

Ivana Lalli Pačelat

Filozofski fakultet Sveučilišta Jurja Dobrile u Puli
Ulica Ivana Matetića Ronjgova 1, HR-52100 Pula
ilalli@unipu.hr

Marija Brkić Bakarić

Odjel za informatiku, Sveučilište u Rijeci
Ulica Radmile Matejčić 2, HR-51000 Rijeka
mbrkic@uniri.hr

Isabella Matticchio

Institut für Kulturanalyse, Abteilung Mehrsprachigkeitsforschung, Alpen-Adria-Universität
Klagenfurt
Universitätsstraße 65-67, A-9020 Klagenfurt am Wörthersee
isabella.matticchio@aau.at

SLUŽBENA DVOJEZIČNOST U ISTARSKOJ ŽUPANIJU: STANJE I PERSPEKTIVE

Službena dvojezičnost pretpostavlja svakodnevno stvaranje usporednih tekstova u dvojezičnim područjima. Slučaj je takav i u Istarskoj županiji, u kojoj se tekstovi obično sastavljaju na hrvatskome, a zatim se prevode na talijanski jezik. Zbog činjenice da je riječ o službenim tekstovima i zbog konteksta uporabe talijanskoga jezika vrlo je važno imati precizno i ujednačeno nazivlje te razvijene jezične tehnologije koje bi omogućile brže i kvalitetnije prevođenje u dvojezičnim institucijama u Istarskoj županiji. Ciljevi su ovoga rada: (1) prikazati ostvarivanje ravnopravne službene uporabe talijanskoga kao manjinskoga jezika u Istarskoj županiji (1a) analizom prevedenih sadržaja na službenim mrežnim stranicama službeno dvojezičnih gradova i općina i (1b) prikazom dosadašnje prakse prevođenja te (2) ukazati na važnost i nužnost razvoja jezičnih tehnologija (2a) prikazom trendova u razvoju jezičnih tehnologija u sličnim dvojezičnim i višejezičnim institucijama, (2b) prikazom pripreme usporednoga korpusa administrativnih tekstova Istarske županije i (2c) analizom postojećega nazivlja provedenom na priređenome korpusu.

Rezultati analize dostupnosti dvojezičnih sadržaja na mrežnim stranicama te analize nazivlja provedene na usporednome korpusu pokazali su da je nužno razviti i upotrebljavati prevoditeljske alate i jezične izvore prilagođene talijanskomu kao manjinskomu jeziku kako bi se olakšalo i ubrzalo prevođenje, a time i omogućilo uspješnije ostvarivanje ravnopravne uporabe talijanskoga kao manjinskoga jezika.

1. Uvod

Službena dvojezičnost ili višejezičnost pretpostavlja proizvodnju velikoga broja usporednih tekstova: prijevoda i izvornika. Zbog činjenice da je riječ o službenoj uporabi jezika u službenim tekstovima neophodno je razviti jezične tehnologije koje bi omogućile brže i kvalitetnije prevođenje u dvojezičnim i višejezičnim institucijama. Radi olakšavanja prevođenja potreba za razvojem jezičnih tehnologija postoji i u Istarskoj županiji, u kojoj su hrvatski i talijanski jezik u ravnopravnoj službenoj uporabi.

Ciljevi su ovoga rada (1) prikazati ostvarivanje ravnopravne službene uporabe talijanskoga kao manjinskoga jezika u Istarskoj županiji (1a) analizom prevedenih sadržaja na službenim mrežnim stranicama dvojezičnih gradova i općina i (1b) prikazom dosadašnje prakse prevođenja te (2) ukazati na važnost i nužnost razvoja jezičnih tehnologija (2a) prikazom trendova u razvoju jezičnih tehnologija u sličnim dvojezičnim i višejezičnim institucijama, (2b) prikazom pripreme usporednoga korpusa administrativnih tekstova Istarske županije i (2c) analizom nazivlja provedenom na priređenome korpusu. Razvoj jezičnih tehnologija, čiji je prvi korak učinjen sastavljanjem usporednoga korpusa, pojednostavnio bi prevođenje u Istarskoj županiji i omogućio stvaranje prijevoda visoke kvalitete, a time i ostvarivanje prava na ravnopravnu službenu uporabu talijanskoga kao manjinskoga jezika.

2. Službena dvojezičnost u Istarskoj županiji

2.1. Ravnopravna službena uporaba talijanskoga kao manjinskoga jezika

Službena dvojezičnost u Istarskoj županiji regulirana je *Ustavnim zakonom o pravima nacionalnih manjina*¹ (NN 155/02, 47/10, 80/10, 93/11), *Zakonom o uporabi jezika i pisma nacionalnih manjina u Republici Hrvatskoj*² (NN 51/00 56/00), *Zakonom o lokalnoj i područnoj (regionalnoj) samoupravi*³ (NN 33/01,

¹ www.zakon.hr/z/295/Ustavni-zakon-o-pravima-nacionalnih-manjina (pristupljeno 30. rujna 2019.).

² <https://zakon.hr/z/1851/Zakon-o-uporabi-jezika-i-pisma-nacionalnih-manjina-u-Republici-Hrvatskoj> (pristupljeno 30. rujna 2019.).

³ [https://zakon.hr/z/132/Zakon-o-lokalnoj-i-podru%C4%8Dnoj-\(regionalnoj\)-samoupravi](https://zakon.hr/z/132/Zakon-o-lokalnoj-i-podru%C4%8Dnoj-(regionalnoj)-samoupravi) (pristupljeno 30. rujna 2019.).

60/01), *Statutom Istarske županije*⁴ (SNIŽ 10/09, SNIŽ 4/13, SNIŽ16/16, SNIŽ 1/17, SNIŽ 2/17, SNIŽ 2/18 i SNIŽ 30/18) te *Statutima gradova i općina*. Prema *Statutu Istarske županije* (čl. 2., 2009.) sedam je gradova i trinaest općina službeno dvojezično: Buje – Buie, Novigrad – Cittanova, Poreč – Parenzo, Pula – Pola, Rovinj – Rovigno, Umag – Umago i Vodnjan – Dignano te Bale – Valle, Brtonigla – Verteneglio, Fažana – Fasana, Funtane – Fontane, Grožnjan – Grisignana, Kaštelir – Labinci – Castelliere – Santa Domenica, Ližnjan – Lisignano, Motovun – Montona, Oprtalj – Portole, Tar – Vabriga – Torre – Abrega, Višnjan – Visignano, Vižinada – Visinada, Vrsar – Orsera. Za detaljan prikaz analize pravnih članaka koji propisuju dvojezičnost u svim dvojezičnim gradovima i općinama vidi Zilli 2017.

Prema posljednjemu objavljenom popisu Državnoga zavoda za statistiku (Popis stanovništva 2013) u Istarskoj su se županiji 12 543 hrvatska državljanina izjasnila da su talijanske narodnosti, što čini 6,03 % stanovništva te županije. Prema tomu popisu 14 205 stanovnika Istarske županije odabralo je talijanski kao materinski jezik.

Prema Statutu Istarske županije hrvatski i talijanski jezik „u ravnopravnoj su službenoj uporabi u radu županijskih tijela u samoupravnom djelokrugu” (čl. 6.) te se „ravnopravna službena upotreba hrvatskoga i talijanskoga jezika ostvaruje kako u radu svih tijela županije u samoupravnom djelokrugu i u postupku pred upravnim tijelima” (čl. 21.) tako i „na dijelu ili cijelom području općina i gradova u Istarskoj županiji, gdje prebivaju pripadnici/pripadnice talijanske nacionalne zajednice” (čl. 23.). U skladu sa Statutom Istarske županije, ali i statutima dvojezičnih gradova i općina, hrvatski i talijanski jezik u ravnopravnoj su službenoj uporabi, a javno objavljeni sadržaji na mrežnim stranicama trebali bi biti sastavljeni dvojezično, odnosno trebali bi biti prevedeni na talijanski jezik.

2.2. Analiza prevedenih sadržaja na mrežnim stranicama

Kako bi se provjerila ravnopravna uporaba talijanskoga kao manjinskoga jezika, analizirani su objavljeni dvojezični sadržaji na službenim mrežnim stranicama svih dvojezičnih gradova i općina. Istražilo se koliko se sadržaja prevodi i koji se

⁴ www.istra-istria.hr/index.php?id=63 (pristupljeno 5. veljače 2020.).

najčešće prevode na mrežnim stranicama službeno dvojezičnih gradova i općina, utječe li vrsta teksta na dostupnost prijevoda te ovisi li dostupnost prijevoda o vrsti ustroja mjesne samouprave. Radi zaštite podataka imena gradova i općina nisu navedena, nego su kodirana tako da su gradovima dodijeljeni kodovi od G1 do G7, a općinama od O1 do O13. Radi jasnijega prikaza i tumačenja rezultata odlučeno je da će se prevedeni sadržaji promatrati podijeljeni u tri skupine tekstova, što nije teorijski motivirano jer su u istu skupinu uvršteni i tekstovi koji pripadaju različitim funkcionalnim stilovima (usp. Frančić, Hudeček i Mihaljević 2006) i tekstnim vrstama u javnoj komunikaciji (usp. Ivanetić 2003). Budući da se različiti sadržaji u nejednakoj mjeri prevode na analiziranim mrežnim stranicama, radi lakše analize tekstovi su podijeljeni u tri skupine: službeni tekstovi, opisni tekstovi i informativni tekstovi. U skupinu službenih tekstova uključeni su tekstovi koji se na pojedinim službenim mrežnim stranicama nazivaju službenim dokumentima, a to su: službene novine (u gradovima G3 i G5 rabi se naziv *službeni glasnik*), statut i proračun. U skupinu opisnih tekstova svrstani su tekstovi koji se odnose na opće podatke, odnosno riječ je o tekstovima koji se ne mijenjaju ili o općim informacijama o gradu ili općini. Na kraju, u skupinu informativnih tekstova okupljeni su tekstovi u kojima je riječ o novostima, javnoj nabavi i natječajima. Zbog svoje uloge takvi se tekstovi često prevode i dovode do češćih posuvremenjivanja mrežnih stranica.

Postojanje talijanskoga prijevoda prikazano je u tablici 1. i bilježeno je sljedećim oznakama: P (prevedeno) ako su tekstovi u cijelosti prevedeni na oba jezika i vremenski usklađeni, DjP (djelomično prevedeno) ako prijevodi na talijanski jezik postoje, ali nisu prevedeni svi sadržaji i NP (nije prevedeno) ako prijevoda na talijanski jezik nema ili je preveden samo naziv.

Tablica 1. Analiza prevedenih tekstova u gradovima

		G1	G2	G3	G4	G5	G6	G7
službeni tekstovi	službene novine	NP	P	NP	P	P	DjP	DjP
	statut	NP	P	P	P	P	P	P
	proračun	DjP	P	DjP	NP	DjP	NP	NP
opisni tekstovi	opći podatci	P	P	P	DjP	P	P	P
informativni tekstovi	novosti	DjP	P	NP	DjP	DjP	P	DjP
	javna nabava	NP	NP	NP	NP	DjP	NP	NP
	natječaji	NP	P	NP	NP	P	DP	P

Na stranicama gradova puno je više prevedenih sadržaja nego na stranicama općina. Od sedam gradova najviše prevedenih tekstova imaju dva grada (G2 i G5), a najčešće su prevedeni tekstovi statut i opći podatci.⁵ Iako ti tekstovi pripadaju različitim tekstnim skupinama (službeni tekstovi i opisni tekstovi), riječ je o sadržajima koji se, za razliku od ostalih, ne mijenjaju često te ih je dovoljno prevesti samo jednom. S druge strane, gotovo ni jedan grad ne prevodi tekstove javne nabave, osim grada G5 koji ih prevodi djelomično. Od službenih dokumenata, kao što je već navedeno, najčešće se prevodi statut, a najrjeđe proračun, koji je u cijelosti preveden samo u jednome gradu (G2). Zanimljivo je primijetiti da tri grada (G1, G3 i G5) ne prevode u cijelosti tekst proračuna, nego samo jedan njegov dio,⁶ a u dvama gradovima na mrežnim stranicama prevedeni su samo nazivi dokumenata. Primjerice, u jednome su gradu (G6) prevedeni samo nazivi dokumenata do 2017. godine, a svi su sadržaji dokumenata na hrvatskome jeziku. Dok se nazivi prevode brže i omogućuju određenu usklađenost mrežnih stranica na obama jezicima, prijevod sadržaja zahtijeva puno više vremena. Da bi svi sadržaji bili istodobno prevedeni, treba imati dovoljan broj prevoditelja s odgovarajućim jezičnim alatima i/ili dovoljna financijska sredstva, pa se pretpostavlja da toga nije bilo. Kad je riječ o statutima, u gradu G1 prevedeno je samo nekoliko općih podataka koji se odnose na sadržaj statuta, dok je u gradu G4 statut preveden i objavljen u *Službenim novinama Grada* u 2009. godini, ali one nisu dostupne na mreži. Službene novine nisu prevedene samo u dvama gradovima (G1 i G3), a u dvama se slučajevima (G6 i G7) djelomično prevode. Primjerice, na mrežnim stranicama grada G6 dostupne su *Službene novine* od 2007. do 2019. godine (u 2019. nije objavljen niti jedan broj na hrvatskome jeziku), a najviše je brojeva na talijanskome jeziku objavljeno u razdoblju od 2007. do 2010. godine. Sličan podatak odnosi se i na grad G7, na čijim su se stranicama prijevodi *Službenih novina* na talijanskome jeziku objavljivali od 2006. do 2014. godine, a u posljednjih pet godina nije objavljen niti jedan broj. Taj podatak može ukazivati na manjak prevoditelja ili financijskih sredstava namijenjenih za prevođenje nakon 2014. godine. Kad je riječ o informativnim tekstovima, primjećuje se da se najmanje prevode tekstovi koji se odnose na javnu nabavu te da se, ako postoje prijevodi, oni odnose isključivo na sam naziv javne nabave ili na nekoliko naziva, kao što je slučaj u gradovima G1 i G7 u kojima su prevedeni nazivi nekoliko poglavlja koja se odnose na javnu nabavu. Natječaji su druga vrsta teksta

⁵ G3 je primjerice sve sadržaje na talijanskome jeziku koji se odnose na grad i njegovo ustrojstvo objavio u PDF-u, dok G4 nije preveo tekstove koji se odnose na opise kulturnih znamenitosti.

⁶ Poput primjerice *Vodiča za građane*, na tal. *Guida per i cittadini*.

koja se rjeđe prevodi, a novosti su (vijesti, događaji i sl.) tekstovi u skupini informativnih tekstova koji se prevode, ali većinom je riječ o djelomičnim prijevodima, što znači da nisu prevedene sve novosti koje su objavljene na hrvatskome jeziku. Tako je primjerice u gradu G7 posljednji prijevod objavljen prije pet mjeseci (30. travnja 2019.). Ako je vijest prema Hrvatskoj enciklopediji „prva obavijest koja se daje ili prima o kakvu nedavnom događaju”,⁷ ona bi trebala biti pravodobno prevedena jer su inače govornici talijanskoga jezika i pripadnici talijanske manjine zakinuti.

Kad je riječ o općinama, kao što je prikazano u tablici 2., valja napomenuti da većina nema mrežne stranice na talijanskome jeziku (O3, O4, O6 – O10, O12, O13). Rijetki prijevodi koji postoje prijevodi su statuta i općih podataka. Što se tiče proračuna i javne nabave, koji su djelomično prevedeni u samo jednoj od trinaest općina (O2), valja napomenuti da su prevedeni sadržaji iz *Popisa postupaka javne nabave*,⁸ ali nisu prevedeni sadržaji iz *Plana javne nabave*,⁹ što može ukazati na strategiju kojom se očito koristi pri odluci koji će se sadržaji prevoditi: najprije se prevode važni dokumenti i oni koji nisu podložni čestim izmjenama. Nadalje, slično stranicama gradova, tekstovi natječaja nisu istodobno objavljeni na obama jezicima. Talijanski se tekstovi uglavnom objavljuju nekoliko mjeseci nakon objave teksta na hrvatskome, pa su govornici talijanskoga jezika primorani čitati tekstove na hrvatskome jeziku. U jednoj je općini zapažen izostanak teksta na hrvatskome jeziku, što pokazuje da postoje i slučajevi, iako iznimno rijetki, u kojima se tekstovi najprije sastavljaju na talijanskome jeziku, a zatim prevode na hrvatski.

Tablica 2. Analiza prevedenih tekstova u općinama

		O1	O2	O3	O4	O5	O6	O7	O8	O9	O10	O11	O12	O13
službeni dokumenti	službene novine	NP	P	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP
	statut	P	P	NP	NP	P	NP	NP	NP	NP	NP	P	NP	NP
	proračun	NP	DjP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP
opisni tekstovi	opći podatci	P	P	NP	NP	P	NP	NP	NP	NP	NP	P	NP	NP
informativni tekstovi	novosti	P	P	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP
	javna nabava	NP	DjP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP
	natječaji	DjP	P	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP

⁷ www.enciklopedija.hr/Natuknica.aspx?ID=64600 (pristupljeno 30. rujna 2019.).

⁸ Na tal. *Piano degli appalti pubblici*.

⁹ Na tal. *Elenco degli appalti pubblici*.

Nakon prikaza analize dostupnosti prevedenih tekstova na mrežnim stranicama službeno dvojezičnih gradova i općina provjereno je statističkom analizom utječe li vrsta teksta na dostupnost prijevoda te ovisi li dostupnost prijevoda o vrsti ustroja mjesne samouprave. Prva je nulta hipoteza da dostupnost prijevoda ne ovisi o vrsti teksta. S pomoću χ^2 -testa ispitana je neovisnost dvaju obilježja. Čak i kad se uzmu u obzir samo nadređene kategorije koje se odnose na vrstu teksta, vrijednosti u pojedinim ćelijama vrlo su niske, pa je potrebno ispraviti neprekidnosti prema Yatesu. Za razinu značajnosti od $\alpha = 5\%$ i četiri stupnja slobode kritična vrijednost χ^2 iznosi 9,49. Izračunat χ^2 iznosi 6,271, stoga se prihvaća nulta hipoteza, tj. analizirani uzorak ne odstupa statistički značajno od nulte hipoteze. Hi-kvadrat testom nezavisnosti na razini značajnosti od 5% uz korekciju neprekidnosti prema Yatesu (kritična vrijednost χ^2 -testa za razinu značajnosti od 5% i dva stupnja slobode iznosi 5,991) utvrđena je statistički značajna razlika u dostupnosti prijevoda jedino između opisnih i informativnih tekstova ($\chi^2 = 6,39$, $p = 0,04$). Naime, informativni tekstovi objavljuju se često i stoga zahtijevaju često prevođenje za razliku od opisnih koje je dovoljno prevesti jednom. Hi-kvadrat testom nije utvrđena statistički značajna razlika u dostupnosti prijevoda između službenih i opisnih tekstova, a ni između službenih i informativnih tekstova. Razlog je vjerojatno činjenica da se određene vrste dokumenata poput službenih novina objavljuju često, slično informativnim tekstovima, dok se primjerice statut mijenja vrlo rijetko, slično opisnim tekstovima. Druga je nulta hipoteza da dostupnost prijevoda ne ovisi o vrsti ustroja mjesne samouprave. Hipoteza se odbacuje (kritična vrijednost χ^2 za razinu značajnosti od 1% i dva stupnja slobode iznosi 9,210) jer je izračunata χ^2 vrijednost od 33,893 ($p \leq 0,001$). Razlika između gradova i općina u dostupnosti prijevoda službenih tekstova statistički je značajna na razini od 1% ($\chi^2 = 15,275$, $p \leq 0,001$), jednako kao razlika između gradova i općina u dostupnosti prijevoda opisnih tekstova ($\chi^2 = 15,934$, $p \leq 0,001$). Hi-kvadrat testom nije dobivena statistički značajna povezanost između dostupnosti prijevoda informativnih tekstova i vrste ustroja mjesne samouprave.

Iz analize prevedenih sadržaja na mrežnim stranicama službeno dvojezičnih gradova i općina Istarske županije i statističke analize proizlazi, dakle, da pripadnici talijanske manjine ne ostvaruju uvijek pravo na ravnopravnu službenu uporabu talijanskoga kao manjinskoga jezika na službenim mrežnim stranicama pojedinih gradova i općina. Većina objavljenih tekstova na hrvatskome jeziku

nije prevedena na talijanski jezik ili pak nije pravodobno prevedena. To se posebice odnosi na informativne tekstove s čijim se objavljivanjem uvelike kasni, pa se ne može govoriti o ravnopravnoj uporabi dvaju jezika. Dostupniji su prijevodi na mrežnim stranicama gradova, a izostaju prijevodi sadržaja na mrežnim stranicama općina: devet općina trenutačno nema objavljen ni jedan tekst na talijanskome jeziku na svojim mrežnim stranicama. Gradovi i općine znatno se razlikuju po dostupnosti prijevoda opisnih i službenih tekstova.

Nadalje, analizom je uočena uporaba raznih strategija kod djelomičnoga prevođenja na talijanski jezik, pa se tako primjerice prevode ili samo nazivi tekstova ili se odabiru najvažnije informacije, koje se potom prevode na talijanski jezik u sažetome obliku. Isto tako, primijećeno je postojanje razdoblja kad se prevodi više ili manje tekstova, što ukazuje na to da je u određenome razdoblju određeni grad ili općina imao veći broj prevoditelja ili veća financijska sredstva predviđena za prevođenje.

Prikazani rezultati analize prevedenih sadržaja na mrežnim stranicama dvojezičnih gradova i općina ukazuju na važnost osvješćivanja problema prevođenja i na nužnost uporabe jezičnih tehnologija i organiziranoga pristupa aktivnosti prevođenja na županijskoj razini kako bi se omogućilo ostvarivanje prava na ravnopravnu uporabu dvaju jezika.

2.3. Prevođenje u Istarskoj županiji – prikaz dosadašnje prakse prevođenja

Za razumijevanje trenutačnoga stanja dostupnosti prijevoda na mrežnim stranicama dvojezičnih gradova i općina bilo je potrebno istražiti i prikazati dosadašnju praksu prevođenja u institucijama Istarske županije.

U Istarskoj se županiji i dvojezičnim gradovima i općinama u većini slučajeva tekstovi sastavljaju na hrvatskome jeziku, a zatim prevode na talijanski jezik. Do 2013. godine uglavnom se nisu upotrebljavale prevoditeljske tehnologije za olakšavanje prevođenja te nije postojala organizirana aktivnost prevođenja. Prvi koraci u osvješćivanju problema povezanih s prevođenjem u dvojezičnim institucijama u županiji i predstavljanju mogućih rješenja učinjeni su u Lalli Pačelat 2013. Vrlo važan korak i prvi konkretni pomoci učinjeni su projektom

S.I.M.P.L.E. (2013. – 2015.). U sklopu toga dvogodišnjeg projekta Istarske županije osnovana je Agencija za potporu dvojezičnosti, koja je, između ostaloga, pružala usluge prevođenja i jezičnoga savjetništva građanima i dvojezičnim institucijama u županiji. Projektom su nabavljeni prevoditeljski alati, organizirana je edukacija prevoditelja i počeli su se upotrebljavati ti alati. U sklopu projekta izrađen je glosar i vodič za dvojezičnost. Pozitivni učinci toga projekta vidljivi su i u dijeljenju prijevodne memorije Istarske županije, kao središnje memorije, ostalim gradovima i općinama, što je dovelo do veće terminološke ujednačenosti prijevoda na razini županije, gradova i općina. Treba spomenuti i aktivno uključivanje predstavnika akademskih i županijskih tijela u mrežu stručnjaka *Rete per l'eccellenza dell'italiano istituzionale* (REI), koja djeluje pri Glavnoj upravi za pismeno prevođenje (DGT) – Službi za talijanski jezik Europske komisije od 2014. godine. Radna skupina *Osservatorio dell'italiano istituzionale fuori d'Italia* (OIIFI), u kojoj sudjeluju i hrvatski predstavnici, a koja djeluje pri toj mreži, brine se o uporabi talijanskoga jezika izvan granica Italije, a posebno se bavi usklađivanjem nazivlja i jezičnim savjetništvom. Važan je pomak učinjen i projektom *Službena dvojezičnost u slovenskoj i hrvatskoj Istri: primjer prevođenja administrativnih tekstova*. Riječ je o znanstvenoistraživačkome projektu u sklopu znanstveno-tehnološke suradnje sa Slovenijom (2018. – 2019.), kojemu je, između ostaloga, cilj bio ponuditi analizu trenutačnoga ostvarivanja prava na ravnopravnu službenu uporabu talijanskoga kao manjinskoga jezika, analizu postojećega nazivlja te prikaz i analizu trenutačne situacije i prakse prevođenja. Jedan od glavnih ciljeva projekta bila je priprema i izgradnja usporednoga korpusa koji će služiti kao temelj daljnjemu razvoju jezičnih tehnologija za prevođenje s hrvatskoga na talijanski i obrnuto.

2.4. Usporedni korpusi u izgradnji jezičnih tehnologija

Na područjima ili u organizacijama u kojima su dva ili više jezika u ravnopravnoj službenoj uporabi neminovna je proizvodnja velikoga broja usporednih tekstova. Usporedni tekstovi često se priređuju u usporedne korpusne (npr. The JRC-Acquis (Steinberger i dr. 2006), The Hansard French-English corpus, The United Nations Parallel Corpus v1.0 (Ziemski i dr. 2016), The Trilingual Allegra-Corpus of German, Italian and Romansh (Scherrer i Cartoni 2012 i dr.)) i

upotrebljavaju se za razvoj jezičnih tehnologija radi podizanja kvalitete prijevoda i smanjivanja troškova prevođenja u dvojezičnim i višejezičnim institucijama i organizacijama, a sve radi očuvanja dvojezičnosti ili višejezičnosti. Usporedni korpusi ključni su za strojno potpomognuto prevođenje jer se rabe za izradu prijevodnih memorija, terminoloških baza, ali su i koristan primarni izvor za pretraživanje i promatranje jezične uporabe (usp. McEnery i Xiao 2007, Kenning, 2010, Somers 2001). Riječ je o iznimno dragocjenome jezičnom izvoru koji se upotrebljava i u razvoju sustava za statističko i neuronsko strojno prevođenje. Kvaliteta korpusa važna je u pripremi sustava za strojno prevođenje i utječe na njegovu uspješnost.

Najbolji je primjer razvoja prevoditeljskih tehnologija, raspoloživih alata i izvora koji olakšavaju i omogućavaju proizvodnju velikoga broja prijevoda Europska komisija. Prikaz dostupnih alata donose Steinberger i dr. (2014) i Europska komisija (2016).

Potreba za razvojem prevoditeljskih tehnologija osim u višejezičnim zajednicama postoji i u svim zajednicama u kojima postoji i službena dvojezičnost na regionalnoj razini, kao primjerice u Istarskoj županiji, a tada se tekstovi uglavnom prevode s većinskih jezika na manjinske. Problem je manjinskih jezika što obično imaju samo manje korpusa, pa je nemoguće izgraditi veće usporedne korpusa na kojima bi se temeljio razvoj jezičnih tehnologija i s pomoću kojih bi se olakšala njihova izgradnja (Trosterud 2002: 116). Unatoč tomu, provodili su se različiti projekti kojima je, između ostaloga, cilj bio razvoj jezičnih tehnologija za manjinske jezike, poput sustava za strojno prevođenje sa španjolskoga na baskijski i obrnuto (Alegria i dr. 2005) i sa španjolskoga na katalonski i obrnuto (Arranz, Comelles i Farwell 2006), zatim oni kojima je cilj bio izgradnja nazivlja i razvoj terminoloških alata i izvora za engleski i galski (Jones i Prys 2006) ili onih za talijanski, njemački i ladinski (Streiter i dr. 2004). Svim tim projektima prethodila je izrada usporednih korpusa.

2.4.1. Postojeći usporedni korpusi koji uključuju hrvatski i talijanski

Iako postoje usporedni korpusi i za hrvatski i za talijanski jezik, oni su većinom u kombinaciji s engleskim jezikom, na što upozoravaju Tadić i dr. (2012) za hrvatski jezik i Calzolari i dr. (2012) za talijanski jezik. Valja napomenuti, međutim,

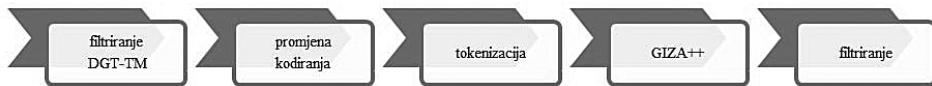
da postoji i nekoliko višejezičnih usporednih korpusa koji uključuju i hrvatski i talijanski jezik, poput primjerice: OPUS2 parallel corpus (Tiedemann 2012), DGT-Translation Memory (Steinberger i dr. 2012), InterCorp (Čermák i Rosen 2012), Bulgarian-X language Parallel Corpus (Koeva i dr. 2012), EUR-Lex Corpus (Baisa i dr. 2016), Eur-Lex judgments corpus (Baisa i dr. 2016), EAC-TM i dr. Međutim, bez obzira na to što neki od tih korpusa uključuju tekstove koji pripadaju administrativnomu funkcionalnom stilu, oni ne mogu u potpunosti zadovoljiti potrebe prevoditelja Istarske županije i ne mogu se smatrati najkvalitetnijim jezičnim izvorom za razvoj jezičnih tehnologija zbog specifičnoga dvojezičnog nazivlja. Budući da postoje usporedni tekstovi koji već imaju razrađeno ili djelomično razrađeno nazivlje prilagođeno manjinskomu statusu talijanskoga jezika, šteta je ne iskoristiti postojeće tekstove za stvaranje usporednih korpusa koji bi se upotrebljavali u daljnjoj izgradnji jezičnih izvora i alata koji bi olakšali prevođenje u Istarskoj županiji.

2.4.2. Priprema korpusa administrativnih tekstova Istarske županije

Budući da u Istarskoj županiji, kao što je već spomenuto, dosad nije bilo nikakvih inicijativa koje bi poticale uporabu usporednih tekstova za razvoj neophodnih jezičnih tehnologija koje bi unaprijedile, ubrzale i olakšale rad prevoditelja i znatno smanjile troškove prevođenja, prvi korak bio je priprema i izgradnja hrvatsko-talijanskoga usporednog korpusa dostupnih tekstova administrativnoga funkcionalnog stila. Korpus koji uključuje tekstove na talijanskome jeziku s nazivljem prilagođenim njegovu manjinskom statusu poslužit će i u daljnjemu razvoju jezičnih tehnologija koje bi zadovoljile specifične potrebe za prevođenjem u dvojezičnim institucijama Istarske županije.

Usporedni korpus sastavljen je od hrvatskih službenih javno dostupnih tekstova i njihovih prijevoda na talijanski jezik. Ručnim vrednovanjem mogućih tekstova i prijevoda koji bi bili uključeni u korpus odabrane su službene novine. Službene su novine odabrane, s jedne strane, jer je riječ o tekstovima koje je najlakše usporediti, a koji su objavljeni na različitim mrežnim stranicama i koje su preveli različiti prevoditelji, pa su pogodni za analizu terminološke raznolikosti. S druge strane, riječ je o najčešće prevedenim službenim dokumentima koji su mrežno dostupni. Zbog tehničkih uvjeta kojima su morali udovoljavati odabrane

su samo službene novine objavljene na mrežnim stranicama Istarske županije, Novigrada, Pule i Umaga. Automatskim postupkom dohvaćeni su odabrani dokumenti različitih oblika. Detaljnije informacije i tehnički podaci o izradi korpusa prikazani su u Brkić Bakarić i Lalli Pačelat 2019. Svrnjivanju korpusa prethodila je izrada rječnika. Rječnik je izrađen izlučivanjem 1267 tmx dokumenata i 284 864 prijevodne jedinice po jeziku iz DGT – TM (Steinberger i dr. 2012). Automatski je generiran rječnik s GIZA++ alatom (Och i Ney 2003). Dohvaćeni rječnik najvjerojatnijih prijevoda od 793 803 natuknice nakon filtriranja sadržavao je 107 269 natuknica. Postupak izrade rječnika prikazan je na slici 1.



Slika 1. Postupak izrade rječnika

Automatski dohvaćeni dokumenti različitih formata svrnjeni su alatom *hunalign* (Varga i dr. 2004). Postupak izrade usporednoga korpusa prikazuje slika 2.



Slika 2. Postupak izrade usporednoga korpusa
(prema Brkić Bakarić i Lalli Pačelat 2019)

Hrvatsko-talijanski usporedni korpus administrativnih tekstova sastavljen je od 423 usporedna dokumenta. Tablica 3. prikazuje broj dohvaćenih dokumenata za svaki jezik i broj svrnjenih dokumenata. Podatci su prikazani za svaki potkorpus zasebno.

Tablica 3. Broj dohvaćenih dokumenata po jeziku
i broj svrnjenih usporednih dokumenata

	broj dokumenata na hrvatskome jeziku	broj dokumenata na talijanskome jeziku	broj svrnjenih dokumenata
G0	660	429	251
G2	126	106	65
G4	65	57	37
G6	286	71	70

Automatskim dohvaćanjem dokumenata s odabranih mrežnih izvora ustanovljeno je da je samo 37 % službenih dokumenata dostupno na talijanskome jeziku.

Broj pojava po svakome potkorpusu prikazan je u tablici 4., a sveukupno hrvatski dio korpusa ima 6,3 milijuna pojava, dok talijanski dio korpusa ima 8,4 milijuna pojava.

Tablica 4. Broj pojava po potkorpusima

potkorpus	broj pojava	
	hrvatski jezik	talijanski jezik
G0	3,2M	3,4M
G2	1,2M	1,7M
G4	858K	1,0M
G6	1,8M	2,3M

Treba napomenuti da se tijekom izrade korpusa različitim postupcima pokušalo doći do boljih rezultata sravnjivanja, što je prikazano u Brkić Bakarić i Lalli Pačelat 2019. Odrađeno je tako ručno vrednovanje preciznosti automatski izrađenoga rječnika na sto istovrijednica odabranih slučajnim odabirom te vrednovanje preciznosti sravnjivanja uporabom različitih postupaka jednakoga broja uparenih rečenica, također odabranih slučajnim odabirom. Rezultati vrednovanja preciznosti pokazali su uglavnom zadovoljavajuće rezultate, za rječnik 79 %, dok je za sravnjivanje, ovisno o postupku, ta vrijednost od 78 % do 93 %. Više o postupcima sravnjivanja i rezultatima vrednovanja v. u Brkić Bakarić i Lalli Pačelat 2019.

U budućem radu nastaviti će se sastavljanje korpusa drugih dostupnih usporednih tekstova s drugih izvora i drugim postupcima radi dobivanja kvalitetnoga usporednog korpusa koji bi služio i za razvoj strojnoga prevođenja. Tako priređeni korpus služio je za planiranu analizu nazivlja.

2.5. Analiza nazivlja

Tekstovi koji čine analizirani korpus pripadaju uglavnom administrativnom funkcionalnom stilu. Poželjne su značajke administrativnoga funkcionalnog stila, između ostaloga, objektivnost, točnost i jasnoća (usp. Frančić i dr. 2006: 254), a tomu svakako ne pridonosi terminološka raznolikost i neujednačenost koja se susreće u tekstovima objavljenim na talijanskome jeziku, o čemu se već upozor-

ravalo u Lalli Pačelat 2013. Dok su se prethodne analize mogle provoditi samo kvalitativno, zahvaljujući izrađenom korpusu, analizi se nazivlja pristupilo i kvantitativno. Za uočavanje terminološke raznolikosti u ovome radu upotrebljavao se u većini slučajeva popis naziva koji je dobiven automatskim crpenjem naziva iz talijanskoga dijela usporednoga korpusa. Crpenje naziva učinjeno je za svaki potkorpus zasebno kako bi se opazila terminološka raznolikost. Dobiveno je tako tisuću naziva za svaki potkorpus. Uporaba odabranih naziva potvrđena je i u korpusu. Za automatsko crpenje naziva i pretragu korpusa upotrijebljen je Sketch Engine. U automatskome crpenju naziva upotrijebljena je mjera ključnosti (engl. *keyness*), čija formula glasi:

$$\frac{f_{pm_{focus}} + n}{f_{pm_{ref}} + n}$$

pri čemu je $f_{pm_{focus}}$ normalizirana frekvencija (po milijunu) riječi iz fokusnoga korpusa, $f_{pm_{ref}}$ normalizirana frekvencija (po milijunu) riječi iz referentnoga korpusa i n jednostavni parametar koji iznosi 1 prema zadanoj vrijednosti. Kao referentnim korpusima koristilo se korpusima *Italian Web 2016* i *Croatian Web*.

U nastavku je prikazano nekoliko primjera nepoželjne terminološke raznolikosti uočene u prijevodima na talijanski jezik. Uz svaku pojavnici u zagradi navedeno je kodirano ime potkorpusa u kojemu je primjer pronađen te podatak o ključnosti i čestoti.

Nerijetko se u prijevodima susreće nepotrebna uporaba dvaju različitih naziva za jedan pojam, poput *zaštite okoliša* koja se prevodi kao *protezione ambientale* (G4: 3,96; 10) i *tutela ambientale* (G2: 62,27; 263; G0: 14,25; 151) ili *predškolskoga odgoja* koji se prevodi kao *istruzione prescolare* (G2: 35,16; 37; G4: 20,24; 25; G6: 58,23; 157) i *istruzione prescolastica* (G6: 34,54; 92), ali i *educazio-ne prescolare* (G4: 16,81; 21; G6: 15,97; 42). Neujednačenost u uporabi nazivlja pronađena je i unutar jedne institucije, poput *enti prescolari* (G6: 18,5; 48) i *enti prescolastici* (G6: 14,49; 37) za *predškolske ustanove* ili *ospedale generale* (G0 42,78; 125), *ospedale pubblico* (G0: 20,69; 66) za *opću bolnicu*. U prva tri primjera riječ je o istovrijednicama i njihova uporaba, osim što nije poželjna zbog težnje ujednačenomu nazivlju, ne utječe na jasnoću prenesene poruke. Za posljednji bi primjer, koji se također pojavljuje u prevedenim talijanskim teksto-

vima kao istovrijednica, bila potrebna i pojmovna analiza kako bi se utvrdilo podudaranje pojmova i prikladnost uporabe tih istovrijednica. Iako su upotrijebljeni nazivi jasni govornicima talijanskoga kao manjinskoga jezika jer se odnose na hrvatski zdravstveni sustav, mogli bi biti nejasni govornicima talijanskoga jezika jer se u današnjemu talijanskom zdravstvenom sustavu takvi nazivi rijetko rabe. Upravo je posljednji primjer ukazao, osim na uporabu neujednačenoga nazivlja, i na problem prevođenja na manjinski jezik. Taj je problem vidljiv i u prevođenju naziva *zdravstveno osiguranje*. Ono se prevodi kao *assicurazione sanitaria* (G4: 256,94; 545; G0: 36,73; 34; G2: 39,8; 54) i kao *previdenza sanitaria* (G6: 139,31; 384). Primjeri istoga konteksta uporabe pronađeni su u različitim potkorpusima, što potvrđuje to da se nazivi upotrebljavaju kao istovrijednice. Taj primjer pokazuje poteškoće u prevođenju na manjinski jezik koji je nacionalni jezik u drugoj državi jer to zahtijeva duboko poznavanje obaju pravnih i drugih sustava kako bi se uspostavila podudarnost između dvaju pojmova (usp. Chiochetti i dr. 2016). Pri prevođenju toga primjera treba provesti temeljitu terminološku analizu, a tek dublje poznavanje zdravstvenoga sustava omogućilo bi uočavanje moguće leksičke i pojmovne neusklađenosti naziva. Međutim, temeljita terminološka analiza nije nužna ako već postoji službeni prijevod koji se preuzima iz službenih izvora. U ovome slučaju službeni prijevod na talijanski jezik već postoji i objavljen je na službenim mrežnim stranicama Hrvatskoga zavoda za zdravstveno osiguranje,¹⁰ gdje stoji odabrana istovrijednica uz pojašnjenje na talijanskome jeziku. To vrijedi i za službeno objavljene dvojezične nazive institucija koji se ne bi trebali prevoditi drugačije, no u korpusu su pronađene različite inačice prijevoda. Primjer je za to naziv *Hrvatski zavod za javno zdravstvo Istarske županije*¹¹ službeno preveden kao *Istituto di sanità pubblica della Regione istriana*, a koji se prevodi i kao *Ente per la salute (pubblica)* (G0: 8,59; 21, G4: 5¹²), *Ente per la sanità (pubblica)* (G0: 10,8; 27) i *Istituto regionale per la salute pubblica* (G3: 6, G: 5). Samo je dio primjera pronađen na popisu za nazive dobiven automatskim crpenjem naziva, a druge su inačice pronađene izravnim pretragom u potkorpusima te je u zagradi naveden samo podatak o čestoti pojavljivanja u tome potkorpusu. U potkorpusu G0 uočena je uporaba dvaju različitih naziva čak i unutar iste institucije i to *Ente per la salute pubblica*

¹⁰ www.hzzo.hr/it/la-tessera-europea-di-assicurazione-malattia/ (pristupljeno 20. rujna 2019.).

¹¹ www.zzjziz.hr/index.php?id=2 (pristupljeno 20. rujna 2019.).

¹² Ti nazivi pronađeni su izravnim pretragom u potkorpusu i navedena im je samo čestota pojavljivanja jer nisu pronađeni na popisu među tisuću kandidata za nazive dobivene automatskim crpenjem naziva.

della Regione i Ente per la sanità pubblica della RI. Taj primjer zorno prikazuje nepostojanje dogovora prevoditelja i nedostatak ili neuporaba terminološke baze koja bi jamčila uporabu ujednačenoga nazivlja. Još je jedan primjer neusklađene uporabe nazivlja unutar iste institucije i naziv *Istarski domovi zdravlja*, za koji su pronađeni sljedeći prijevodi: *Case di salute istriane* (G0: 10,44; 26) i *Case della salute dell'Istria* (G0: 15,52; 40), koji su pronađeni na popisu naziva, te *Case istriane della salute* (G0: 122), *Case della salute istriane* (G0: 14) i *Casa della salute istriana* (G0: 1), koji su pronađeni izravnom pretragom u potkorpusu G0. Potvrde navedenih primjera donosi slika 3.

4 a del Consiglio d'amministrazione delle	Case istriane della salute	a conformare le attività dell'istituto
4 ute a conformare le attività dell'istituto	Case istriane della salute	alle disposizioni della Legge sulla t
15.000,00 0,00 0,00 Arredo dei vani	Casa della salute istriana	– aiuto medico urgente di Umago 3 USCI
IV Ante Ivančić, direttore della	Case di salute istriane	, viene nominato a membro del Comand
glio d'amministrazione dell'ente sanitario	Case di salute istriane	sulla concessione del diritto di servitù a t
irezione, Sezione per la medicina sociale;	Case della salute istriane	– servizio infermiere a domicilio, medici di
extraospedaliera della tossicodipendenza;	Case della salute istriane	– Ambulatori per la tossicodipendenza ed
Delibera del Consiglio d'amministrazione delle	Case della salute	dell'Istria sull'indebitamento per acquistare me
te DELIBERA sul consenso alla Delibera delle	Case della salute	dell'Istria sulla scelta dell'offerta migliore per la

Slika 3. Potvrde istovrijednica za naziv *Istarski domovi zdravlja* u potkorpusu G0

Provjereno je također je li se naziv na hrvatskome jeziku s vremenom mijenjao kako bi se opravdala terminološka raznolikost na talijanskome jeziku, no pretragom korpusa nisu pronađene različite inačice toga naziva.

Valja napomenuti i da na mrežnim stranicama Istarskih domova zdravlja postoji službeni talijanski naziv *Case della salute dell'Istria*, koji će se, pretpostavlja se, u budućim prijevodima rabiti te se u talijanskim tekstovima više neće susretati neprihvatljiva kreativnost i sloboda u prevođenju naziva institucija za koje već postoji službeni prijevod.

Sklonost prevoditelja zadržavanju zastarjeloga naziva i neprilagođavanju promjenama koje su uslijedile u nazivu na izvornome jeziku pokazuje primjer prijevoda za naziv *gradsko vijeće*. Pri prevođenju toga naziva na talijanski jezik neki se prevoditelji oslanjaju na stari naziv *općinsko vijeće*, pa se tako u nekim tek-

Osim potvrde pojavljivanja različitih prijevodnih rješenja, na slici 2. mogu se uočiti i različite inačice toga naziva. Pri pisanju su naziva, naime, vidljiva različita pravopisna rješenja (u ovome slučaju različito je pisanje velikoga i maloga početnog slova), što dodatno ukazuje na nužnost uporabe prevoditeljskih alata i jezičnih izvora koji bi riješili pravopisne nedoumice prevoditelja i umanjili mogućnost pojavljivanja nepotrebne terminološke raznolikosti.

Iz ove analize nazivlja očita je potreba za normiranjem nazivlja na talijanskome jeziku koje se odnosi na specifičnosti koje talijanskomu jeziku donosi status manjinskoga jezika u Hrvatskoj. Ta se potreba očituje najviše u područjima koja su pod mjesnom samoupravom i onima koja izravno ovise o hrvatskome pravnom sustavu i razlikuju se od područja u Italiji. Najpotrebnije je stoga usustavljanje nazivlja javne uprave, zdravstva, socijalnoga rada, odgoja i obrazovanja, ali i drugih područja kako ne bi došlo do nesporazuma u komunikaciji na talijanskome jeziku zbog nepodudaranja u pojmovima i zbog sličnosti ili razlika u nazivima koji se upotrebljavaju u talijanskome jeziku u Italiji i u talijanskome kao manjinskome jeziku u Hrvatskoj. Nazivlje na talijanskome jeziku trebalo bi biti jasno svim govornicima talijanskoga jezika, a posebno bi trebalo odgovarati potrebama talijanskoga kao manjinskoga jezika u Istarskoj županiji. Terminološka nepreciznost i raznolikost, zapažena zahvaljujući uporabi korpusnolingvističkoga pristupa i prethodnoj pripremi usporednoga korpusa, osim na potrebu za temeljitom pojmovnom analizom i normiranjem nazivlja upućuje i na potrebu izrade terminoloških izvora i uporabe alata za strojno potpomognuto prevođenje koji bi olakšali prevođenje i umanjili mogućnost uporabe neujednačenoga nazivlja. Isto tako, pokazalo se da prevoditeljima nedostaju smjernice, organizacija prevođenja ili drugi oblici umrežavanja koji bi zahtijevali od prevoditelja uporabu zajedničkih izvora i alata te doprinijeli uporabi ujednačenoga nazivlja radi stvaranja prijevoda visoke kvalitete koji bi u konačnici bili jasni svim govornicima talijanskoga jezika.

3. Zaključak

Ako se uzme u obzir to da je od automatski dohvaćenih službenih novina s odabranih mrežnih stranica samo 37 % službenih dokumenata dostupno i na

talijanskome jeziku te da je analiza sadržaja na službenim mrežnim stranicama dvojezičnih gradova i općina pokazala da se veći dio sadržaja ne prevodi, ne prevodi pravodobno ili se prevodi samo djelomično, očito je da nedostaje prevoditelja ili financijskih sredstava za prevođenje. Međutim, ako se uzmu u obzir primjeri sličnih dvojezičnih i višejezičnih institucija u Europi, jasno je da je prevoditeljima neophodna i podrška jezičnih tehnologija, alata i izvora koji bi olakšali i ubrzali prevođenje te smanjili troškove prevođenja. U Istarskoj županiji, do početka projekta u sklopu kojega je i ovaj rad napisan, nije bilo inicijativa kojima bi cilj bio razvoj jezičnih tehnologija. Ovim projektom sastavljen je prvi usporadni korpus, koji se, osim za lingvističke analize, može upotrebljavati za strojno potpomognuto prevođenje, a budući da će se nastavljati nadograđivati, služiti će i kao važan izvor i u razvoju strojnoga prevođenja. Analiza postojećega nazivlja provedena na korpusu pokazala je da je potrebno sustavno planirati izgradnju, provjeravanje i normiranje talijanskoga nazivlja hrvatskoga pravnog sustava i svih njemu podređenih sustava. Budući da je riječ o nazivlju na manjinskome jeziku u Hrvatskoj, koji ima status nacionalnoga jezika u Italiji, neophodno je pri izgradnji, usklađivanju i obradi nazivlja uzeti u obzir razlike između dvaju različitih pravnih sustava kako bi komunikacija prema korisnicima bila jasna i nedvosmislena i kako bi se, gdje je potrebno, jasno razlikovale posebnosti dviju država. Neophodno je stoga usklađivanje postojećega nazivlja i izrada terminološke baze. Uporaba zajedničkih jezičnih izvora i alata u Istarskoj županiji umanjila bi mogućnost uporabe neujednačenoga nazivlja i uklonila uočenu terminološku raznolikost nepoželjnu za administrativni funkcionalni stil i ujedno doprinijela bržemu i uspješnijemu prevođenju.

Literatura

- ALEGRIA, INAKI i dr. 2005. Open Architecture for Transfer-based Machine Translation between Spanish and Basque. *Proceedings of the MT Summit X Workshop on Open-Source Machine Translation*. Asia-Pacific Association for Machine Translation (AAMT). Phuket. 7–14. <https://pdfs.semanticscholar.org/d346/7010dd32f2f317f66cdc0bb532fcb045a97b.pdf> (pristupljeno 30. rujna 2019.).
- ARRANZ, VICTORIA; COMELLES, ELISABET; FARWELL, DAVID. 2006. Speech-to-Speech Translation for Catalan. *Proceedings of the Lesser Used Languages and Computer Linguistics Conference (LULCL 2005)*. Ur. Ties, Isabella. Accademia Europea Bolzano. Bolzano.

BAISA, VIT i dr. 2016. European Union Language Resources in Sketch Engine. *The Proceedings of tenth International Conference on Language Resources and Evaluation (LREC 16)*. European Language Resources Association (ELRA). Portorož. 2799–2803. www.aclweb.org/anthology/L16-1445 (pristupljeno 30. rujna 2019.).

BRKIĆ BAKARIĆ, MARIJA; LALLI PAČELAT, IVANA. 2019. Parallel Corpus of Croatian-Italian Administrative Texts. *Proceedings of the Workshop Human-Informed Translation and Interpreting Technology*. Association for Computational Linguistics. Shoumen. doi.org/10.26615/issn.2683-0078.2019_002.

CALZOLARI, NICOLETTA i dr. 2012. *The Italian language in the digital age / La lingua italiana nell'era digitale*. Springer Verlag. Berlin. www.meta-net.eu/whitepapers/e-book/italian.pdf (pristupljeno 30. rujna 2019.).

CHIOCCHETTI, ELENA; RALLI, NATASCIA; STANIZZI, ISABELLA. 2006. *Normazione: aspetti giuridici e linguistici*. MediAzioni: Rivista online di studi interdisciplinari su lingue e culture. www.mediazioni.sitlec.unibo.it/images/stories/PDF_folder/document-pdf/3%20chiocchetti.pdf (pristupljeno 30. rujna 2019.).

ČERMÁK, FRANTI; ROSEN, ALEKSANDR. 2012. The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics* 17/3. 411–427. doi.org/10.1075/ijcl.17.3.05cer.

EUROPSKA KOMISIJA. 2016. *Translation tools and workflow*. Publication Office of the European Union. Luxemburg. doi.org/10.2782/703257.

FRANČIĆ, ANĐELA; HUDEČEK, LANA; MIHALJEVIĆ, MILICA. 2006. *Normativnost i višefunkcionalnost u hrvatskome jeziku*. Hrvatska sveučilišna naklada. Zagreb.

IVANETIĆ, NADA. 2003. *Uporabni tekstovi*. Zavod za lingvistiku Filozofskog fakulteta Sveučilišta u Zagrebu. Zagreb.

JONES, DEWI B.; PRYS, DELYTH. 2006. The Welsh National Online Terminology Database. *Proceedings of the Lesser Used Languages and Computer Linguistics Conference (LULCL 2005)*. Ur. Ties, Isabella. Accademia Europea Bolzano. Bolzano. 149–169.

KENNING, MARIE-MADELEINE. 2010. What are parallel and comparable corpora and how can we use them?. *The Routledge handbook of corpus linguistics*. Ur. O'Keeffe, Anne; McCarthy, Michael. Routledge. London. 487–500. doi.org/10.4324/9780203856949.ch35.

KOEVA, SVETLA i dr. 2012. Bulgarian X-language Parallel Corpus. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*. European Language Resources Association (ELRA). Istanbul. 2480–2486. www.lrec-conf.org/proceedings/lrec2012/pdf/587_Paper.pdf (pristupljeno 30. rujna 2019.).

LALLI PAČELAT, IVANA. 2013. L'italiano adattato all'amministrazione croata: problem e prospettive del bilinguismo amministrativo nella Regione Istriana. *Italica Belgradensia. Atti del Convegno Internazionale Oltre i confini: aspetti transregionali e interculturali dell'italiano*. Ur. Vučo, Julijana; Milinković, Snežana. Filološki fakultet Univerziteta. Beograd. 148–161.

McENERY, TONY; XIAO, RICHARD. 2007. Parallel and comparable corpora: What are they up to?. *Incorporating Corpora: Translation and the Linguist* (Translating Europe). Ur. Andermann M., Gunilla; Rogers, Margaret. Multilingual Matters. Clevedon. 1–13.

HRVATSKI ZAVOD ZA STATISTIKU. 2013. *Popis 2011: Jer zemlju čine ljudi – Popis stanovništva, kućanstva i stanova 2011. Stanovništvo prema državljanstvu, narodnosti, vjeri i materinskom jeziku*. Statistička izvješća. www.dzs.hr/Hrv_Eng/publication/2012/SI-1469.pdf (pristupljeno 30. rujna 2019.).

OCH, FRANZ J.; NEY, HERMANN. 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics* 29/1. 19–51. doi.org/10.1162/089120103321337421.

SCHERRER, YVES; CARTONI, BRUNO. 2012. The Trilingual ALLEGRA Corpus: Presentation and Possible Use for Lexicon Induction. *Proceedings of the 8th international conference on Language Resources and Evaluation (LREC 2012)*. Ur. Calzolari, Nicoletta. European Language Resources Association (ELRA). Istanbul. 2890–2896. www.lrec-conf.org/proceedings/lrec2012/pdf/685_Paper.pdf (pristupljeno 30. rujna 2019.).

SOMERS, HAROLD. 2001. Bilingual Parallel Corpora and Language Engineering. *Proceedings of the Workshop on Language Engineering for South-Asian Languages*. 1–16. www.emille.lancs.ac.uk/lesal/somers.pdf (pristupljeno 30. rujna 2019.).

STEINBERGER, RALF i dr. 2006. The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*. Genova. 2142–2147. https://arxiv.org/ftp/cs/papers/0609/0609058.pdf (pristupljeno 8. listopada 2020.).

STEINBERGER, RALF i dr. 2012. DGT-TM: A Freely Available Translation Memory in 22 Languages. *Proceedings of the 8th international conference on Language Resources and Evaluation (LREC 2012)*. European Language Resources Association (ELRA). Istanbul. 454–459. www.lrec-conf.org/proceedings/lrec2012/pdf/814_Paper.pdf (pristupljeno 30. rujna 2019.).

STEINBERGER, RALF i dr. 2014. An Overview of the European Union's highly multilingual parallel corpora. *Language Resources and Evaluation* 48/4. 679–707. doi.org/10.1007/s10579-014-9277-0. STEINBERGER, RALF i dr. 2014. An Overview of the European Union's highly multilingual parallel corpora. *Language Resources and Evaluation* 48/4. 679–707. doi.org/10.1007/s10579-014-9277-0.

STREITER, OLIVER; STUFLESSER, MATHIAS; TIES, ISABELLA. 2004. Cle, an aligned tri-lingual latin-italian-german corpus. Corpus design and interface. *First Steps in Language Documentation for Minority Languages*. Ur. Carson-Berndsen, Julie. SALT MIL. Lisbon. 84–87.

TADIĆ, MARKO; BROZOVIĆ-RONČEVIĆ, DUNJA; KAPETANOVIĆ, AMIR. 2012. *The Croatian language in the digital age / Hrvatski jezik u digitalnom dobu*. Springer Verlag. Heidelberg. www.meta-net.eu/whitepapers/e-book/croatian.pdf (pristupljeno 30. rujna 2019.).

TIEDMANN, JÖRG. 2012. Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*. European Language Resources Association (ELRA). Istanbul. 2214–2218. www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf (pristupljeno 30. rujna 2019.).

TROSTERUD, TROND. 2002. Parallel corpora as tools for investigating and developing minority languages. *Parallel corpora, Parallel worlds* (Language and computers: studies in practical linguistics 43). Ur. Borin, Lars. Rodopi. Amsterdam. 111–122. doi.org/10.1163/9789004334298_007.

VARGA, DÁNIEL i dr. 2007. PARALLEL CORPORA FOR MEDIUM DENSITY LANGUAGES. *Recent Advances in Natural Language Processing IV: Selected papers from RANLP 2005*. Ur. Nicolov, Nicolas i dr. John Benjamins Publishing Company. Amsterdam. 247–258. doi.org/10.1075/cilt.292.32var.

ZIEMSKI, MICHAŁ; JUNCZYS-DOWMUNT, MARCIN; POULIQUEN, BRUNO. 2016. The United Nations parallel corpus v1. 0. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Ur. Calzolari, Nicoletta. European Language Resources Association. Portorož. 3530–3534.

ZILLI, SILVANO (ur.). 2017. *La Comunità nazionale italiana negli statuti delle città e dei comuni bilingui della Regione Istriana – Talijanska nacionalna zajednica u statutima dvojezičnih gradova i općina u Istarskoj županiji*. Tutti i diritti sono riservati a norma di legge. Rovinj. <http://consiglio-cni-istria.hr/wp-content/uploads/2018/06/Knjiga-dvojezicnost-1.pdf> (pristupljeno 2. veljače 2020.).

The Official Bilingualism in the Istrian County: State of the Art and Perspectives

Abstract

Official bilingualism assumes the daily creation of parallel texts in bilingual areas. This is also the case in Istria County, where texts are usually written in Croatian and then translated into Italian. Given the official nature of texts and the context of the Italian language usage, it is extremely important to have a precise and uniform terminology, as well as well-developed language technologies, which would allow a faster and more accurate translation processes in bilingual institutions of Croatia. The aims of this paper are (1) to investigate the equal use of both Croatian and Italian as a minority language in the Istrian County through (1a) an analysis of the translated content on the official websites of the officially bilingual cities and municipalities and through (1b) an overview of current translation practices, and (2) to highlight the importance and necessity of the development of language technology through (2a) the presentation of trends in the development of language technology in similar bilingual and multilingual institutions,

(2b) the presentation of the parallel corpus compiled from administrative texts of the Istrian County and (2c) the analysis of existing terminology conducted on the compiled corpus. The analysis of the availability of bilingual content on websites and the terminology extracted from the compiled parallel corpus indicates the need to develop and use translation tools and linguistic resources tailored to the Italian language as a minority language, in order to facilitate and accelerate the translation activity, thus allowing a more efficient use of Italian as a minority language.

Ključne riječi: jezične tehnologije, prevođenje, dvojezičnost, usporedni korpusi

Keywords: language technology, translation, bilingualism, parallel corpora

