

UDK 811.163.41'373:004

Nikola Ljubešić

Institut Jožef Stefan

nikola.ljubestic@ijs.si

Maja Miličević Petrović

Sveučilište u Beogradu

m.milicevic@fil.bg.ac.rs

Tanja Samardžić

Sveučilište u Zürichu

tanja.samardzic@uzh.ch

JEZIČNA AKOMODACIJA NA TWITTERU: PRIMJER SRBIJE

U ovom radu istražujemo fenomen jezične akomodacije kod srpskih korisnika Twittera analizirajući geokodirane poruke objavljene u razdoblju između 2013. i 2016. godine na području Bosne i Hercegovine, Crne Gore, Hrvatske i Srbije. Jezičnu produkciju korisnika Twittera opisujemo s pomoću 16 varijabli za koje je poznato da variraju među govornicima policentričnog makrojezika BCMS. Uspoređujemo jezičnu produkciju mobilnih srpskih korisnika Twittera s produkcijom nemobilnih srpskih korisnika, kao i produkciju mobilnih korisnika u Srbiji i izvan nje. Dok prva analiza djelomično podržava teoriju akomodacije, druga analiza ne daje nikakve naznake tog fenomena.

Ključne riječi: društvene mreže, geokodirane poruke, BCMS

In this paper we investigate the phenomenon of linguistic accommodation among Serbian Twitter users by analysing geo-encoded Twitter messages published between 2013 and 2016 in the area of Bosnia, Croatia, Montenegro and Serbia. We describe the linguistic production of Twitter users via 16 variables that are known to vary among the speakers of the pluricentric BCMS language. We compare that production of mobile Serbian Twitter users to that of non-mobile Serbian Twitter users, and by comparing the mobile users' language production inside and outside Serbia. While the first analysis shows support for accommodation, the second analysis yields no signal for that phenomenon.

Keywords: social media, geo-encoded messages, BCMS

1 Uvod

Poznato je da jezični izrazi variraju ovisno o kontekstu: govornici odabiru različite izraze ovisno o tome komuniciraju li s prijateljima, kolegama s posla ili nekim koga su tek upoznali. Posebna vrsta takve varijabilnosti, zvana jezična akomodacija, pojavljuje se kad promijenimo način izražavanja da bismo ga približili varijanti koju prepoznajemo kao drugačiju od naše, npr. promijenimo naglasak, dijalekt ili čak jezik. Ta se promjena obično događa u prisutnosti neke druge jezične varijante i može biti namjerna, ali i automatska. O njezinu djelovanju i mehanizmima naširoko se raspravljalo u literaturi

iz područja na raskrižju lingvistike, sociologije i psihologije (Giles 1991, Pickering i Garrod 2004).

Fenomen jezične akomodacije proučavao se u različitim okruženjima, poput regionalnih varijanti engleskoga kao što su kanadski i britanski engleski (Chambers 1992), švicarskih njemačkih dijalekata (Ruch 2018), varijanti španjolskoga (Erker i Otheguy 2016), češkoga (Wilson 2011) te drugih jezika. Sveobuhvatan pregled istraživanja ove teme Ruch i de Benito Moreno (u tisku) ukazuje na to da su kratkoročni učinci jezične akomodacije jasno utvrđeni, dok su njezini dugoročni učinci i uloga u općenitim promjenama u jeziku usprkos mnogim raspravama i dalje pretežito nepoznati. Kratkotrajni učinci većinom se istražuju u kontroliranim eksperimentalnim okruženjima, a nedostatni dokazi o dugoročnim učincima najčešće proizlaze iz opservacijskih podataka prikupljenih uz malo kontrole nad kontekstom u kojem su nastali.

Prema Ruch i de Benito Moreno (u tisku), prevladavajuće je gledište na jezičnu akomodaciju funkcionalno, s fokusom na ulogu tog fenomena u uspostavljanju društvenih odnosa. Sam se mehanizam manje proučavao, iako se raspravljalo o njegovim potencijalnim vezama s poznatim mehanizmima poput udešavanja (engl. priming) i oponašanja (engl. imitation). Osim toga, istraživanja jezične akomodacije često se usredotočuju na fonetske pojave, tj. promjene u izgovoru glasova, dok su druge lingvističke značajke, a pogotovo gramatika, manje ispitivane.

Glavna tema ovog rada jezična je akomodacija u porukama na Twitteru koje su objavili govornici policentričnog makrojezika koji uključuje bosanski, crnogorski, hrvatski i srpski (BCHS). Komponente ovog makrojezika čine četiri standardna jezika (od kojih svaki ima svoj ISO jezični kod), koji su utemeljeni na zajedničkom južnoslavenskom novoštokavskom dijalektu, koji je u prošlosti bio standardiziran kao dijasistem pod imenom srpskohrvatski jezik. BCHS je posebno zanimljiv za istraživanje jezične akomodacije zbog snažnih političkih sila te pitanja identiteta i kulturne raznolikosti koji su od velike važnosti za ovaj jezični prostor.

Naš je cilj izmjeriti stupanj jezične akomodacije kod govornika BCHS-a koristeći se podacima ekstrahiranim iz geokodiranih objava na Twitteru. Prvo je pitanje koje postavljamo prilagođavaju li govornici BCHS-a svoju jezičnu produkciju kada su u kontaktu s govornicima drugih jezika BCHS-a. Da bismo odgovorili na to pitanje, usredotočujemo se na Srbiju i uspoređujemo ostvarenje 16 lingvističkih značajki za koje je poznato da variraju među jezicima BCHS-a: (1) u porukama mobilnih i nemobilnih korisnika, tj. u porukama korisnika koji su u kontaktu nasuprot korisnika bez kontakta s drugim standardnim jezikom, i (2) u porukama mobilnih korisnika objavljenim unutar prebivališta (Srbija) nasuprot objava na području nekog od drugih standarda (Bosna i Hercegovina, Crna Gora, Hrvatska).

Razdoblje koje se može proučavati na temelju opažanja s Twittera može biti vrlo dugo, što omogućuje proučavanje dugoročnih učinaka na sistematičniji način nego što se to radilo dosad. Druga prednost korištenja podataka s Twittera činjenica je da

možemo upotrebljavati metode obrade prirodnog jezika za prikupljanje potvrda o raznim vrstama značajki iz spontane jezične produkcije, čime možemo proširiti znanje o samom mehanizmu akomodacije. Međutim, Twitter kao izvor podataka donosi i mnoga ograničenja, od kojih je najistaknutiji problem pitanje koliko su ti podaci reprezentativni za sve govornike na određenom području ili za određeni jezik, pa čak i za sveukupnu jezičnu produkciju određenih govornika.

2 Povezana istraživanja

Twitter (<https://twitter.com>) je društvena mreža na kojoj korisnici komuniciraju i objavljuju poruke zvane tvitovi (engl. tweets), čiji je sadržaj isprva bio ograničen na 140 znakova, ali je taj broj znakova nedavno udvostručen. Twitter je često korišten izvor informacija za različite vrste istraživanja zbog toga što ima dobru infrastrukturu za prikupljanje podataka (aplikacijsko programsko sučelje koje omogućuje automatsko prikupljanje podataka) i relativno permisivne licence za korištenje podataka u istraživanjima, pri čemu se uzimaju u obzir i pitanja privatnosti te vlasništva autora nad podacima. Važno je napomenuti da su podaci s Twittera također bogati metapodacima, a jedan određeni metapodatak od ključnog je značaja za naše istraživanje: geolokacija s koje je određeni tvit poslan, izražena u obliku geografske širine i dužine. Iako mnogi korisnici nevoljko daju svoju geolokaciju, s obzirom na sveukupnu količinu dostupnih podataka, moguće je prikupiti i prilično velike količine geokodiranih poruka.

Podaci s Twittera u prethodnim istraživanjima su se već koristili za proučavanje lingvističkih varijacija u odnosu na geografske čimbenike, pretežito u području računalne lingvistike. Kao što se može i očekivati, većina istraživanja bavi se (američkim) engleskim jezikom. Eisenstein, O'Connor, Smith i Xing (2010) predlažu model koji uči (s umjerenim uspjehom) povezivati određenu temu s određenim geografskom regijom. U jednom od kasnijih istraživanja, Doyle (2014) pokazuje da prostorna distribucija lingvističkih značajki ekstrahiranih s Twittera odgovara distribucijama prethodno utvrđenima s pomoću tradicionalnih dijalektoloških metoda. Eisenstein, O'Connor, Smith i Xing (2014) modeliraju prostornu raspršenost novih lingvističkih značajki kroz vrijeme, pokazujući da je ona pod snažnim utjecajem demografskih čimbenika. Problem neravnomjerne prostorne distribucije i nedostatka podataka u tim se istraživanjima rješava sofisticiranim statističkim modelima koji uključuju latentne varijable i razne transformacije originalnih pobrojavanja.

Povezana istraživanja koja se bave drugim jezicima prilično su rijetka. Gonçalves i Sánchez (2014) pokušavaju teritorijalno grupirati globalne varijetete španjolskog jezika, no umjesto toga nailaze na prevladavajuću podjelu između urbanih i ruralnih govornika. Scheffler, Gontrum, Wegel i Wendler (2014) pokušavaju pridružiti tvitove na njemačkom jeziku jednoj od sedam regija računajući vjerojatnost pojavljivanja riječi u određenoj regiji, ali ne uzimaju u obzir potencijalne varijacije u temi. U našem prethodnom istraživanju (Ljubešić i dr., u tisku), koje je najrelevantnije za ovaj rad, usredotočujemo se na BCHS i proučavamo prostornu distribuciju 16 lingvističkih značajki koje se spominju u gotovo svim radovima koji se bave razlikama unutar

BCHS-a. Rad pokazuje da za te značajke lingvističke granice u velikoj mjeri odgovaraju granicama među državama (pogotovo u slučaju Hrvatske i Srbije), no nikada u potpunosti. Rezultati tog istraživanja čine temelj nekih od glavnih pretpostavki u ovom radu (vidjeti odjeljak 3.3).

Kao što se vidi iz gornjeg prikaza, lingvistička istraživanja na podacima s Twittera uglavnom se bave geografskom rasprostranjenošću lingvističkih značajki, dok su pitanja poput jezične akomodacije mnogo manje istražena. Jedno istraživanje koje proučava akomodaciju na podacima s Twittera proveli su Danescu-Niculescu-Mizil i dr. (2011), koji ekstrahiraju razgovore s Twittera u svrhu praćenja stilističke akomodacije, pri čemu se ne uzima u obzir geografska rasprostranjenost.

U istraživanju prikazanom u ovom radu koristimo se geografskim podacima i usredotočujemo na pitanje akomodacije. Nastavljamo s povezanim nizom istraživanja na BCHS-u, stavljajući ovaj put u fokus činjenicu da je BCHS izuzetno zanimljiv slučaj za istraživanje akomodacije, s jedne strane zbog lingvističke bliskosti jezika, a s druge strane zbog snažnih političkih sila te pitanja identiteta i kulturne raznolikosti koja prevladavaju na ovom jezičnom prostoru. Budući da su političke sile posebno ojačale tijekom krvavog raspada Jugoslavije krajem prošlog stoljeća, zanimljivo je istražiti pojavljuje li se akomodacija usprkos njima u današnjoj komunikaciji na Twitteru.

3 Istraživanje

3.1 Istraživačka pitanja i hipoteze

Cilj je ovog istraživanja izmjeriti stupanj jezične akomodacije kod govornika BCHS-a koji prebivaju u Srbiji, a koji također putuju u druge države gdje se govore jezici BCHS-a. Na temelju saznanja iz literature i s obzirom na lingvističku bliskost jezika koji čine BCHS, može se očekivati da se akomodacija pojavljuje. Međutim, razni sociolingvistički čimbenici također bi mogli imati utjecaja u ovom još politički osjetljivom kontekstu.

Naše glavno istraživačko pitanje jest sljedeće: prilagođavaju li srpski korisnici Twittera svoj jezični izričaj kada su u dodiru s ostalim varijantama BCHS-a? Pokušavamo odgovoriti na to pitanje uspoređujući produkciju mobilnih srpskih korisnika Twittera s produkcijom nemobilnih srpskih korisnika te uspoređujući jezičnu produkciju mobilnih korisnika unutar i izvan Srbije. Pretpostavljamo da će odgovor na naše pitanje biti potvrđan te konkretno predviđamo (1) da će jezična produkcija mobilnih korisnika biti bliža drugim jezicima BCHS-a od produkcije nemobilnih korisnika te (2) da će među mobilnim korisnicima jezična produkcija biti bliža drugim jezicima BCHS-a u tvitovima koji su nastali izvan Srbije nego u tvitovima koji su nastali u Srbiji.

Mobilnost korisnika utvrđena je na temelju geolokacijskih podataka pridruženih korisnikovim porukama na Twitteru. Pri proučavanju jezične produkcije usredotočujemo se na 16 lingvističkih varijabli, koje su sve opsežno istražene u kontekstu

međuvarijetetnih razlika u BCHS-u (te su varijable opisane u odjeljku 3.3). U ovom se istraživanju koristimo istim početnim skupom podataka i oslanjamo na rezultate našeg prethodnog istraživanja (Ljubešić i dr., u tisku).

3.2 Podaci

Analize mobilnosti korisnika započinjemo s polaznim skupom podataka na jezicima BCHS-a koji je prikupljen s pomoću alata TweetCat (Ljubešić, Fišer i Erjavec 2014). Taj alat namijenjen je prikupljanju podataka s Twittera kod jezika s malim brojem govornika. U postupku prikupljanja podataka koji je trajao od lipnja 2013. do kraja 2016. prikupljeni su podaci od 70 107 korisnika koji su objavili sveukupno 38 726 488 tvitova. Za potrebe naših istraživanja zadržani su samo podaci geokodirani u BiH, Hrvatskoj, Crnoj Gori i Srbiji, čime je skup podataka sveden na 17 172 korisnika i 1 755 525 tvitova. Nakon što su relevantne varijable ekstrahirane iz geokodiranih tvitova, uklonjeni su tvitovi koji ne sadrže relevantne podatke ni za jednu varijablu, čime se skup podataka smanjio na 13 102 korisnika i 693 111 tvitova.

Za potrebe ovog istraživanja, također smo uklonili sve tvitove korisnika koji su imali manje od 50 tvitova, pretpostavljajući da procjenjivanje naših varijabli na manjem broju tvitova po korisniku ne bi bilo ni izbliza pouzdano. Osim toga, izbacili smo sve korisnike koji su objavljivali manje od dvije trećine tvitova u jednoj državi. Tu smo odluku donijeli zbog jedne od svojih glavnih pretpostavki, prema kojoj svaki korisnik prebiva u određenoj državi, koju određujemo kao državu iz koje su poslani barem dvije trećine korisnikovih geokodiranih tvitova.

Zadnja dva ograničenja smanjila su početni skup podataka na 489 295 tvitova koje je objavilo 3 083 korisnika. Distribucija odabranih korisnika po zemljama boravišta jest sljedeća: 2516 korisnika iz Srbije, 272 korisnika iz Crne Gore, 167 korisnika iz BiH i 128 korisnika iz Hrvatske. Vrlo neravnomjerna distribucija po državama glavni je razlog za ograničavanje našeg istraživanja na srpske korisnike Twittera (i na njihova 404 823 tvita): u trenutnom skupu podataka nema dovoljno podataka da bi se provele slične analize na korisnicima iz drugih država gdje se govori neki od jezika BCHS-a. Međutim, postupak prikupljanja iz kojeg je proizašao ovaj skup podataka i dalje je u tijeku, pa očekujemo da će u budućnosti biti moguće i te analize.

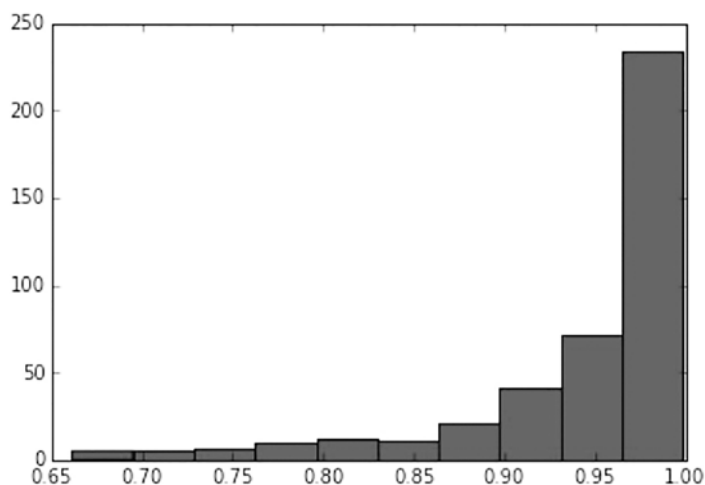
Mobilnost korisnika središnji je čimbenik u našem istraživanju jer pretpostavljamo da mobilnost vodi do kontakta s govornicima drugih jezika BCHS-a. Korisnika smatramo mobilnim ako u nekom trenutku tijekom našeg prikupljanja podataka objavljuje tvitove izvan svoje zemlje boravišta. Ako korisnik ne objavljuje izvan zemlje boravišta tijekom postupka prikupljanja podataka, smatramo ga nemobilnim. Tablica 1. pokazuje broj i postotak mobilnih korisnika za svaku od četiri države u našem skupu podataka. Srbija ima najmanji postotak mobilnih korisnika Twittera, a iza nje slijede Hrvatska, BiH i Crna Gora. Postoji više mogućih objašnjenja za te razlike, od kojih je jedno činjenica da su korisnici Twittera u Srbiji mlađi nego u ostalim državama. Međutim, za ovo istraživanje najvažnija je informacija to da je oko 17 % korisnika Twittera u

Srbiji mobilno. Treba napomenuti i da u ovom istraživanju ne pratimo kamo se mobilni korisnici kreću (u BiH, Hrvatsku ili Crnu Goru).

Zemlja boravišta	Broj mobilnih korisnika	Postotak među svim korisnicima
Bosna i Hercegovina	64	38,3 %
Hrvatska	38	29,7 %
Crna Gora	121	44,5 %
Srbija	417	16,6 %

Tablica 1: Broj i postotak mobilnih korisnika po zemlji boravišta.

Konačno, da bismo bolje razumjeli razinu mobilnosti korisnika u ovom istraživanju, na Slici 1 prikazujemo histogram postotka tvitova objavljenih unutar zemlje boravišta za srpske mobilne korisnike. Prisjetimo se da smo iz analize izbacili korisnike koji su objavili manje od dvije trećine tvitova u istoj državi, pa minimalna vrijednost na Slici 1 iznosi 66 %. Na slici se jasno vidi da većina mobilnih korisnika objavljuje većinu tvitova u svojoj zemlji boravišta, a samo mali broj tvitova izvan nje.



Slika 1: Distribucija postotka tvitova koje su mobilni korisnici objavili u Srbiji (na osi x prikazan je postotak, a na osi y broj korisnika koji odgovara određenom postotku).

3.3 Lingvističke varijable

Analize jezične akomodacije provodimo na 16 kategoričkih lingvističkih varijabli, od kojih svaka ima dvije moguće vrijednosti. Te su varijable opisane u Tablici 2 i ilustrirane primjerima iz početnog skupa podataka. U nastavku ovog pododjeljka varijable su i detaljnije opisane.

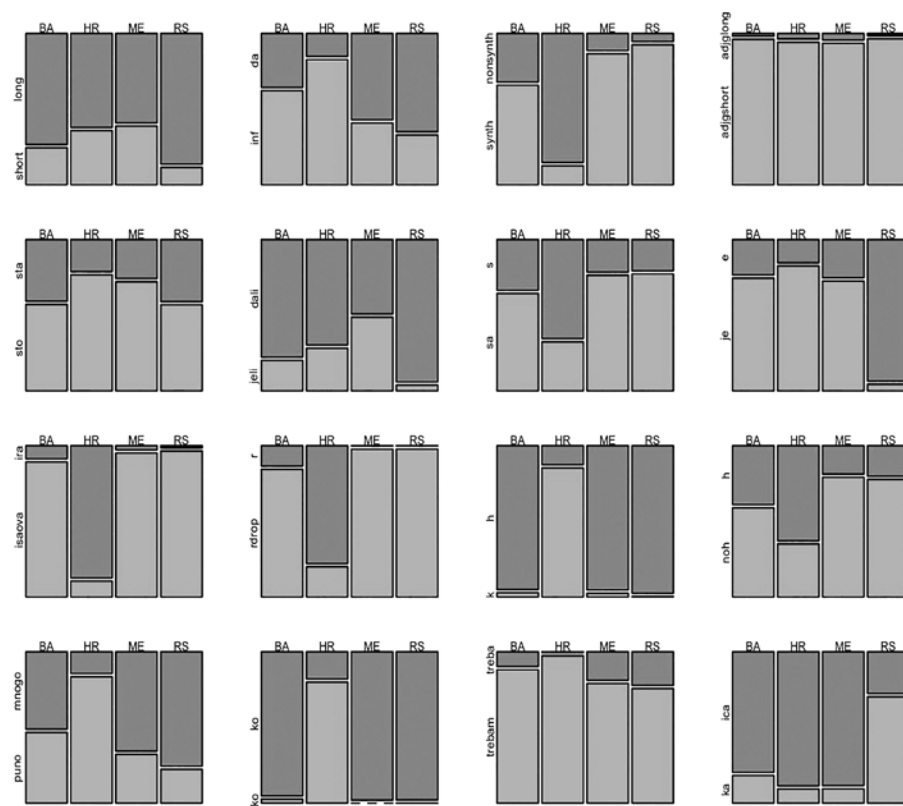
Vrsta varijable	Naziv varijable	Razine varijable i primjeri uporabe
Fonetska varijabla	e:je	e: <i>Ja sam uvek za sve kriva NARAVNO</i> je: <i>Uvijek mi je bio ritam suprotan od ljudi oko mene</i>
	rdrop:nodrop	rdrop: <i>@pop_Joil takodje! Neki hrono fazan pasti mora!!</i> nodrop: <i>@IKopric s najguscom, sjajnom kosom takodjer ;-)</i>
	k:h	k: <i>gledam 15 minuta #vb pocecu da pricam kao Ava IDEM SE OKUPAT JEST PA UCIT KEMIJU</i> h: <i>pobegnem s hemije, i da li naucim za sledeci cas? Nee</i>
	h:noh	h: <i>Kafa se nece sama skuhati. Pikaner se MORA nechim ugasiti!</i> noh: <i>A znam da ce se skuvati tokom noci i da ce sve ovo završiti na podu</i>
Leksička varijabla	sto:sta	<i>sto: Nestala struja baterija prazna, što ću da radim noćas kukala mi majka</i> <i>sta: Šta ovo puca u Makarskoj?</i>
	dali:jeli	dali: <i>Da li i vi cupate obrve na terasi da vas ceo komsiluk vidi?</i> jeli: <i>Je li hladno napolju?</i>
	s:sa	s: <i>Ljubi mi se s tobom, ali me mrzi da ustajem</i> sa: <i>Nisam u kontaktu sa bivšim devojkama, ko zna zašto je to dobro</i>
	mnogo:puno	mnogo: <i>@GaleMadafaka hahahaha vama je mnogo dosadnooo</i> puno: <i>Izgubio sam 0 followera ali opet nemam ih puno</i>
	ko:tko	ko: <i>@jelena_m_zmaj ali ne moze svako dobro da pise</i> tko: <i>@zvjer moze svatko, to nije upitno, vec, da li i ti?</i>

Morfo- sin- taktička varijabla	inflong:short	inflong: <i>Mama očekuje da cu napustiti moj dragi krevet i uciti hemiju...hah</i> short: <i>imam poriv teatralno napustit prostoriju. #fpzg</i>
	da:inf	da: <i>trenutci sa njim ne mogu da se mere ni sa jednim drugim</i> inf: <i>Kada je dovoljno mračno, mogu se vidjeti zvijezde! (Perzijska izreka)</i>
	synth:nosynth	synth: <i>Umrecu od gladi ako se rucak ne stvori odma sad na stolu</i> nosynth: <i>Nurija Begic Nuki UMRIJET CU ZA NJOM 2014Uzivo: http://t.co/XoJTI4tNdf</i>
	adjgshort:long	adjgshort: <i>mama odvrnula muziku od ranog jutra, nmg</i> long: <i>Sto volim da mi se dovuku mala deca od ranoga jutra</i>
	ira:isaova	ira: <i>Upravo sudjelujem na Kongresu o vanjskoj politici koji organizira @EPPWomen te zaklada @KASonline. http://t.co/QgL94sN5j3</i> isaova: <i>Zasto se ovako nesto organizuje dva dana pre prijemnog? (@ Gradski Trg - Spomenik Milutin w/ @mlazovic) http://t.co/hC6nhaugxh</i>
	treba:trebam	treba: <i>@BaneJevticBane ti si paranoik...a treba da ga se bojiš, al šta te bole</i> trebam: <i>@hrvatskitelekom stick podigne program, ali kad ga trebam vezati trazi postavke za hrvatski telekom</i>
	ica:ka	ica: <i>Profesorice mozete me pustit ranije moram na aerodrom?</i> ka: <i>Zamalo ženi kosu da spalim, mislila sam da je profesorka geografije hahaha</i>

Tablica 2: Lingvističke varijable korištene u istraživanju.

Središnji kriterij pri izboru varijabli bila je lingvistička relevantnost, koja je utvrđena pregledom literature (uključujući sljedeće radove: Pešikan, Jerković i Pižurica, 2010, Barić, Lončarić, Malić, Pavešić, Peti, Zečević i Znika, 1997, Halilović, 2004, Čirgić, Pranjeković i Silić, 2010, Tošović, 2008 i druge). Također smo uzeli u obzir mogućnost automatske ekstrakcije i dovoljnu zastupljenost u skupu podataka. Postupak ekstrakcije u većini se slučajeva temeljio na posebnim leksikonskim datotekama, tj. popisima oblika riječi koji su relevantni za istraživanje. U slučajevima u kojima je ta metoda bila neprimjenjiva koristili smo se regularnim izrazima.

U prethodnom istraživanju (Ljubešić i dr., u tisku) već smo utvrdili koje su vrijednosti dominantne u kojoj od četiriju država koje proučavamo (BiH, Hrvatska, Crna Gora i Srbija). U ostatku ovog odjeljka ukratko opisujemo logiku odabira varijabli i prostorne distribucije njihovih vrijednosti (prikazano na Slici 2, uzeto iz Ljubešić i dr., u tisku). Počinjemo s fonetskim varijablama, nakon čega opisujemo leksičke varijable te završavamo s morfosintaktičkima.



Slika 2: Distribucija vrijednosti 16 proučanih varijabli po državama (BA = BiH, HR = Hrvatska, ME = Crna Gora, RS = Srbija).

Varijabla **e:je** odnosi se na dva fonetska refleksa praslavenskog suglasnika jata, *e* (npr. u riječima »mleko« i »pjesma«) i (*i*)*je* (»mlijek«, »pjesma«). Kao što se vidi na Slici 2 (kraj drugog reda), *e* je vrlo dominantan u porukama na Twitteru objavljenima u Srbiji, dok je *je* tipičan za ostale tri zemlje. Varijabla **rdrop:nodrop** motivirana je činjenicom da se u nekim riječima jezika BCHS-a *r* na kraju riječi može pojaviti ili ispustiti; prva je opcija dominantna u hrvatskom jeziku, a druga u srpskom, bosanskom i crnogorskom. Neke od riječi koje promatramo su »juče(r)«, »veče(r)« i »takođe(r)«. Alternacija **k:h** događa se na počecima riječi grčkog podrijetla koje su počinjale slovom

hi (χ), pa imamo parove riječi poput »kemija«/»hemija« i »kirurg«/»hirurg«. U hrvatskom se jeziku dosljedno koristi slovo *k*, dok se u srpskom, bosanskom i crnogorskom koristi *h*. Posljednja fonetska varijabla povezana je s prisutnošću ili odsutnošću slova *h* (**h:noh**). Konkretnije, slovo *h* ponekad se izostavlja na početku riječi, a ponekad se unutar riječi zamjenjuje drugim slovom (najčešće *j* ili *v*). Neki od primjera ove alternacije sljedeći su: »hrđa«/»rđa«, »snaha«/»snaja« i »gluh«/»gluv«. Opcije koje u sebi sadrže *h* tipične su za hrvatski, a opcije bez *h* (*noh*) pojavljuju se u srpskom i crnogorskom, dok bosanski više naginje riječima bez *h*.

Što se tiče leksičkih varijabli, **sta:sto** odnosi se na standardni oblik upitne zamjenice »što«, koja u hrvatskom, bosanskom i crnogorskom glasi »što«, a u srpskom šta« (u srpskoj literaturi »što« se također navodi, no vrlo rijetko). Iako se ispostavlja da je »što« dominantna vrijednost u svim državama (najvjerojatnije zbog primjera korištenja te riječi kao odnosne zamjenice i skraćenog oblika riječi »zašto«, koje u postupku ekstrakcije nismo mogli zaobići), vidljivo je da se »šta« više koristi u Srbiji i BiH nego u Hrvatskoj i Crnoj Gori. Varijabla **dali:jeli** odnosi se na skup glagola »biti« i čestice »li« (»je li«) i skup čestica »da li«, koji se upotrebljavaju u upitnim rečenicama. Podaci s Twittera pokazuju da je oblik »da li« dominantan u svim državama, pri čemu se »je li« najčešće koristi u Crnoj Gori, iako je prema standardima u Bosni i Hrvatskoj propisan oblik »je li«. Sljedeća varijabla, **s:sa**, odnosi se na pisanje prijedloga »s« i »sa«. U standardnom hrvatskom jeziku, izbor između »s« i »sa« ovisi o fonetskim pravilima – »sa« se treba koristiti ispred glasova »s«, »š«, »z« i »ž«; ispred suglasničkih skupina »ks« i »ps« te ispred instrumentala zamjenice »ja« (»sa mnom«), dok se »s« koristi u svim drugim slučajevima. U standardnom srpskom jeziku izbor se najčešće prepušta govornicima. Podaci s Twittera potvrđuju da je »s« dominantna vrijednost u hrvatskom, a »sa« u ostala tri jezika. Upitna zamjenica »ko« u srpskom, bosanskom i crnogorskom te upitna zamjenica »tko« u hrvatskom čine varijablu **ko:tko**. Budući da je »ko« homonimna riječ, usredotočujemo se samo na izvedene oblike »ni(t)ko« i »i(t)ko«, čija distribucija odgovara očekivanjima – »tko« je dominantna vrijednost u Hrvatskoj, a »ko« u ostalim državama. Zadnja leksička varijabla, **mnogo:puno**, tiče se priloga količine »mnogo« i »puno«. Oba se oblika koriste u svim varijantama BCHS-a, ali je oblik »puno« osobito karakterističan za hrvatski, a »mnogo« za srpski, crnogorski i do određene mjere bosanski.

Prva morfosintaktička varijabla, **inflong:short** povezana je s infinitivnim oblicima. Puni infinitivni oblik glagola u svim jezicima BCHS-a završava na -ti (»čitati«) ili -ći (»doći«), ali je u nekim varijantama, a pogotovo u hrvatskom jeziku, uobičajeno skraćivanje infinitiva izbacivanjem slova »i« na kraju riječi (»čitat«, »doć«). Kao što je prikazano na Slici 2, puni oblik infinitiva dominantan je u sva četiri jezika, ali se krnji infinitiv rjeđe koristi u srpskom nego u drugim jezicima. Sljedeća varijabla, **synth:nosynth**, odnosi se na činjenicu da futur prvi u srpskom jeziku većinom ima sintetički oblik, što znači da je pomoćni glagol »htjeti« spojen s glavnim glagolom (npr. »čitacu«), dok se u hrvatskom koristi analitički oblik, tj. infinitiv glagola i pomoćni glagol dvije su odvojene riječi (»čitat ću«). Sintetički oblici dominiraju u crnogorskom i (u manjoj mjeri) u bosanskom. Varijabla **adjgshort:long** povezana je s

činjenicom da se pri sklonidbi pridjeva u BCHS-u ponekad može dodati samoglasnik na kraj riječi radi lakšeg izgovora i/ili stilske obilježnosti. Najbolji je primjer ove pojave dodavanje nastavka -a u muškom rodu jednine pridjeva (npr. »novoga«), što se češće radi u standardnom hrvatskom nego u standardnom srpskom jeziku. Varijabla **ira:isaova** odnosi se na posuđenice izvedene od međunarodnih glagola: u hrvatskom se uglavnom koristi glagolski sufiks -ira (npr. »promovirati« i »registrirati«), dok sufiksi -isa i -ova prevladavaju u srpskom jeziku (»promovisati«, »registrovati«). Sufiksi -isa i -ova prevladavaju i u crnogorskom i bosanskom. Varijabla **ica:ka** tiče se sufiksa koji se upotrebljavaju u tvorbi imenica koje označavaju ženske vršitelje radnje i koje se djelomično preklapaju, a djelomično razlikuju u jezicima BCHS-a. Sufiks -ica (npr. u riječi »nastavnica«) prisutan je u svim jezicima, ali je dominantan samo u hrvatskom i bosanskom, dok je u srpskom sufiks -ka (»čitateljka«) također vrlo učestao. Međuvarijetetne razlike između sufiksa -ica i -ka najčešće se pojavljuju u riječima čija se osnova završava na -r (npr. »profesorica«/»profesorka«), zbog čega smo promatrali samo riječi koje završavaju s -rica i -rka te uočili da je nastavak -rka dominantan u Srbiji, a nastavak -rica u ostale tri zemlje.

Varijabla **treba:trebam** odnosi se na činjenicu da se modalni glagol »trebati« u srpskom jeziku često upotrebljava impersonalno. Razlog tome je preskriptivna tradicija koja zabranjuje konstrukcije poput »trebam da idem« i zahtijeva oblik »treba da idem«. U hrvatskom se jeziku personalni oblici normalno koriste te nakon njih dolazi infinitiv (npr. »trebam ići«). Podaci s Twittera pokazuju da su personalni oblici zapravo dominantni posvuda, ali se u srpskom impersonalniji oblik »treba« koristi više nego u drugim jezicima. Posljednja je sintaktička značajka koju promatramo sastav nekih kompleksnijih predikata, koji u srpskom najčešće imaju dopunu »da« + prezent glagola (npr. »volim da čitam«), dok se u hrvatskom obično koristi infinitiv glagola (»volim čitati«). Ta je razlika izražena u varijabli **da:inf**, kojom se utvrdila dominantnost oblika s »da« u srpskom i crnogorskom te dominantnost oblika s infinitivima u hrvatskom i bosanskom.

U smislu navedenih značajki, akomodaciju opažamo kao smanjenje dominantnosti, tj. frekventnosti dominantne verzije u Srbiji kao rezultat akomodacije ostalim jezicima (bosanskom, hrvatskom i crnogorskom). Napominjemo da je kod dvije varijable, sto:sta i treba:trebam, dominantna razina utvrđena iz podataka s Twittera različita od one koja se navodi u literaturi (barem djelomično zbog ograničenja našeg automatskog postupka ekstrakcije). U svrhu ove analize (uspoređivanja uporabe određenih vrijednosti varijabli kod mobilnih i nemobilnih korisnika), svejedno smo se usredotočili na nedominantne oblike »šta« i »treba«.

3.4 Analize

Sve analize provodimo kako bismo pronašli odgovor na svoje glavno istraživačko pitanje: provode li srpski korisnici Twittera akomodaciju u svojem jeziku kada su u kontaktu s drugim varijantama BCHS-a. Podatke promatramo iz dvije perspektive: prvo *iz perspektive korisnika* uspoređujemo jezičnu produkciju mobilnih korisnika

(koji objavljuju tvitove i iz Srbije i iz drugih država gdje se govore jezici BCHS-a) s produkcijom nemobilnih korisnika (koji objavljuju tvitove samo iz Srbije). Nakon toga proučavamo situaciju iz *perspektive tvitova* koje objavljuju mobilni korisnici tako što uspoređujemo jezičnu produkciju mobilnih korisnika unutar i izvan Srbije.

Odredimo prvo detaljnije hipoteze opisane u odjeljku 3.1: kod perspektive usmjerene na korisnike očekujemo da će jezična produkcija mobilnih korisnika s obzirom na naše odabrane varijable biti bliža drugim jezicima BCHS-a od produkcije nemobilnih korisnika (tj. da mobilni korisnici manje upotrebljavaju vrijednosti koje su dominantne u Srbiji od nemobilnih korisnika). Kod perspektive usmjerene na tvitove očekujemo da će jezična produkcija među mobilnim korisnicima s obzirom na naše odabrane varijable biti bliža drugim jezicima BCHS-a u tvitovima koji su nastali izvan Srbije nego u tvitovima koji su nastali u Srbiji (tj. da se vrijednosti dominantne u Srbiji manje upotrebljavaju u tvitovima koji su objavljeni izvan Srbije nego u onima koji su objavljeni u Srbiji). Prvu hipotezu ispitujemo na neuparenim uzorcima, tj. na mjerenjima za svaku od 16 varijabli za mobilne i nemobilne korisnike. U testiranju druge hipoteze koristimo se uparenim uzorkom, tj. mjerenjima za svaku od 16 varijabli za svakog korisnika u slučajevima kada on objavljuje tvitove izvan i unutar Srbije.

Mjerenja koja ekstrahiramo iz podataka pokazuju koliko su zastupljene vrijednosti koje su tipičnije za srpski jezik u pojedinoj varijabli. Uzmimo sljedeći primjer iz perspektive usmjerene na korisnike: ako se u 234 od 563 tvita koje je korisnik objavio pojavljuje varijabla e:je (tj. ako ti tvitovi sadrže barem jednu riječ u kojoj je moguća alternacija e:je, npr. »mleko«), a dominantni ekavski oblik (vrijednost e) pojavljuje se u 224 tvita, računamo omjer kao $224 / 234 = 0,957$, što znači da taj određeni korisnik upotrebljava ekavsku varijantu u 95,7 % od svih relevantnih slučajeva. Kod perspektive usmjerene na tvitove, omjer računamo odvojeno za tvitove koje je korisnik objavio u Srbiji i one koje je objavio u drugoj državi gdje se govori neki jezik BCHS-a.

Nakon što izračunamo omjere za sve varijable za pojedinačne mobilne i nemobilne korisnike, kao i za tvitove koje su pojedinačni mobilni korisnici objavili u Srbiji i izvan nje, računamo omjere po skupinama korisnika (mobilni i nemobilni korisnici, 1. hipoteza) i vrstama tvitova (tvitovi objavljeni u Srbiji i izvan nje, 2. hipoteza) te ispitujemo postavljene hipoteze statističkim metodama. U prvoj analizi koja uključuje dva nezavisna uzorka primjenjujemo niz Wicoxonovih testova sume rangova (jedan za svaku lingvističku varijablu), dok za drugu analizu, provedenu na uparenim uzorcima, primjenjujemo Wilcoxonov test rangova sa predznacima za zavisne uzorke. Upotrebljavamo neparametrijske testove zbog toga što naši podaci nisu normalno distribuirani. Međutim, u rezultatima prikazujemo (parametrijske) prosjeke umjesto (neparametrijskih) medijana kako bismo pružili jasniji uvid u slučajeve gdje su razlike male i ne bi bile vidljive u medijanima.

Varijabla (dominantna vrijednost ispisana podebljanim slovima)	Broj nemobilnih korisnika	Broj mobilnih korisnika	Omjer za nemobilne korisnike	Omjer za mobilne korisnike	Razlika između omjera	Statistika W	p-vrijednost	CLES
ica:ka	416	70	0,79	0,657	0,133	12 547	0,012	0,569
e:je	2099	417	0,976	0,887	0,089	31 8370	< 0,001	0,636
inflong:short	2091	414	0,883	0,84	0,043	368 840	< 0,001	0,574
dali:jeji	1672	335	0,97	0,927	0,043	258 040	< 0,001	0,539
mnogo:puno	1690	330	0,773	0,738	0,036	263 640	0,09	---
da:inf	2098	417	0,664	0,63	0,034	375 130	< 0,001	0,571
treba:trebam	1947	379	0,227	0,201	0,026	345 040	0,040	0,532
s:sa	2096	417	0,813	0,79	0,023	404 430	0,016	0,537
h:noh	1926	380	0,776	0,761	0,015	355 360	0,352	---
ira:isaova	1612	330	0,987	0,977	0,010	257 760	0,012	0,515
synth:nosynth	1915	379	0,947	0,94	0,007	348 930	0,112	---
adjshort:long	2025	410	0,986	0,98	0,006	403 720	0,092	---
rdrop:modrop	1670	350	0,998	0,995	0,003	291 310	0,478	---
ko:tko	1431	293	1,0	1,0	0,0	209 930	0,523	---
k:h	838	169	0,996	1,0	-0,004	71 318	0,271	---
sto:sta	1225	260	0,424	0,432	-0,008	158 730	0,934	---

Tablica 3: Rezultati usporedbi nemobilnih i mobilnih korisnika.

Varijabla (dominantna vrijednost ispisana podebljanim slovima)	Broj uparenih podatkovnih točaka	Omjer unutar Srbije	Omjer izvan Srbije	Razlika između omjera	Statistika V	p-vrijednost	CLES
mnogo :puno	49	0,769	0,662	0,107	327	0,123	---
e :je	354	0,879	0,818	0,062	18 658	0,776	---
s :sa	212	0,788	0,738	0,05	8574,5	0,455	---
da :inf	255	0,637	0,613	0,025	16 699	0,666	---
synth :nosynth	62	0,959	0,941	0,018	90,0	0,862	---
adjshort :long	100	0,98	0,972	0,008	61,0	0,737	---
inflong :short	168	0,824	0,823	0,002	3857,5	0,018	0,312
dali :jeli	33	0,782	0,782	0,0	40,5	0,752	---
rdrop :nodrop	39	1,0	1,0	0,0	0,0	Nije primjenjivo	---
k :h	6	1,0	1,0	0,0	0,0	Nije primjenjivo	---
ko :tko	17	1,0	1,0	0,0	0,0	Nije primjenjivo	---
ica :ka	1	1,0	1,0	0,0	0,0	Nije primjenjivo	---
h :noh	66	0,776	0,779	-0,003	505,5	0,706	---
ira :isaova	47	0,996	1,0	-0,004	0,0	0,371	---
treba :trebam	73	0,204	0,209	-0,005	677,5	0,703	---
sto :sta	84	0,389	0,399	-0,01	1369	0,923	---

Tablica 4: Rezultati usporedbe tvitova koje su mobilni korisnici objavili u Srbiji i onih koje su objavili izvan Srbije.

Konačno, za statistički značajne razlike računamo i veličinu učinka metodom Common Language Effect Size (CLES). Općenito, veličine učinka kvantifikacijski su način prikazivanja praktične važnosti rezultata, a u društvenim i humanističkim znanostima raste njihova primjena u istraživanjima uz *p*-vrijednost o statističkoj značajnosti razlike (koja kvantificira vjerojatnost da je izmjerena razlika slučajna). CLES je vrlo jednostavna i intuitivna mjera kojom je kodirana vjerojatnost dobivanja veće vrijednosti za element iz prvog uzorka nego za element iz drugog uzorka ako se nasumično odabire par elemenata, po jedan iz svakog uzorka. CLES vrijednost od 0,5 označava razinu slučajnosti (vjerojatnost dobivanja veće vrijednosti iz prvog uzorka je 50 %). Ako je vrijednost CLES-a manja od 0,5, to znači da postoji veća vjerojatnost dobivanja veće vrijednosti iz drugog uzorka nego iz prvog uzorka.

3.5 Rezultati

Za svaku od dvije skupine analiza prikazujemo sljedeće podatke: naziv varijable, broj(eve) relevantnih opažanja (oni se razlikuju od varijable do varijable), relevantne omjere dominantne vrijednosti varijable (po skupini korisnika ili lokaciji objavljivanja tvita), razliku između dvaju omjera, statistiku testa, podatak o statističkoj značajnosti neparametrijskog testa te mjeru veličine učinka (gdje je to primjenjivo).

Prvo prikazujemo rezultate analize usmjerene na korisnike s fokusom na našu hipotezu da je jezična produkcija mobilnih korisnika bliža drugim jezicima BCHS-a nego jezična produkcija nemobilnih korisnika. U Tablici 3 prikazani su rezultati te analize, s varijablama poredanima u padajućem nizu na temelju razlike između omjera za nemobilne i mobilne korisnike.

Na primjeru jedne od najistaknutijih varijabli, *e*:je, objasniti ćemo kako prikazane vrijednosti treba čitati. Vrijednost ove varijable koja je dominantna u Srbiji jest *e* (ekavski oblik, npr. u riječi »mleko«). Međutim, uz pretpostavku prisutnosti jezične akomodacije (zbog toga što se u ostale tri države koristi ijekavica), očekujemo da će omjer vrijednosti *e* biti manji kod mobilnih nego kod nemobilnih korisnika. Bilježimo da postoji 2099 nemobilnih i 417 mobilnih korisnika za koje imamo mjerenja za ovu varijablu. To znači da možemo izmjeriti varijablu *e*:je za svih 2516 korisnika u našem skupu podataka, što nije čest slučaj (npr. za varijablu *ica:ka* samo smo od $416 + 70 = 486$ korisnika dobili relevantne podatke). Nakon toga prikazujemo omjer uporabe vrijednosti varijable *e* kod mobilnih i nemobilnih korisnika: nemobilni korisnici upotrebljavaju ekavske oblike u 97,6 % slučajeva, dok je taj broj za mobilne korisnike 88,7 %. U sljedećoj je koloni u tablici prikazana razlika između tih dvaju omjera ($0,976 - 0,887 = 0,089$). Potom bilježimo rezultat Wilcoxonova testa sume rangova kojim ispituje nultu hipotezu da se ta dva uzorka (omjeri uporabe vrijednosti varijable *e* kod nemobilnih i mobilnih korisnika) ne razlikuju. Budući da je vjerojatnost toga vrlo mala ($p < 0,001$), možemo odbaciti nultu hipotezu i zaključiti da postoji značajna razlika između nemobilnih i mobilnih korisnika po pitanju ove varijable. Na kraju prikazujemo mjeru veličine učinka, tj. CLES, koja iznosi 0,636 i govori nam da će, ako odaberemo jednog nasumičnog nemobilnog i jednog nasumičnog mobilnog korisnika, postojati vjerojatnost

od 63,6 % da nemobilni korisnik ima veći omjer uporabe vrijednosti varijable e nego mobilni korisnik.

Kao što je vidljivo u Tablici 3, značajne razlike između nemobilnih i mobilnih korisnika pronađene su za polovicu (osam) varijabli, svugdje s osrednjim veličinama učinka u rasponu od 0,515 do 0,636. Činjenicu da su značajne razlike pronađene baš za te konkretne varijable ne može se objasniti na temelju njihove lingvističke vrste zbog toga što su sva tri tipa varijabli (fonetski, leksički i morfosintaktički) jednako zastupljena. Međutim, vidljiva je određena doza pravilnosti u prostornoj distribuciji njihovih vrijednosti u državama u kojima se govore jezici BCHS-a. Naime, kod ovih varijabli dominantna vrijednost varijable u Srbiji razlikuje se od dominantnih vrijednosti varijabli u svim trima ili barem dvjema državama BCHS-a (obično u Hrvatskoj i Bosni) (vidjeti Sliku 2). Razlika je posebno očita kod varijabli e:je i ica:ka, koje imaju najveću razliku između omjera, a slijedi ih varijabla da:inf. U nekoliko slučajeva razlika ne postoji u dominantnoj vrijednosti, nego u mjeri u kojoj se u srpskom koristi određena varijabla u usporedbi s jezicima drugih država (inflong:short, treba:trebam, dali:jeli). S druge strane, »neznačajne« varijable većinom su one u kojima srpski dijeli dominantnu vrijednost s crnogorskim i bosanskim (za razliku od hrvatskog, npr. ko:tko, k:h, rdop:nordrop) ili one za koje se ne mogu vidjeti jasne teritorijalne razlike (npr. sto:sta). Za te bi se varijable trebalo provesti dodatno istraživanje u kojem bi se proučavala mobilnost korisnika u određene države i time izbjegla poništavanja između država koja se sada vjerojatno događaju.

Nastavljamo s drugom analizom, u kojoj se usredotočujemo na hipotezu da će tvitovi koje su mobilni korisnici objavili izvan Srbije biti bliskiji drugim jezicima BCHS-a od tvitova koje su isti korisnici objavili u Srbiji. Rezultati te analize prikazani su u Tablici 4. Tablica je organizirana na sličan način kao i Tablica 3, s razlikom da je u njoj prikazan samo jedan broj korisnika, tj. broj korisnika za koje imamo mjere varijabli i u Srbiji i izvan Srbije te da su omjeri upareni. Podsjetimo da u ovom slučaju provodimo neparametrijski Wilcoxonov test rangova sa predznacima za zavisne uzorke da bismo ispitali nultu hipotezu prema kojoj dva uzorka dolaze iz iste distribucije, tj. razlika mjera u uparenim uzorcima ima simetričnu distribuciju oko nule.

Rezultati druge analize ne podupiru hipotezu da postoji akomodacija. Jedina značajna razlika dobivena je za varijablu inflong:short, ali radi se o divergenciji, a ne konvergenciji. Naime, dominantna vrijednost dugog infinitiva upotrebljava se manje u tvitovima objavljenima u Srbiji (iako to nije vidljivo iz omjera, očito je iz vrijednosti CLES-a, koja iznosi 0,312, dakle manje od 0,5). Po pitanju deskriptivnih indikatora, sedam varijabli daje podatak očekivanog predznaka (pozitivnu razliku između omjera). Te varijable uključuju i e:je i da:inf što potvrđuje njihovu relativnu sklonost akomodaciji. Međutim, razlike tih vrijednosti nisu značajne (relativno visoka p-vrijednost). Jedna od najistaknutijih varijabli iz prve analize, ica:ka, nalazi se u donjoj polovici tablice u drugoj analizi, i to samo s jednim korisnikom koji ju je proizveo i u Srbiji i izvan nje. Trebalo bi imati na umu da je većina vrijednosti varijabli u ovoj analizi temeljena na malom broju opažanja i da je manjak podataka vjerojatni razlog za neuočavanje ili

neznačajnost razlika. To je posebno vidljivo u četiri slučaja u kojima je razlika između omjera jednaka nuli, što se vrlo vjerojatno ne bi dogodilo kod većeg uzorka.

4 Rasprava

U odjeljku 3.4 ustanovili smo da rezultati prve analize donekle potvrđuju prisutnost jezične akomodacije kod mobilnih korisnika iz Srbije (u usporedbi s nemobilnim korisnicima). Druga analiza, s druge strane, ne otkriva nikakve razlike u jezičnoj produkciji mobilnih korisnika u Srbiji i izvan nje. Općenito, dokazi koji podupiru jezičnu akomodaciju nisu vrlo snažni, ali u osjetljivom sociopolitičkom kontekstu, poput onog u državama gdje se govore jezici BCHS-a, njezino postojanje svejedno je vrijedno pozornosti. Zanimljiva je činjenica da, u slučajevima kada se akomodacija događa, ona postaje obilježje govornikove jezične produkcije bez obzira na njegovu trenutačnu lokaciju. Takva situacija mogla bi ukazivati na dugoročne učinke. Struktura ovog istraživanja ne dopušta nam da čvršće potvrdimo tu pretpostavku, ali bi je bilo zanimljivo istražiti detaljnije u budućim istraživanjima, pogotovo zbog toga što se kontekst na koji smo se usredotočili razlikuje od konteksta tipičnih istraživanja govornika koji su se trajno preselili u područje gdje se govori drugi varijetet jezika (vidjeti Ruch i Benito Moreno, u tisku).

U pogledu lingvističkih varijabli koje su najsklonije akomodaciji, u našem se istraživanju pokazalo da su to one varijable po kojima se Srbija najjasnije ističe od ostale tri države, bez obzira na to je li varijabla fonetska, leksička ili morfosintaktička. Nismo razmatrali neke druge lingvističke čimbenike koji se često spominju u literaturi, poput razumljivosti (engl. *intelligibility*, koju ne sputava nijedna naša varijabla), uočljivosti (engl. *salience*, o kojoj trenutačno nemamo podataka) ili sinkronijskih varijacija u srpskom jeziku. Vezano uz posljednji čimbenik možemo spomenuti da je uporaba glagola »trebati« možda podložna akomodaciji zbog čisto preskriptivne naravi pravila o njegovoj impersonalnoj uporabi.

Također treba napomenuti da je na naše rezultate vrlo vjerojatno utjecalo nekoliko metodoloških i provedbenih odluka. Prva važna odluka tiče se načina na koji smo odredili zemlju boravišta. Udio od dvije trećine tvitova objavljenih u zemlji boravišta uistinu je proizvoljan (kao što bi bio i bilo koji drugi broj) i ne možemo znati jesu li svi korisnici koji su zadovoljili taj kriterij zbilja iz određene države ili su tamo samo proveli kraći period. Drugim riječima, nismo imali načina za razlikovanje zemlje u kojoj netko boravi trajno od one u kojoj boravi privremeno. Drugo, način na koji smo definirali mobilne korisnike, ubrajajući među njih one koji su imali vrlo malen postotak tvitova objavljenih izvan Srbije, možda je bio previše popustljiv i moguće je da je doprinio nedostatku razlika u našoj drugoj analizi. Međutim, količina podataka koju smo imali za mobilne korisnike bila je prilično ograničena, a manje ograničavajući kriterij za definiranje mobilnih korisnika bio je jedini način da se barem djelomično izbjegne problem pomanjkanja podataka.

Na kraju, zanemarivanje informacije o državi prema kojoj je mobilnost bila usmjerena onemogućilo nam je uspoređivanje jezične akomodacije između država, što je potrebno ako bismo htjeli detaljnije proučiti sociopolitičke čimbenike i stavove govornika, koji imaju veliki utjecaj na procese akomodacije.

5 Zaključak i smjernice za buduća istraživanja

U ovom smo radu proveli dvije analize na skupu geokodiranih poruka s Twittera koje su objavili srpski korisnici kako bismo izmjerili jezičnu akomodaciju koja prati mobilnost korisnika. Dok je prva analiza barem djelomično poduprla početnu hipotezu, druga nije pokazala efekte akomodacije.

Naše je istraživanje očito samo prvi korak u proučavanju akomodacije među jezicima BCHS-a s pomoću podataka s Twittera i obrade prirodnog jezika te postoje brojni mogući smjerovi za buduća istraživanja. Kao prvo, trebalo bi provesti razlikovnu analizu temeljenu na određenoj državi prema kojoj je mobilnost usmjerena: razumno je pretpostaviti, na primjer, da će srpski korisnik drugačije akomodirati jezik kada posjećuje Crnu Goru nego kad posjećuje Hrvatsku, zbog toga što srpski dijeli više dominantnih vrijednosti varijabli s crnogorskim nego s hrvatskim, a drugačiji je i sociopolitički kontekst. Kao drugo, u ovoj smo fazi mjerili akomodaciju samo u slučajevima kada su korisnici bili fizički mobilni, tj. kada su objavljivali tvitove iz zemalja gdje se govore jezici BCHS-a, a u kojima inače ne borave. Druga je vrsta mobilnosti koju treba uzeti u obzir »komunikacijska mobilnost«, tj. komunikacija s korisnicima Twittera koji su iz drugih država gdje se govore jezici BCHS-a. Kao treće, kada skup podataka postane dovoljno velik, ovo istraživanje treba proširiti na druge države gdje se govore jezici BCHS-a.

Priznanja

Ovo istraživanje potpomognuto je projektom Izvori, metode i alati za razumijevanje, prepoznavanje i razvrstavanje različitih oblika društveno neprihvatljivog diskursa koji financira Javna agencija za istraživačku aktivnost Republike Slovenije (ARRS J7-8280) te projektom Standardni srpski jezik: sintaksička, semantička i pragmatička proučavanja, koji financira Ministarstvo prosvjete, znanosti i tehnološkog razvoja Republike Srbije (178004).

LITERATURA

- Eugenija BARIĆ, Mijo LONČARIĆ, Dragica MALIĆ, Slavko PAVEŠIĆ, Mirko PETI, Vesna ZEČEVIĆ i Marija ZNIKA, 1997: Hrvatska gramatika, 2. izd. Zagreb: Školska knjiga.
Jack K. CHAMBERS, 1992: Dialect acquisition. *Language* 68/4. 673–705.
Adnan ČIRGIĆ, Ivo PRANJKOVIĆ i Josip SILIĆ, 2010: Gramatika crnogorskoga jezika. Podgorica: Ministarstvo prosvjete i nauke Crne Gore.

- Cristian DANESCU-NICULESCU-MIZIL, Michael GAMON i Susan DUMAIS, 2011: Mark my words! Linguistic style accommodation in social media. Proceedings of the International World Wide Web Conference. Hyderabad, India. 745–54.
- Gabriel DOYLE, 2014: Mapping dialectal variation by querying social media. Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics. Gothenburg: Association for Computational Linguistics. 98–106.
- Jacob EISENSTEIN, Brendan O’CONNOR, Noah A. SMITH i Eric P. XING, 2010: A latent variable model for geographic lexical variation. Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. Cambridge, MA: Association for Computational Linguistics. 1277–87.
- Daniel ERKER i Ricardo OTHÉGUY, 2016.: Contact and coherence: Dialectal leveling and structural convergence in NYC Spanish. *Lingua* 172–73. 131–46.
- Senahid HALILOVIĆ. 2004.: *Pravopis bosanskoga jezika za osnovne i srednje škole*. Zenica: Dom štampe.
- Howard GILES, Justine COUPLAND i Nikolas COUPLAND, 1991: Accommodation Theory: Communication, context, and consequence. Ur. H. Giles, J. Coupland i N. Coupland: *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press. 1–68.
- Bruno GONÇALVES i David SÁNCHEZ, 2014: Crowdsourcing dialect characterization through Twitter. *PLoS ONE* 9/11. <https://doi.org/10.1371/journal.pone.0112074>
- Nikola LJUBEŠIĆ, Darja FIŠER i Tomaž ERJAVEC, 2014: TweetCaT: A tool for building Twitter corpora of smaller languages. Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14). Reykjavik, Iceland. 2279–283.
- Nikola LJUBEŠIĆ, Filip KLUBIČKA, Željko AGIĆ i Ivo-Pavao JAZBEC, 2016.: New inflectional lexicons and training corpora for improved morphosyntactic annotation of Croatian and Serbian. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). Paris: European Language Resources Association (ELRA). 23–28.
- Nikola LJUBEŠIĆ, Maja MILIČEVIĆ PETROVIĆ i Tanja SAMARDŽIĆ: Borders and boundaries in Bosnian, Croatian, Montenegrin and Serbian: Twitter data to the rescue. *Journal of Linguistic Geography*. U tisku.
- Mitar PEŠIKAN, Jovan JERKOVIĆ i Mato PIŽURICA, 2010: *Pravopis srpskoga jezika*. Novi Sad: Matica srpska.
- Martin J. PICKERING i Simon GARROD, 2004: Toward a mechanistic psychology of dialogue. *Behavioral i Brain Sciences* 27. 169–90.
- Hanna RUCH, 2018: The role of acoustic distance and sociolinguistic knowledge in dialect identification. *Frontiers in Psychology* 9. Članak 818. <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00818/abstract>
- Hanna RUCH i Carlota DE BENITO MORENO: Linguistic Accommodation. Ur. H. Ruch, E. van Gijn, Max Wahlström, A. Hasse: *Language Contact*. Language Science Press.
- Tatjana SCHEFFLER, Johannes GONTRUM, Matthias WEGEL i Steve WENDLER, 2014: Mapping German tweets to geographic regions. Proceedings of the NLP4CMC Workshop at Konvens. Bochum: Bochumer Linguistische Arbeitsberichte. 26–34.

Branko Tošović, 2008: Gramatičke razlike između srpskog, hrvatskog i bošnjačkog jezika (preliminarium). Ur. T. Berger i B. Golubović: *Morphologie – Mündlichkeit – Medien: Festschrift für Jochen Raecke*. 311–22.

James WILSON, 2011.: Types of dialect accommodation in first-generation contact between adult speakers of mutually intelligible but regionally different varieties. *Multilingua* 30. 177–220.

POVZETEK

V prispevku obravnavamo fenomen jezikovnega prilagajanja med srbskimi uporabniki Twitterja, in sicer analiziramo tvite s podatki o geolokaciji, ki so bili objavljeni med letoma 2013 in 2016 na območju Bosne, Hrvaške, Črne gore in Srbije. Jezikovna produkcija uporabnikov Twitterja je opisana s 16 spremenljivkami, ki pogosto variirajo med govorce pluricentričnega jezika na območju Bosne, Hrvaške, Črne gore in Srbije. Teh 16 spremenljivk lahko razvrstimo v tri jezikoslovne tipe: fonetične (npr. uporaba praslovanskega samoglasnika jat'), leksikalne (npr. uporaba vprašalnega zaimka 'kaj') in oblikoskladenjske (npr. sintetična ali analitična oblika prihodnjika). V raziskavi izvedemo dve primerjavi: med produkcijo mobilnih in nemobilnih srbskih uporabnikov Twitterja ter med jezikom mobilnih uporabnikov, ki objavljajo znotraj ali zunaj Srbije. Kot mobilne upoštevamo tiste uporabnike, ki so v času zbiranja podatkov sporočila na Twitterju objavljali tudi izven Srbije. Za ločevanje med uporabniki iz Srbije in tistih iz drugih držav, kjer govorijo pred omenjeni pluricentrični jezik, definiramo državo prebivanja, in sicer tako, da morata biti vsaj dve tretjini vseh tvitov objavljeni v specifični državi, vse ostale uporabnike, ki temu kriteriju ne zadoščajo, pa zanemarimo. Prva analiza je pokazala pomembne razlike med mobilnimi in nemobilnimi uporabniki Twitterja za osem spremenljivk, vse pa imajo srednje velik učinek. Teh osem spremenljivk ne izkazuje nobenega vzorca glede na jezikoslovne tipe, vseeno pa je teritorialna razporeditev vrednosti teh spremenljivk pravilna, in sicer je večina spremenljivk s pomembnimi razlikami tistih, ki so najrazličnejše uporabljene v Srbiji na eni strani in v drugih obravnavanih državah na drugi strani. Druga analiza, tj. primerjava uporabe jezika mobilnih uporabnikov Twitterja znotraj in izven Srbije, ni pokazala neposrednega dokaza za jezikovno prilagajanje. Čeprav pri nobeni spremenljivki ni bilo pričakovanih pomembnih razlik, pa z opisnega vidika nekatere spremenljivke, ki so bile najmočnejši pokazatelj jezikovnega prilagajanja pri prvi analizi, izkazujejo enak trend, vendar ta ni statistično pomemben. Razlog bi lahko pripisali temu, da so vrednosti spremenljivk pri tej analizi temeljile na majhnemu številu pojavitev in da je zaradi pomanjkanja podatkov nekatere temeljne razlike težje identificirati.