustoličavanja pomenute moći, pri čemu se, konstantnom i stabilnom upotrebom istih jezika u određenim socijalnim kontekstima (kao što je internet), konstituiše jezička ideologija. Radom koji sledi nastojaću da suzim uporište tematskog okvira navedenih jezičkih ideologija današnjice, koji se prevalentno bavi društvenom hijerarhijom uspostavljenom kroz jezički diskurs u određenoj zajednici, na domen digitalnih prostora, gde su jezičke ideologije podvrgnute znatno manjem nadzoru. Imajući na umu da su jezičke ideologije duboko ukorenjene metalingvističke konceptualizacije jezika (Blommaert, 2006), kao i da se oslanjaju na društvene faktore poput obrazovanja, porodice i rada koji utiču na vrednovanje izbora jezičkog varijeteta, jezik neminovno dobija svoju kvantitativno merljivu vrednost na osnovu tržišta rada i njegovih potreba. Institut za ekonomiju rada na Univerzitetu u Bonu[3] ističe da jezik ima ekonomsku vrednost (Grenier, 2021: 3) i kao sredstvo komunikacije, ali i kao vid kulturne manifestacije (Hofstede, 2010). Kulturna vrednost najčešće nije izražena kvantitativno, no podjednako je važna za opstanak jedne jezičke zajednice (Granier, 2021: 7). Budući da ekonomske koristi od zajedničkog jezika prevazilaze ličnu dobit, vlade, među ostalim društvenim akterima, mogu podsticati ili nametati upotrebu jezika koji neki članovi zajednice ne bi prvobitno izabrali kao svoj primarni jezički izbor. Nedavna analiza jezičke politike iz perspektive pravičnosti zauzima provokativan stav da se „širenju trenda poboljšanja jezičke kompetencije na engleskom jeziku ljudi ne treba odupreti niti je preokrenuti, već prihvatiti i ubrzati“[4] iz razloga ne samo ekonomske efikasnosti, nego i pravičnosti (Van Parijs, 2011). Dakle, monolingvizam i primat jedinstvenog zajedničkog jezika na tržištu navodno ne mogu da štete očuvanju jezičkog diverziteta i sveopštem kulturnom bogatstvu, prenosivim kroz prevashodno jezičke prakse (usmene i pisane), niti su u opoziciji sa sistemom ekonomskog vrednovanja jednog jezika. Međutim, pod pretpostavkom da je ovaj stav tačan, šta se dešava kada na tržište rada uvedemo aspekte poput međunarodne saradnje ili globalne prodaje proizvoda/usluga koji potiču iz drugačijeg jezičkog ambijenta ili konkurentnost firmi

[2] Mnogi istraživači izrazili su zabrinutost zbog ogromne dominacije engleskog jezika na internetu, dok je drugi ističu kao povoljnu za dalji razvoj globalizacije (v. Danet & Herring, 2007).

[3] IZA Institute of Labor Economics initiated by Deutsche Post Foundation

[4] Prevod autora

različitog porekla u određenoj industriji, ili razmenu znanja? Tada prethodno pomenuta socioekonomska pravičnost upotrebe jezika gubi svoj prioritet.

AKADEMSKA PRODUKCIJA U DIGITALNOM PROSTORU

Za potrebe ovog rada, usredsrediću se na jedno specifično tržište rada, usko određeno upotrebom jezika: akademsku zajednicu. Uvodeći značajan pojam sistema mašinskog prevođenja (eng. *Machine Translation System*) koji se uveliko koristi u sferi današnje internet edukacije kao potencijalno rešenje za etničke manjine čije je očuvanje jezika u digitalnom prostoru ugroženo (Daneteds, 2007: 21), izvesno je da performans lošijeg kvaliteta takvog MT sistema utiče na mnoge jezike i njihovu manifestaciju, među kojima su akademska produkcija i dostupnost akademske građe na različitim jezicima na platformama poput pretraživača *Google Scholar (*Burlot & Yvon, 2017). Naučni istraživači i saradnici širom sveta koji, spletom raznoraznih društvenih faktora (obrazovanje, porodica, radno iskustvo, nacionalna jezička politika, jezik nacionalne manjine), nisu bili dovoljno izloženi engleskom jeziku, a s pravom su zaradili svoju poziciju u naučnoistraživačkim i univerzitetskim institucijama, onemogućeni su da dalje razvijaju svoj akademski kapacitet, jer im međunarodna naučna građa dostupna na internetu, koja je u praksi isključivo na engleskom jeziku (samostalna pretraga domena *Google Scholar* osnovni je dokaz iznesenog stava), nije u potpunosti razumljiva. U takvim situacijama, govornici lokalnih jezika (ali ne samo oni) mogu pribeći mašinskom prevođenju kako bi bolje savladali određeni tekst. Međutim, kvalitet mašinskog prevoda sa engleskog na, primera radi, svahili jezik, nije ni približno jednak kvalitetu mašinskog prevoda sa engleskog na francuski jezik (Araújo, 2019: 9). Odbrambeni argument softverskih kompanija uvek je isti: manjak jezičkog uzorka na osnovu kojeg bi se poboljšao kvalitet.

ZNAČAJ MAŠINSKOG PREVODA

Shodno istraživanjima koje je sprovela prevodilačka kompanija *Weglot*, kvalitet mašinskog prevoda koji u svojoj kombinaciji sa engleskim ima španski, nemački, francuski, italijanski ili portugalski jezik, znatno je viši u poređenju sa prevodom u kombinaciji engleskog i bengali jezika[5]. Na to koliko je trivijalan argument o manjku jezičkog uzorka ukazuje činjenica da govornika francuskog jezika kao maternjeg ima 64 miliona, prema podacima Etnologa (eng. *Ethnologue*) iz 2020. godine, dok prema istom izveštaju, govornika

[5] Nimdzi Insights, Weglot (2022). *The State of Machine Translation for Websites: A Comparative Study of the Top Machine Translation Engines*.

bengali jezika kao maternjeg ima 159 miliona.[6] Zaključak koji se sam nameće svedoči da jezički uzorak nije u direktnoj vezi s kvalitetom prevoda, jer da je to slučaj, indijski jezički varijeteti davno bi premašili performans MT sistema na jezicima Evropske unije, čijih govornika zajedno ima tri puta manje nego govornika indijskog potkontinenta (EU ima 477 miliona stanovnika, dok ih Indija ima 1,4 milijarde). Ni izloženost mobilnom telefonu, koji je neophodan da bi se prikupio jezički uzorak za potrebe poboljšanja kvaliteta mašinskog prevoda, ne ide u prilog prethodno pomenutom argumentu, budući da je broj korisnika pametnih telefona u Indiji u 2020. godini dostigao 748 miliona[7], dok internet praksa pokazuje da MT sistemi odlično funkcionišu sa mnogo manjim uzorkom, uzevši u obzir prevod među evropskim jezicima. Takođe, rezultati Izveštaja o stanju jezika na internetu[8] pokazuju da 90% internet korisnika sa afričkog govornog područja mora da se pozove na svoj drugi jezik prilikom internet pretrage[9]. Nedostatak zalaganja za optimizaciju MT sistema posredno utiče na stagniranje, efemernost i kvalitet naučnokulturnog razvoja jedne jezičke zajednice. Određena količina internet sadržaja (inicijalna pretraga, kao i materijal na sajtovima) nudi automatski generisani prevod (u skladu sa jezičkim potrebama i geolokacijom), koji je, sveopšte sudeći, generisan loše. Kvalitet prevoda nanovo utiče na ponudu medijsko-informativnog sadržaja na internetu (imajući u vidu da se sve veći broj vesti stvara korišćenjem veštačke inteligencije i automatski generisanog prevoda), samim tim i na izloženost šireg dela zajednice relevantnim podacima, a takođe utiče i na usporavanje procesa integrisanja korisnika u globalnu internet zajednicu i praćenje međunarodnih novosti. Neretko stanovnici zemalja sa opresivnim režimom na vlasti nemaju gotovo nikakav pristup informacijama od značaja za politički život, odnosno do tih informacija bi mogli da dođu putem inostranih novinskih agencija, nevladinih organizacija i udruženja, koji se oslanjaju jedino na kontaktni jezik kako bi uspostavili komunikaciju sa (čitalačkom) publikom. Engleski se nameće kao *lingua franca* (Crystal, 2003; Mares, 2016), čak i tamo gde su u zvaničnoj upotrebi drugi kontaktni jezici, budući da je upitno da li je stanovnicima Demokratske republike Kongo ili Kameruna (u kojima se trenutno dešavaju veliki etnički

[6] Summary by language size. Ethnologue

[7] Podaci istraživanja koje je sprovela međunarodna kompanija *Statista* (https://www.statista.com/statistics/262946/most-common-languages-on-the-internet/)

[8] *State of the Internet's Languages report* (https://internetlanguages.org/en/numbers/a-platform-survey/)

[9] Termin „drugi jezik" nije bliže preciziran u ovom izveštaju, što ostavlja prostor za tumačenje termina i kao drugog službenog jezika u upotrebi u određenoj afričkoj državi, i kao stranog jezika koji je usvojen kroz obrazovanje.

sukobi i napadi[10]) lakše da razumeju informacije od značaja za njihov svakodnevni život na engleskom ili pak na francuskom, koji je zvanični jezik u 21 afričkoj zemlji. S druge strane, dok svetski priznate novine *New York Times* imaju u ponudi prevod svoje stranice samo na španski i kineski jezik, renomirani časopis *Guardian* nema ni toliko, već se celokupan sadržaj nudi jedino na engleskom jeziku, a isto važi i za druge uvažene redakcije, poput *The Economist* ili *Deutsche Welle* (koji generiše sadržaj isključivo na engleskom jeziku, iako je u pitanju nemačka novinska agencija).

ZASTUPLJENOST LOKALNIH JEZIKA U DIGITALNOM PROSTORU

Suštinski, manjak reprezentacije, odnosno zastupljenosti lokalnih ili makar kontaktnih jezika alternativnih engleskom, na internetu ostavlja ozbiljne tragove. Neke od dalekosežnijih posledica otelovljene su u širenju netolerantnosti, segregacije, kao i govora mržnje na najbržem komunikativnom kanalu današnjice. Na osnovu podataka koje je izneo *Deutsche Welle* u emisiji *Shift* posvećenoj tehnologiji[11] (kao i *BBC News, CNN* i druge medijske kompanije), Gugl prevodilac je skoro dodao 24 nova jezika u svoju uslugu, poput kečua, kojim se služi oko 8–10 miliona ljudi u Latinskoj Americi ili bhojpuri, koji koristi oko 50 miliona ljudi u severnoj Indiji. Međutim, nejasno je da li je ovaj gest učinio pomak ka većoj inkluziji ili je puki gest lažne jezičke demokratičnosti. Najosnovniji problemi sa kojima se suočava digitalno društvo, poput privilegovanih jezičkih zajednica na mreži, govora mržnje, pristrasnosti i netolerantnosti, koje internet provajderi još uvek nisu prevazišli, nisu rešeni ovim postupkom. Procenjuje se da je 60–70% veb sadržaja napisano na engleskom jeziku, zavisno od izvora, dok su manji jezici *de facto* nedovoljno zastupljeni na mreži (Flammia & Saunders, 2009: 1899). Kako tvrde različite istraživačke kompanije i vebsajtovi (*Whose Knowledge?*, *Statista, Visual Capitalist, Web Technology Surveys*), iako postoji oko 7000 jezika širom sveta, samo 500 od njih generiše internet sadržaj, što znači da mnogi ljudi moraju da se okrenu drugom jeziku da bi pregledali ili koristili aplikacije. Na osnovu iznesenih podataka, dovodi se u pitanje nužnost postojanja ovakvog stanja i potreba za iznalaženjem mehanizama koji bi potpomogli generisanje internet sadržaja na lokalnim jezicima.

ODGOVORNOST SOFTVERSKIH KOMPANIJA

Mogući tvorci nedovoljno delotvornih mehanizama su tehnološke kompanije i investitori koji bi trebalo da ulažu u unapređivanje softvera dizajniranih i prilagođenih

[10] (v. Tull, 2017)

[11] *DW Shift* (https://www.dw.com/en/shift-living-in-the-digital-age/program-15566070)

lokalnim jezicima, umesto engleskom. Ukoliko se ne sprovodi kontinuirana optimizacija softvera za generisanje i prevođenje jezičkih podataka, nedostupnost nepristrasnih, kvalitetnih višejezičnih informacija na internetu uticaće na usporavanje globalne integracije zajednica iz zemalja u razvoju, što će doprineti evidentnom stagniranju u istoimenoj nikad završenoj fazi razvoja. Međutim, odgovor internet kompanija uvek je isti: postoje onlajn prevodioci (ili automatski prevodi nekih internet stranica). Ali kada oni zakažu, ne zbog tehničke nemogućnosti nego zbog nedovoljnog ulaganja u njihovo poboljšanje, postaje vidno da takva jezička distribucija utiče na osnaživanje postojećih ideologija diskriminacijom drugih jezika, nacija, kultura. Primera radi, govor mržnje u Mjanmaru podstakao je ekstremno nasilje protiv muslimanske manjine Rohingja 2017. godine[12]. Kako prenosi međunarodni magazin *Amnesty International*, Rohingja izbeglice su pokrenule tužbu protiv kompanije *Meta* 2021. godine, smatrajući da Fejsbuk nije obezbedio moderiranje sadržaja (eng. *content moderation*) za lokalne jezike, a u Mjanmaru ih ima oko 100, pod optužbom da je mreža dozvolila širenje dezinformacija i utrla put ekstremnom nasilju. Iako je u *Meti* zaposleno petnaest hiljada moderatora, činjenica je da se moderiranje sadržaja uglavnom vrši na engleskom, dok se drugi jezici često zanemaruju. Svakako, nedovoljna zastupljenost lokalnih jezika na mreži ima i druge nedostatke. Neke aplikacije, kao što je platforma *Google Maps*, nekompatibilne su za korišćenje na manje zastupljenim jezicima. Gugl mape ponuđene su u različitim jezičkim varijantama, ali ono što je zapravo dostupno varira u velikoj meri. Može se pristupiti rezultatima pretrage na engleskom širom sveta, dok su, na primer, rezultati za hindi, jezik koji govori pola milijarde ljudi, ograničeni na određene regione. Izvan tih regiona, korisnici moraju da pređu na engleski. Takođe, sadržaj dostupan na manjim jezicima može biti pristrasan. I dalje je izuzetno teško pronaći obrazovni i politički diversifikovani sadržaj na indonežanskim i indijskim jezičkim varijetetima[13]. Isključivanje manjih jezika znači isključivanje kultura, što ne bi smelo da se dešava ni u stvarnom, ni u digitalnom svetu, jer „jezici su nosioci naših kultura, kolektivnog pamćenja i vrednosti. Oni su suštinska komponenta svačijeg identiteta i gradivni blok naše raznolikosti i životnog nasleđa“[14].

[12] *Amnesty International*. (2022). Preuzeto sa https://www.amnesty.org/en/latest/news/2022/09/myanmar-facebooks-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report/

[13] *State of the Internet's Languages report*. (2022). Preuzeto s*a https://internetlanguages.org/en/*

[14] *United Nations Educational, Scientific and Cultural Organization* – UNESCO (2006). Preuzeto sa https://ich.unesco.org/en/ich-and-mother-languages-00555, prevod autora.

MAŠINSKO PREVOĐENJE I JEZIČKA PRAKSA NA INTERNETU

Dakle, da li je mašinsko prevođenje na internetu odraz istinske društvene jezičke prakse u upotrebi danas? Na zakonodavnom nivou, već se uveliko pokreće kampanja obavezivanja tehnoloških kompanija kao što je *Meta* na moderiranje govora mržnje na više različitih jezičkih varijeteta, a ne samo na engleskom, uz ograničavanje pristrasnosti kompanija koje se bave analizom jezika, premda istrajava nedoumica da li su softverski konglomerati uistinu onemogućeni da optimizuju učinak MT sistema na jezicima rasprostranjenim u zemljama u razvoju. Uprkos mogućnosti prikupljanja velikog jezičkog uzorka od govornika lokalnih jezika i mogućnosti korišćenja njihovog diskursa za proveru kvaliteta, softverske kompanije i dalje ne ulažu u bolji performans svojih tehnoloških rešenja na nekolonijalnim jezicima. Naposletku, sporno je da li navedeni tehnološki giganti iz Silicijumske doline[15] imaju koristi od zakržljalog performansa svojih softvera za MT sisteme na lokalnim jezicima u digitalnom prostoru.[16] Kako bi se izbegle spekulacije u vezi sa vidljivim i prikrivenim motivima dejstvovanja tehnoloških kompanija, naučnostručna zajednica radi na sanaciji navedenih jezičkih praksi na internetu, među kojima su i softverski mehanizmi za razumevanje ljudskog jezika.

NLP TEHNIKA

Softverski derivat, koji podrazumeva mašinsko procesiranje prirodnog jezika i koji napreduje vrtoglavom brzinom, stoji na raspolaganju kao ključni faktor u rešavanju tehničkih prepreka koje utiču na istrajavanje pristrasnih ili diskriminatornih jezičkih markera na internetu. Naime, obrada prirodnih jezika (eng. *Natural Language Processing*) je softverska procedura u okviru veštačke inteligencije kojom se unakrsno prevode podaci sa ljudskog jezika na kompjuterski, odnosno mašinski jezik (Beheraeds, 2023). NLP alati generišu i kontrolišu, ne samo celokupan sadržaj koji korisnici vide na internetu, već i prevod tog sadržaja, odnosno predstavljaju srž softvera mašinskih prevodilaca. Veštačka inteligencija (VI) *a priori* uči od čoveka, tako da je na njemu zadatak da poduči VI korektnim načinima komunikacije i upotrebe jezika u načelu (Branković, 2017). Radi slikovitijeg shvatanja

[15] „Silicijumska dolina“, u kalifornijskom zalivu u Americi, prednjači kao najistaknutije i najskuplje američko tehničko središte, sa fondom talenata od skoro 380.000 tehnoloških radnika okupljenih iz celog sveta; izvor: *Visual Capitalist*, 21. septembar 2022.

[16] Internet popularnost jezika je očigledno kontradiktoran pojam u ovom kontekstu. Ako se kao parametar uzme broj stanovnika, španski je drugi najpopularniji jezik na svetu sa oko 493 miliona govornika. Uprkos tome, na internetu je svega 4% internet sadržaja generisano na španskom jeziku (na prvom mestu je engleski sa 60,4%, dok je na drugom mestu ruski jezik sa 8,5% generisanog sadržaja na internetu).

dejstvovanja ovih alata korisno je iskoristiti primer Tvitera. Ukoliko stvarna praksa na Tviteru pokazuje da jezik korisnika obiluje stigmom, nije u pitanju greška kompjutera nego operatera koji su na taj način podučili kompjuter da postupa, odnosno dizajnirali NLP softver (Hovy, 2018). Ako je primetno da internet sadržaj generisan na srpskom jeziku ne koristi ženske agentivne sufikse, krivica pripada isključivo struci, koja takve oblike nije uvela u algoritam kojim se NLP služi. Istim principom rukovodi se i softver mašinskog prevoda. Da bi ponuda kvalitetnih i značajnih informacija na internetu bila višejezična, inženjeri moraju osnažiti temelje svojih NLP alata radi boljeg ukrštanja prevodnih ekvivalenata između izvornog i ciljnog teksta, što podrazumeva „hranjenje“ softvera digitalizovanim podacima preuzetim iz savremenih gramatika, rečnika, pravopisa i korpusa.

BUDUĆNOST NLP TEHNIKE

Nedavna istraživačka saznanja upućuju na činjenicu da se budućnost NLP tehnike može sagledati u novom, pravednijem svetlu, time što će suštinski prerasti u razumevanje prirodnog jezika (eng. *Natural Language Understanding*), umesto u njegovo puko procesiranje (Cambria & White, 2014). Jedan od najvećih izazova sa kojim se inženjeri NLP softvera suočavaju primetan je u obilju kako semantičke, tako i morfosintaksičke dvosmislenosti pri generisanju jezičkih podataka. Veštačka inteligencija do sada nije uspela da usvoji tehnike učenja svojstvene čoveku koje podrazumevaju razlučivanje značenja u dvosmislenim delovima rečenice. Ona takođe nema kapacitete konceptualizacije semantičke sprege između materijalnog i apstraktnog, kao ni mogućnost prepoznavanja određenih stilskih figura poput ironije, premda se intenzivno radi na unapređivanju takvih alata[17]. Druga mogućnost prevazilaženja osetnih posledica supremacije engleskog jezika u digitalnim prostranstvima podrazumeva promenu perspektive. Naime, naučna zajednica ohrabruje istraživače, inženjere, programske dizajnere i druge relevantne učesnike u lancu jezičke produkcije na internetu da prilagode svoju delatnost u skladu sa istaknutim principom jezičke pravičnosti. Ako bi trebalo sprovesti istraživanje većeg obima i dužeg trajanja koje se služi metodološkim izvorima na internetu i onlajn anketiranjem, takva istraživanja bi nužno trebalo da budu višejezična, kao i da svoja pitanja postave na više jezika kojima raspolaže ciljna grupa. Isto važi i za programe koji se uvode na tržište, za ponudu vebsajtova, kao i za informativno-medijski sadržaj. Sve što se plasira na internetu tehnički ima kapacitete generisanja na više različitih jezika. Pitanje je da li postoji volja, odnosno motivacija za takav poduhvat.

[17] V. Reyes, et al., 2013.

PRIMERI LOŠE NLP PRAKSE

Gugl pretraživač je u potpunosti osposobljen da paralelno prikaže rezultate na više različitih jezika, ili makar na dva jezika koja su u upotrebi u nekoj od zemalja (Wangeds, 2009), međutim, algoritmi po kojima pretraživač funkcioniše rukovode se mnogim faktorima, između ostalih i frekventnošću, odnosno pretpostavkom da, ako je najveći broj korisnika na internetu kliknuo rezultat na engleskom, a ne, primera radi, na indonežanskom jeziku, onda pretraživač „smatra" da nema potrebe da se među prvim rezultatima ponudi išta drugo, budući da će korisnik sigurno razumeti sadržaj na engleskom, čak iako je pretraživač osposobljen da geolocira korisnikovu poziciju, te da zaključi kako se nalazi u Indoneziji u momentu pretrage[18]. U krajnjoj liniji, ako korisnik i razume sadržaj na engleskom, zašto bi bio uskraćen za potpunije razumevanje teksta, naročito ukoliko tekst već postoji na njegovom maternjem jeziku? Upravo vodeći se ovom mišlju, mnogi korisnici, iz straha da pretraživač neće pružiti dovoljno kvalitetnu i široku ponudu rezultata na njihovom izvornom jeziku, često automatski u pretragu unose upit (eng. *query*) upravo na engleskom jeziku, bez obzira na ortografsku ispravnost unesenog upita, jer se izuzetno radilo na ulaganju u povećanje kapaciteta softvera da prepozna zadati upit sa pravopisnim greškama generisanim na engleskom[19]. U duhu predložene opcije o podizanju svesti po pitanju jezičke dominacije na internetu, diskusiju o upotrebi maternjeg i kontaktnih jezika valjalo bi začeti već na osnovnim emancipatorskim nivoima, promenom institucionalnog okvira koji takvu upotrebu podstiče, odnosno obeshrabruje, bez ikakve aluzije na isključivanje iz savremenih tokova. Globalizacija nikada u teoriji nije podrazumevala izopštavanje drugih jezika i kultura, naprotiv, jedan od osnovih ciljeva umrežavanja čitavog sveta u jedinstveno tržište jeste doprinos boljoj vidljivosti i reprezentaciji izvornih kultura i jezičkih varijeteta[20].

[18] Kompanija *Google* je na svom vebsajtu objavila celokupan vodič o parametrima koje uzima u obzir prilikom korisničke pretrage podataka na internetu i na koji način optimizuje pretragu. Za više informacija videti: https://www.google.com/search/howsearchworks/our-approach/

[19] *TELUS International* je jedan takav primer; u pitanju je globalni provajder rešenja za digitalno korisničko iskustvo, sa fokusom na tehničku podršku i poboljšanje digitalnih procesa. Kompanija je osnovana 2005. godine kao podružnica *TELUS Corporation*, kanadske telekomunikacione kompanije, i od tada je prerasla u samostalan biznis sa preko 50.000 zaposlenih u više od 50 zemalja. Određena grupa zaposlenih ima pretežni zadatak da ocenjuje performans rada pretraživača upravo na osnovu grešaka koje bi korisnik, nemajući maternje poznavanje engleskog jezika, mogao da počini, kako bi mu, uprkos pravopisnim greškama, pretraživač, zauzvrat, pružio željenu informaciju.

[20] Sheshrao Chavhan, V. (2016). *Globalization and its Impact on the Cultural Diversity*.

INICIJATIVA *LANGUAGE JUSTICE* KAO PRIMER DOBRE PRAKSE

Predložene opcije za sanaciju posledica jezičke supremacije engleskog u digitalnim prostorima već su materijalizovane u inovativnoj i autentičnoj inicijativi pod nazivom *Language Justice*, koju je pokrenulo udruženje *Whose Knowledge*, a koje predstavlja globalnu kampanju za sakupljanje naučnih podataka zajednica koje su marginalizovane na internetu, sačinjenu od profesora, istraživača i volontera širom sveta. Pokazatelj dobre prakse umrežavanja ljudi sa istaknutom svešću o datoj problematici jeste konvencija organizovana 23. oktobra 2019. godine pod nazivom „Dekolonizacija internetskih jezika“[21], čiji je cilj otvaranje dugoročne panel diskusije na goruću temu dominacije kolonijalnih jezika na internetu i zapostavljanja jezika marginalizovanih zajednica, odnosno promene paradigme sakupljanja digitalnog znanja. Kako prenosi izveštaj, konvenciji je prisustvovalo 30 učesnika koji su sopstvenim identitetom istakli socijalni diverzitet koji lokalni jezici poseduju: 65% učesnika su bile žene ili nebinarne/trans osobe, polovina je došla sa svetskog Juga, a većina je govorila više od jednog jezika. Jedan od osnovnih zaključaka konvencije upućuje na činjenicu da nije potrebno mnogo kako bi se izneta ideja pretočila u delo, i da u korenu leži doprinos emancipaciji, odnosno širenju moderne perspektive edukacije koja neće potirati lokalni jezik spram kolonijalnog.[22] Nakon uspešno završenog zasedanja, 2022. godine je usledila saradnja sa Centrom za internet i društvo[23] i Oksfordskim internet institutom[24] i napisan je prvi izveštaj o stanju jezika na internetu. Izveštaj je poslužio kao izuzetna smernica za istraživačke poduhvate, ali i za statističku obradu važnih podataka na internetu, kao i za podizanje svesti o značaju sakupljanja celovitog znanja na izvornom jeziku. Udruženje aktivno radi na širenju kampanje o jezičkom diverzitetu putem dostupnih medija, platformi društvenih mreža i spajanja važnih aktera radi doprinosa poboljšanju vidljivosti lokalnih jezika na internetu. U duhu inicijative i predloženih progresivnih postupaka, zaključna misao ovog rada poistovećuje se sa mišlju velikog političkog aktiviste, bivšeg predsednika Južnoafričke Republike, ali pre svega, ogromnog borca za mir, toleranciju, društvenu jednakost i pravdu, mišlju koja je istaknuta na samoj stranici inicijative *Language Justice*:

[21] *Decolonizing the Internet's Language – Summary Report (*28. februar 2020.) Preuzeto sa https://whoseknowledge.org/resource/dtil-report/

[22] Idem

[23] Centre for Internet & Society

[24] The Oxford Internet Institute

„Ako razgovarate sa čovekom na jeziku koji on razume, prodirete mu u glavu. Ako sa njim razgovarate na njegovom izvornom jeziku, prodirete mu u srce.“[25] – Nelson Mandela.

ZAKLJUČAK

Dominacija jednog jezika ne bi smela da bude tumačena pod prizmom invazivnog dejstvovanja na druge jezike, jer njihovo očuvanje ne zavisi od globalnih jezika već od samih govornika i svakodnevnih digitalnih praksi koje bi pospešile vidljivost i prisutnost sopstvenih jezičkih varijeteta na internetu. Engleski je stekao reputaciju globalnog jezika zbog snažne ekonomije i vojne moći zemalja u kojima se govori (Weiss, 2005: 6). Međutim, globalizacija neće zaustaviti napredak lokalnih jezika ako se prate primeri zvaničnih institucija poput UNESCO-a, koji kontinuirano radi na promovisanju lokalnih jezika u digitalnoj sferi kako bi se očuvao kulturni diverzitet zagovaranjem jednakog pristupa digitalnim informacijama (Paolillo, 2005: 45). Svaki pojedinačni govornik, a istovremeno i korisnik interneta, trebalo bi da sledi isti princip odabirom upotrebe maternjeg, umesto stranog jezika u digitalnom prostoru. Međutim, postizanje ovog cilja neizbežno se oslanja na dostupnost tehnološke podrške. Ukoliko veruju u osnovne postulate ljudskih prava i pravičnosti, te ako žele da nastave da imaju ogroman broj korisnika od kojih profitiraju, velike tehnološke kompanije bi morale osigurati potpunu i prilagođenu upotrebu lokalnih jezika na internetu, što uključuje lingvistički optimiziranu pretragu i osnažen sistem mašinskog prevođenja. Tehnološke inovacije navedenog tipa biće ključne u osiguravanju relevantne pozicije lokalnih jezika u digitalnoj sferi. Mašinsko prevođenje se kontinuirano poboljšava, ali još uvek postoji potreba za dubljim razvojem i usavršavanjem kako bi se osigurala visoka stopa kvaliteta i sveobuhvatna pouzdanost. Osim toga, potrebno je osigurati dostupnost ovih tehnologija, kao i bolju informisanost društva o njihovoj upotrebi kako bi se podstakla šira upotreba lokalnih jezika na internetu.

LITERATURA

Araújo, M., Pereira, A., & Benevenuto, F. (2020). A comparative study of machine translation for multilingual sentence-level sentiment analysis. *Information Sciences*, 512, 1078–1102.

Bellini, P. (2022). State of the Internet’s Languages Summary report. *Internet Languages*. Dostupno preko: https://internetlanguages.org/media/pdf-summary/EN-STIL-SummaryReport.pdf [14.3.2023]

[25] Prevod autora (izvor: *Whose Knowledge?*)

Bhutada, G. (2021, May 26). The most used languages on the internet. *Visual Capitalist*. Dostupno preko: https://www.visualcapitalist.com/the-most-used-languages-on-the-internet/ [14.3.2023]

Blommaert, J. (2006). Language policy and national identity. *An introduction to language policy: Theory and method*, pp. 238–254.

Branković, S. (2017). Veštačka inteligencija i društvo. *Srpska politička misao*, 2, 13–32.

Burlot, F., & Yvon, F. (2017, September). Evaluating the morphological competence of machine translation systems. *2nd Conference on Machine Translation* (WMT17), Vol. 1, pp. 43–55.

Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational intelligence magazine*, 9 (2), pp. 48–57.

Danet, B., & Herring, S. C. (Eds.). (2007). *The multilingual Internet: Language, culture, and communication online*. Oxford University Press on Demand.

Deng, L., & Liu, Y. (2018). A joint introduction to natural language processing and to deep learning. *Deep learning in natural language processing*, pp. 1–22.

Ditter, R. (2022, June 9). Why Google adding your language to Google translate isn't good enough. *DW Shift: Technology*. Dostupno preko: https://www.youtube.com/watch?v=h_VNB66xWEA [10.3.2023]

Ethnologue, Languages of the World. (2002). Dostupno preko: https://www.ethnologue.com/ [25.4.2023]

Fagnani, R. (2022, September 29). Myanmar: Facebook's systems promoted violence against Rohingya; Meta owes reparations. *Amnesty International*. Dostupno preko: https://www.amnesty.org/en/latest/news/2022/09/myanmar-facebooks-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report/ [14.3.2023]

Flammia, M., & Saunders, C. (2007). Language as power on the Internet. *Journal of the American Society for Information Science and Technology*, 58(12), 1899–1903.

Granov, A. (2010). *Digitalizacija jezika i razvoj jezičkih tehnologija u funkciji digitalizacije kulturnog nasleđa*. Beograd: Matematički fakultet Univerziteta u Beogradu.

Grenier, G., Zhang, W. (2021). The value of language skills. *IZA World of Labor,* 205

Heller, M. (1997). Autonomy and interdependence: language in the world. *International Journal of Applied Linguistics*, 7(1), pp. 79–85.

Hofstede, G., Hofstede, G. J. i Minkov, M. (2010). *Cultures and Organizations: Software of the Mind.* Revised and Expanded. McGraw–Hill. New York. 1, A

Hovy, D. (2018, June). The social and the neural network: How to make natural language processing about people again. *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pp. 42–49.

Katarina, W., Barbara, J., & Roman, K. (2021). Human-computer interaction in translation activity: Fluency of machine translation. *Вестник Российского университета дружбы народов. Серия: Психология и педагогика*, 18(1), pp. 217–234.

Melve, A. (2023, September 15). Machine translation quality: How good is it and how to get started. *Weglot*. Dostupno preko: https://weglot.com/blog/machine-translation-quality/ [14.3.2023]

Munkova, D., Hajek, P., Munk, M., & Skalka, J. (2020). Evaluation of machine translation quality through the metrics of error rate and accuracy. *Procedia Computer Science*, 171, pp. 1327–1336.

Nepoznati autor (2022, July). The State of Machine Translation for Websites: A Comparative Study of the Top Machine Translation Engines. *Weglot & Nimdzi*. Dostupno preko: https://26501464.fs1.hubspotusercontent-eu1.net/hubfs/26501464/the-state-of-machine-translation-for-websites-report.pdf [14.3.2023]

Nepoznati autor (2022, May 12). Google Translate adds 24 new languages. *BBC News: Technology.* Dostupno preko : https://www.bbc.com/news/technology-61416757 [14.3.2023]

Nguyen, D., Doğruöz, A. S., Rosé, C. P., & De Jong, F. (2016). Computational sociolinguistics: A survey. *Computational linguistics*, 42(3), pp. 537–593.

Olive, J., Christianson, C., & McCary, J. (Eds.). (2011). *Handbook of natural language processing and machine translation: DARPA global autonomous language exploitation*. Springer Science & Business Media.

Paolillo, J. (2005). Language diversity on the internet: Examining linguistic bias. In UNESCO Institute for Statistics (Ed.), *Measuring linguistic diversity on the Internet* (pp. 43–89). Paris: UNESCO.

Petrosyan, A. (2023). Common languages used for web content. *Statista*. Dostupno preko: https://www.statista.com/statistics/262946/most-common-languages-on-the-internet/ [14.3.2023]

Pousada, A. (2019). The Role of the Internet in Language Preservation and Revitalization. *15to Congreso Puertorriqueño de Investigacionen la Educación, University of Puerto Rico, Río Piedras.*

Pozo, C. & Appasamy, Y. (2018). Language Justice. *Whose Knowledge?* Dostupno preko: https://whoseknowledge.org/ [14.3.2023]

Prahl, B., & Petzolt, S. (1997). *Translation problems and translation strategies involved in human and machine translation: Empirical studies*. Universitat Hildesheim Institut fur Angewandte Sprachwissenschaft Computerlinguistik.

Reyes, A., Rosso, P., & Veale, T. (2013). A multidimensional approach for detecting irony in twitter. *Language resources and evaluation*, 47, pp. 239–268.

Riney, T. J. (1998). Pre-colonial systems of writing and post-colonial languages of publication. *Journal of Multilingual and Multicultural Development*, 19(1), 64–83.

Routley, N. (2022). The Biggest Tech Talent Hubs in the U.S. and Canada. *Visual Capitalist*. Dostupno preko: https://www.visualcapitalist.com/biggest-tech-talent-hubs-in-us-and-canada/ [25.4.2023]

Sheshrao Chavhan, V. (2016). Globalization and its Impact on the Cultural Diversity. *The International Journal of Humanities and Social Science Research*.

Sinha, A. (2020, December 31). Global Civil Society Coalition launches website to promote Access to Knowledge. *Centre for Internet & Society*. Dostupno preko: https://cis-india.org/a2k/blogs/global-civil-society-coalition-launches-website-to-promote-access-to-knowledge [14.3.2023]

Tull, D. M. (2018). The limits and unintended consequencesof UN peace enforcement: the Force Intervention Brigade in the DR Congo. *International Peacekeeping*, 25(2), pp. 167–190.

Van Parijs, P. (2011). *Linguistic Justice for Europe and for the World*. Oxford University Press.

Wang, X., Broder, A., Gabrilovich, E., Josifovski, V., &Pang, B. (2009, February). Cross-language query classification using web search for exogenous knowledge. In: *Proceedings of the Second ACM International Conference on Web Search and Data Mining* (pp. 74–83).

Weiss, E.H. (2005). *The elements of international English style*. Armonk, NY: M.E. Sharpe.

Woolard, K. A. (1985). Language variation and cultural hegemony: Toward an integration of sociolinguistic and social theory. *American ethnologist*, 12(4), pp. 738–748.

Yimakan, H. (2006). *Living Heritage and mother languages*. United Nations Educational, Scientific and Cultural Organization – UNESCO. Dostupno preko: https://ich.unesco.org/en/ich-and-mother-languages-00555 [25.4.2023].

NEVENA N. MANIĆ

THE ROLE OF LANGUAGE IDEOLOGIES IN DIGITAL SPACES: A CASE STUDY OF ENGLISH LANGUAGE SUPREMACY IN ONLINE TRANSLATION

Summary: This paper explores the impact of language ideologies on social hierarchy in digital spaces, with a particular focus on the dominance of the English language in online translation. Drawing on existing literature and case studies, the paper argues that language ideologies play a significant role in shaping power relations and linguistic diversity in digital spaces. The author introduces the concept of Machine Translation (MT) systems as a potential solution for ethnic minorities to preserve their languages in the digital space. The paper also highlights the importance of linguistic diversity in online research and the need for researchers to adopt a more inclusive approach to language use. The findings of this paper have important implications for policymakers, educators, and researchers who are interested in promoting linguistic diversity and social justice in digital spaces.

Key words: Language ideologies, Digital spaces, English language supremacy, Machine Translation (MT), Language preservation

Datum prijema: 1.9.2023.

Datum ispravki: 15.11.2023.; 30.11.2023.; 5.12.2023.

Datum odobrenja: 5.12..2023.