

Nataša Logar
Fakulteta za družbene vede, Ljubljana
UDK 811.163.6(038):004.738.52

Jezikovni viri na klik in dotik

V prispevku razmišljamo o jezikovnih virih in tehnologijah, ki so postali nujni del opremljenosti vsakega jezika v digitalni dobi. Na izbirnem tečaju bomo na konkretnih primerih in problemsko spoznavali v prispevku našete ter druge vire, orodja in aplikacije, pa tudi za različne namene uporabne spletne strani o slovenščini.

jezikovni viri, jezikovne tehnologije, digitalna doba, slovenščina, prihodnost jezika

The paper discusses language resources and technologies that have become a necessary part of any language in the digital age. During the course we will examine language resources, tools and applications mentioned in the paper and others, as well different websites on Slovene that are useful for a variety of purposes.

language resources, language technologies, digital age, Slovene, language future

1 Uvod

V letih 2002 in 2003 je UNESCO zaprosil mednarodno skupino jezikoslovcev, naj oblikuje referenčni okvir, po katerem bi bilo mogoče določiti vitalnost jezikov. Nastal je dokument *Jezikovna vitalnost in ogroženost (Language Vitality and Endangerment, 2003)*, v katerem je kot osnova za oceno vitalnosti jezikov opredeljenih devet kriterijev. Med njimi so *absolutno število govorcev, medgeneracijski jezikovni prenos, odnos govorcev do lastnega jezika* itn. Eden od devetih kriterijev je tudi *odziv na nova področja rabe in nove medije*.

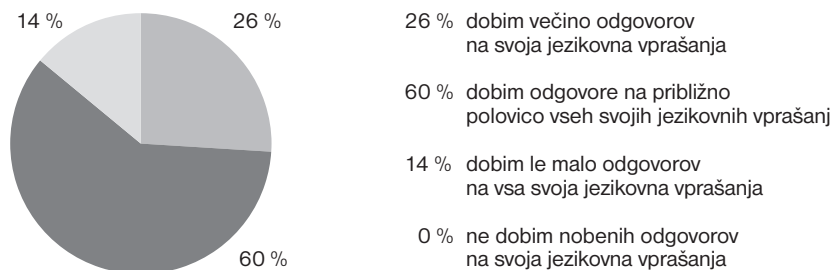
Od takrat je minilo le deset let, a v teh desetih letih je postal ta kriterij ključen, natančneje: vitalnost jezikov se je začela meriti po njihovi prisotnosti na internetu – celo še ožje: na Wikipediji. Prim. Hladnik (7. maj 2013: 8): »S kvaliteto [...] člankov [na Wikipediji] se meri vitalnost jezikov in njihova sposobnost preživetja. Jeziki brez wikiskupnosti so obsojeni na izumrtje.«

2 Prihodnost: jezikovne potrebe mladih govorcev

Ko govorimo o prihodnosti jezika, imamo v mislih predvsem govorce, ki so danes otroci in mladostniki. Generacije, rojene po letu 1990, so vstopale v digitalni svet brez priročnikov za uporabo ali posebnih tečajev, klikanje in komuniciranje z napravami prek dotika je zanje samoumevno ter intuitivno. Kakršnokoli iskanje informacij, ki zahteva večji napor, je le še prisilna izbira, hipnost odgovora pa ima prednost pred njegovo zanesljivostjo. Da na ta način iščejo tudi odgovore na jezikovna vprašanja, smo potrdili z anketo.

Anketo smo izvedli februarja in marca 2012, in sicer med študenti Fakultete za družbene vede (FDV) in Filozofske fakultete v Ljubljani (FF). Šlo je za študente tržnega komuniciranja in odnosov z javnostmi (1. in 2. letnik dodiplomskega študija; 75 študentov) ter študente medjezikovnega posredovanja (1., 2. in 3. letnik dodiplomskega študija; 99 študentov).

Vsi študenti so na vprašanje *Ali ste v zadnjih 6 mesecih iskali podatek o zapisu besede, prevod ali kateri drug jezikovni podatek z iskalnikom na svetovnem spletu (npr. z google.com, najdi.si ipd.)?* odgovorili z »da«. Slika 1 nadalje prikazuje deleže odgovorov, ki smo jih dobili, ko smo študente prosili, naj izberejo enega od zaključkov trditve: *Že zgolj z zadetki, ki jih dobim z uporabo iskalnika po celotnem spletu ...*



Slika 1: Zadovoljstvo študentov z zadetki, ki jih dobijo na celotnem spletu, ko tam iščejo odgovore na svoja jezikovna vprašanja

Študenti so bili torej v 86 % v polovici ali celo večini primerov zadovoljni z odgovori na jezikovna vprašanja, ki jim jih je ponudil *google.com* ali kateri drugi brskalnik, pri čemer so tam iskali predvsem razlago besed (91 %) in prevode (83 %), nekoliko manjkrat pravilen zapis (68 %) ter v dobri četrtini primerov slovnične podatke (26 %). Seveda so nam potrdili, da odgovore na jezikovna vprašanja iščejo tudi po *slovarjih* na spletu, pri čemer to počnejo tudi preko pametnih telefonov, dlančnikov, tabličnih računalnikov ali podobnih naprav (torej ne le prek osebnih računalnikov), in sicer v 15 % (študenti FDV) oz. 23 % (študenti FF). Gre za dve leti star podatek in predvidevati je mogoče, da bi bili danes deleži teh odgovorov večji.

Poleg *načina* dostopanja do odgovorov na jezikovna vprašanja na svetovnem spletu, ki ga je pokazala naša anketa, pa so pomembne tudi *potrebe in pričakovanja* mladih govorcev glede spletnih jezikovnih virov. Te je pokazala nedavna anketa, ki je bila prav tako izvedena med študenti Oddelka za prevajalstvo FF in študenti FDV, v tem primeru študenti novinarstva. Rezultati so bili predstavljeni na *Posvetu o novem slovarju sodobnega slovenskega jezika*, ki ga je 12. februarja 2014 organiziralo Ministrstvo za kulturo RS, in bodo objavljeni v Gorjanc (2014). Pokazalo se je, da se mladi zelo dobro zavedajo multimedijskosti, hiperpovezovanosti, količinske neomejenosti, možnosti stalnega posodabljanja podatkov in drugih lastnosti spletne forme, da prilagoditev tem lastnostim pričakujejo tudi od jezikovnih virov (vprašanja so se nanašala na prihodnji slovar slovenščine) ter da so v tem pogledu izredno zahtevni uporabniki.

3 Stanje: jezikovna opremljenost

Opremljenosti evropskih jezikov za digitalno dobo je posebno raziskavo leta 2012 posvetila Mreža odličnosti, ki jo sestavlja Tehnološko združenje za večjezično Evropo – na kratko: META-NET (www.meta-net.eu). Gre za povezavo raziskovalcev, komercialnih ponudnikov tehnologij,

zasebnih in korporativnih uporabnikov jezikovnih tehnologij, poklicnih strokovnjakov na področju jezikov ter drugih nosilcev interesov na področju informacijske družbe iz (trenutno) 34 držav, njihov namen pa je »nujno potrebna pospešitev razvoja jezikovnih tehnologij kot sredstva za udejanjenje vizije Evrope, združene v en sam digitalni trg in informacijski prostor« (prim. tudi CLARIN, www.clarin.eu; zasnova CLARIN-a je opisana v Váradi idr. 2008). Raziskava združenja META-NET, objavljena kot serija belih knjig s skupnim naslovom *Evropski jeziki v digitalni dobi*, je pokrila 31 evropskih jezikov. Ključna ugotovitev raziskave je bila naslednja: nobeden od jezikov ni *odlično* jezikovnotehnološko podprt, po posameznih segmentih (strojno prevajanje, procesiranje govora, besedilna analiza, govorni in besedilni viri) je dobro podprta le angleščina. Naslednjo, manjšo skupino sestavljajo jeziki, ki so *zmerno* dobro podprti (predvsem francoščina, španščina, nemščina, italijanščina, nizozemščina, češčina), preostala velika skupina pa je bodisi le *fragmentarno podprta* bodisi *zelo slabo podprta* oz. sploh *nepodprta*. Slovenščina je v družbi slednjih: po treh jezikovnotehnoloških segmentih je le fragmentarno podprta, pri strojnem prevajanju pa celo zelo slabo podprta (META-NET; Krek 2012a; Krek 2012b).

Kot je sintetično prikazal Krek (2013), so mnogi evropski jeziki prek financiranja v okviru Evropske unije svojo osnovno jezikovno infrastrukturo (leksikoni, korpusi, jezikovni viri nasploh) zgradili že v devetdesetih letih 20. stoletja. Slovenščina se je tem jezikom priključila s sodelovanjem Tomaža Erjavca z Instituta "Jožef Stefan" v projektu MULTEXT-EAST v letih 1995–1997, nato pa s prvim referenčnim korpusom, tj. korpusom FIDA, leta 2000 (Erjavec idr. 1998).

Kljub slabi celostni jezikovnotehnološki podprtosti slovenščine, ki smo jo izpostavili v prvem odstavku tega poglavja, pa velja poudariti, da je do danes nastalo že kar nekaj jezikovnotehnoloških orodij in aplikacij za procesiranje slovenskega jezika, od katerih jih je večina že tudi prosto dostopna: označevalnika ToTaLe in Obeliks (Erjavec idr. 2005; Grčar idr. 2012), skladenjski razčlenjevalnik (Dobrovoljc idr. 2012), aplikacija za pridobivanje dobrih zgledov iz korpusov GDEX (Kosem idr. 2011), slovarski urejevalnik Termania (www.termania.net; Romih, Krek 2012), sintetizator govora (Amebis, Govorec), luščilnik terminoloških kandidatov (Vintar 2010), luščilnik definicij iz specializiranih korpusov (Pollak 2014) itn. Seveda pa za laične uporabnike našeta orodja niso zares uporabna. Kar rabijo in pričakujejo tisti, ki bodo sicer odgovore na jezikovna vprašanja poiskali na Googlu, so jezikovni opisi ter priročniki, črkovalniki, slovnični pregledovalniki in prevajalniki ter terminološki portali z enim iskalnim okencem in dobro možnostjo filtriranja zadetkov. Vmesna stopnja do nekaterih od njih so korpusi (referenčni pisni in govorni), leksikoni besednih oblik in leksikalne baze. Od teh so že na voljo npr. korpus Gigafida in korpus Kres (Logar Berginc idr. 2012), Korpus Gos (Verdonik, Zwitter Vitez 2011), Besana in Sloleks (Holozan idr. 1991; Erjavec idr. 2008), leksikalna baza za slovenščino v omejenem obsegu 2.500 gesel (Gantar 2011) in slovenski wordnet sloWNet (Fišer idr. 2013).

4 Načrti: jezikovni viri za uporabnike vseh generacij

Osrednje mesto v slovenističnih razpravljanih ima v zadnjem letu nov slovar slovenščine. Za del jezikoslovcev ni dvoma: novi temeljni leksikalni opis slovenskega jezika mora biti metodološko zasnovan za digitalni medij, izdelan z najnovejšimi leksikografskimi orodji in aplikacijami ter brezplačno spletno dostopen (prim. www.sssj.si; Krek idr. 2013). Naslednji jezikovni opis, ki bi ga

najprej potrebovali jezikoslovna znanost in stroka, v poljudni obliki pa nato še vsa druga javnost, je posodobljena slovenska slovnica. Zgled, kako je mogoče slovnico za učne potrebe osnovnošolcev in dijakov zasnovati na spletu, je bil že oblikovan (www.slovenscina.eu/portali/pedagoski-slovnici-portal; Arhar Holdt 2012), za široko uporabnost pa bi bilo treba tam obstoječim 24 temam čim prej začeti dodajati nove. Odgovore na različne pravopisno-stilistične zadrege je na spletu mogoče najti v jezikovnokotičkarskih in podobnih svetovalnicah (Hribar 2009; Marušič, Žaucer 2009), ki pa so razpršene ter večinoma odvisne od samoiniciativnosti in dobre volje izvajalcev. Izrecno je bil kot tovrstni spletni priročnik zasnovan *Slogovni priročnik* (<http://slogovni.slovenscina.eu>; Krek idr. 2013), a zaenkrat ostaja v demonstracijski različici v obsegu 15 gesel. Tudi slovnici pregledovalnik za slovenščino že obstaja (Amebis, Besana), ni pa še na voljo v nekomercialni obliki. Dobre zglede torej že imamo, a zgolj ti za zadovoljstvo s tovrstno opremljenostjo slovenščine nikakor ne zadostujejo.

Resolucija o Nacionalnem programu za jezikovno politiko 2014–2018, sprejeta v Državnem zboru RS 15. julija 2013, v poglavju 2.2 *Jezikovna opremljenost* predvideva številne ukrepe, s katerimi bi lahko bili preseženi šibka jezikovnotehnološka podprtost slovenščine in slaba opremljenost z jezikovnimi viri, ki jih govorci slovenščine, pa tudi tisti, ki se slovenščine učijo kot tujega jezika, potrebujejo že danes. Osrednji cilj tega poglavja je zato sprejetje posebnega *Akcijskega načrta za jezikovno opremljenost*, ki bo konkretiziral »izzive, ki so potrebni hitrega in učinkovitega ukrepanja.« Besedilo akcijskega načrta je v času oddaje tega prispevka že pripravljeno za javno obravnavo; če se bo uresničil vsaj njegov dokajšnji del, so obeti dobri.

5 Sklep

Revija *Forbes* je 21. marca 2014 objavila podatek, da bo svetovni jezik prihodnosti – francoščina. Šele na drugem in tretjem mestu naj bi bili angleščina in mandarinščina. Študijo je sicer opravila investicijska banka in sloni na nekoliko vprašljivi metodologiji, zato je že sam avtor prispevka do podatka skeptičen (gl. tudi Richmond, 26. marec 2014). Vendarle pa je objava povedna na drugi način: tako zelo samoumevno se nam zdi, da bo angleščina lingua franca tudi leta 2050, da nas vsakršen drugačen podatek predvsem začudi.

Enako kot o svetovnem jeziku prihodnosti ni veliko dvoma (ali pa vsaj ne veliko alternativnih odgovorov; prim. npr. Graddol 2004), tudi v jezikovnonačrtovalskih premislekih o prihodnosti posameznih jezikov na področju jezikovne opremljenosti ni veliko različnih ciljev. Kot smo videli, je cilj lahko celo en sam: slediti angleščini in vsak jezik opremiti z infrastrukturo, ki bo v prihodnosti »predvsem digitalna, spletna in mobilna« (Resolucija o Nacionalnem programu za jezikovno politiko 2014–2018: 23).

Če vitalnost slovenščine merimo po njeni prisotnosti na Wikipediji, imamo, kot je zapisal Hladnik (7. maj 2013), razlog za veselje:

- slovenska Wikipedija s 500 aktivnimi wikipedisti (proti 15.000 angleškim) in 133.000 članki (proti 3,9 mio. v angleščini) /je/ samo 30-krat manjša od angleške; populacijska (400 mio. : 2 mio.) in ekonomska razmerja kažejo za slovenščino in Slovenijo veliko večji razkorak
- slovenščina je na 34. mestu med 283 jeziki, glede na število govorcev bi bilo /potemtakem/ njeno pravično mesto 150.

Vendar pa je slovenski jezik treba tudi na vseh drugih področjih opremiti novim družbenim in tehnološkim okoliščinam primerno ter govorcem slovenščine o njem ponuditi raznovrstne informacije na lahko dostopen način in v ustrezno vizualizirani obliki – in to govorcem vseh generacij. Računalnik namreč redno uporablja že vsak tretji Slovenec, starejši od 60 let. Ali kot je povedala 73-letna upokojenka v prispevku na Valu 202 (20. oktober 2012):

Z računalnikom imaš cel svet v svoji dnevni sobi. [...] Če bi šla rada v galerijo, odprem Louvre, poiščem sobo z impresionisti in se sprehajam med njimi. Na ta način sem obnovila tako angleščino kot nemščino. Štos je pa v tem, da ti ni treba iti gledat v slovar, ker imaš slovar na računalniku.

Literatura

- ARHAR HOLDT, Špela, 2012: *Pedagoška korpusna slovnica*. www.rtvsllo.si
CLARIN: www.clarin.eu
- DOBROVOLJC, Kaja, KREK, Simon, RUPNIK, Jan, 2012: Skladenjski razčlenjevalnik za slovenščino. Tomaž Erjavec, Jerneja Žganec Gros (ur.): *Zbornik Osme konference Jezikovne tehnologije*. Ljubljana: Institut "Jožef Stefan". 42–47.
- ERJAVEC, Tomaž, GORJANC, Vojko, STABEJ, Marko, 1998: Korpus FIDA. *Jezikovne tehnologije za slovenski jezik/ Mednarodna multikonferenca Informacijska družba – IS'98*. Ljubljana: Institut "Jožef Stefan". 124–127.
- ERJAVEC, Tomaž idr., 2005: Massive Multi-lingual Corpus Compilation: Acquis Communautaire and ToTaLe. *Proceedings of the 2nd Language & Technology Conference*. Poznan. 32–36.
- ERJAVEC, Tomaž idr., 2008: *Specifikacije za leksikon (besednih oblik) – projekt Sporazumevanje v slovenskem jeziku*. http://projekt.slovenscina.eu/Media/Kazalniki/Kazalnik3/SSJ_Kazalnik_3_Specifikacije-leksikon_v1.pdf
- FIŠER, Darja, PIASECKI, Maciej, BRODA, Bartosz, 2013: Grounding sloWNet on Slovene corpus data. *Slovenščina 2.0* 1/2. 82–112.
- Forbes, 21. marec 2014: *Want to Know the Language of the Future? The Data Suggests it Could Be...French*. <http://www.forbes.com/sites/pascalemanuelgobry/2014/03/21/want-to-know-the-language-of-the-future-the-data-suggests-it-could-be-french>
- GANTAR, Polona, 2011: Leksikalna baza za slovenščino: komu, zakaj in kako (naprej)? *Jezikoslovni zapiski* 17/2. 77–92.
- GORJANC, Vojko, 2014: Slovar slovenskega jezika v digitalni dobi. Prispevek na *Posvetu o novem slovarju sodobnega slovenskega jezika*. Ljubljana: Ministrstvo za kulturo RS, 12. februar 2014. V tisku.
- GRADDOL, David, 2004: The Future of Language. *Science* 303/5662. 1329–1331.
- GRČAR, Miha, KREK, Simon, DOBROVOLJC, Kaja, 2012: Obeliks: statistični oblikoskladenjski označevalnik in lematizator za slovenski jezik. Tomaž Erjavec, Jerneja Žganec Gros (ur.): *Zbornik Osme konference Jezikovne tehnologije*. Ljubljana: Institut "Jožef Stefan". 89–94.
- HLADNIK, Miran, 7. maj 2013: Wikipedija pod Slovenci. www.ff.uni-lj.si/oddelki/slovenistika/mh/Wikipedija_zajesenice.pptx; http://sl.wikiversity.org/wiki/Desetletnica_Wikipedije
- HOLOZAN, Peter, 1991: *Besana*. Ljubljana: Graf inženiring. <http://besana.amebis.si>
- HRIBAR, Nataša, 2009: Al' prav se piše ... – spletna razglabljanja o slovenskem jeziku. Marko Stabej (ur.): *Infrastruktura slovenščine in slovenistike. Obdobja 28*. Ljubljana: Znanstvena založba Filozofske fakultete. 171–176.
- KOSEM, Iztok, HUSAK, Miloš, MCCARTHY, Diana, 2011: GDEX for Slovene. *Proceedings of eLex 2011*. Ljubljana: Trojina, zavod za uporabno slovenistiko. 150–159.
- KREK, Simon, 2012a: *Predstavitev bele knjige Slovenski jezik v digitalni dobi*. http://videlectures.net/is2012_krek_bela_knjiga
- KREK, Simon, 2012b: *Slovenski jezik v digitalni dobi/The Slovene Language in the Digital Age*. www.meta-net.eu/whitepapers/e-book/slovene.pdf
- KREK, Simon, 2013: Vsebina projekta in rezultati. *Zaključna konferenca projekta Sporazumevanje v slovenskem jeziku* http://videlectures.net/zakljucnakonferencassj2013_krek_vsebina
- KREK, Simon idr., 2013: Online Style Guide for Slovene as a Language Resources Hub. *Proceedings of eLex 2013*. Ljubljana, Talin: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut. 379–391. http://eki.ee/elex2013/proceedings/eLex2013_26_Krek+etal.pdf
- KREK, Simon, KOSEM, Iztok, GANTAR, Polona, 2013: *Predlog za izdelavo Slovarja sodobnega slovenskega jezika*. www.sssj.si

- Language Vitality and Endangerment*, 2003. Pariz: UNESCO. <http://unesdoc.unesco.org/images/0018/001836/183699E.pdf>
- LOGAR BERGINČ, Nataša idr., 2012: *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. Ljubljana: Trojina, zavod za uporabno slovenistiko; Fakulteta za družbene vede.
- META-NET: www.meta-net.eu; www.meta-net.eu/whitepapers/key-results-and-cross-language-comparison
- POLLAK, Senja, 2014: *Polavtomatsko modeliranje področnega znanja iz večjezičnih korpusov: Doktorska disertacija*. Ljubljana: Filozofska fakulteta.
- Resolucija o Nacionalnem programu za jezikovno politiko 2014–2018*. www.uradni-list.si/1/objava.jsp?urlid=201362&stevilka=2475
- RICHMOND, Ben, 26. marec 2014: *The Language of the Future is French, a New Study Concludes*. <http://motherboard.vice.com/read/french-is-supposed-to-be-the-language-of-the-future-vraiment>
- ROMIH, Miro, KREK, Simon, 2012: Termania – prosto dostopni spletni slovarski portal. Tomaž Erjavec, Jerneja Žganec Gros (ur.): *Zbornik Osme konference Jezikovne tehnologije*. Ljubljana: Institut "Jožef Stefan". 163–166.
- Termania: www.termania.net
- Val 202 (20. oktober 2012): *Ko za računalnik sede 93-letnica ...* <http://val202.rtvlo.si/2012/10/reakcija-ko-za-racunalnik-sede-93-letnica>
- VÁRADI, Tamás idr., 2008: CLARIN: Common Language Resources and Technology Infrastructure. *LREC*. Marrakech. 1244–1248.
- VERDONIK, Darinka, ZWITTER VITEZ, Ana, 2011: *Slovenski govorni korpus Gos*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- VINTAR, Špela, 2010: Luščenje terminologije iz angleško-slovenskih vzporednih in primerljivih korpusov. Špela Vintar (ur.): *Slovenske korpusne raziskave*. Ljubljana: Znanstvena založba Filozofske fakultete. 37–53.
- ŽAUCER, Rok, MARUŠIČ, Lanko, 2009: Jezikovno svetovanje, praksa in ideali. Marko Stabej (ur.): *Infrastruktura slovenščine in slovenistike. Obdobja 28*. Ljubljana: Znanstvena založba Filozofske fakultete. 449–456.