

Zakaj ne z eno poizvedbo hkrati po različnih korpusih? (Troje korpusnih preverb pod primerjalnim drobnogledom)

Helena Dobrovoljc,* Urška Vranjek Ošlak**

*Inštitut za slovenski jezik Frana Ramovša ZRC SAZU
Novi trg 4, 1000 Ljubljana
Fakulteta za humanistiko Univerze v Novi Gorici
Vipavska 13, 5000 Nova Gorica
helena.dobrovoljc@zrc-sazu.si

**Inštitut za slovenski jezik Frana Ramovša ZRC SAZU
Novi trg 4, 1000 Ljubljana
urska.vranjek@zrc-sazu.si

Povzetek

Prispevek predstavlja utemeljitev potrebe po skupnem vmesniku za trenutno dostopne korpusne slovenskega jezika. Potreba je orisana s pomočjo preverb izbranih jezikovnih vprašanj: (1) uresničevanja preglasa pri pregibanju samostalnikov moškega spola, ki se končujejo na izglasni govorjeni c, (2) aktualnost normativno-stilističnih napotkov pri rabi predloga *prek* in vezljivosti glagola *upravljati*; (3) socialno- in funkcijskozvrstno prevrednotenje samostalnika *tenisač* in pridevniškega sopomenskega para *duševen – mentalen*. Zaradi različnosti besedil, na katerih temeljijo posamezni korpusi sodobne slovenščine, se tudi rezultati poizvedb pričakovano razlikujejo. Pri pravopisnih in leksikografskih poizvedbah, ki temeljijo na ovrednotenju obstajajočih različic jezikovne rabe, bi integracija korpusov s skupnim iskalnim vmesnikom olajšala poizvedbe in poleg tega ponujala vsebinske in statistične primerjalne analize.

One query to search different corpora? (Comparative Scrutiny of Three Corpora Checks)

This paper presents an argumentation of the need for a common interface of the currently available Slovenian corpora. This need is outlined through corpus queries on selected language issues: (1) realization of inflectional progressive vowel fronting (*o* to *e*) in masculine nouns ending in pronounced final *c*, (2) relevance of normative and stylistic guidelines in the use of the preposition *prek* and in the valency of the verb *upravljati*; (3) re-valuation of register for the noun *tenisač* and the adjective synonyms *mental* – *dušev*. Due to the diversity of texts on which individual corpora of modern Slovenian are based, the results of these queries differ as expected. For orthographic and lexicographic queries based on the evaluation of existing versions of language use, the integration of corpora with a common search interface would facilitate queries and in addition provide for substantive and statistical comparative analysis.

1. Uvod

Sodobne zadrege jezikovnih uporabnikov jezikoslovje usmerja z napotili, ki se vsebinsko opirajo na jezikoslovna prepričanja zadnjih desetletij prejšnjega stoletja. Pri posodobitvi jezikovnih priručnikov knjižnega jezika in ugotavljanju normativne vrednosti jezikovnih pojavov sodelavci programske skupine »Slovenski jezik v sinhronem in diahronem pogledu« proučujemo aktualne jezikovnosistemske uresničitve v različnih položajih ter izkoriščamo jezikovnotehnoške možnosti pri preverjanju jezikovne rabe po korpusih sodobne slovenščine.

Empirični pristopi pri ugotavljanju rabe jezikovnih prvin so v elektronski dobi dobili nove perspektive, ki jih tradicionalne metode niso omogočale. Korpusno jezikoslovje ne ponuja le obsežnih zbirk raznovrstnih besedil, ki jih je mogoče na različne načine povezovati, temveč prinaša možnosti posploševanja, predvidevanja in napovedovanja jezikovnega vedenja in vzorcev. Od vsega najbolj pomembna je opustitev pristranskosti in

subjektivnosti, ki izvira iz favoriziranja jezikovnega čuta posameznika.

Zaradi zvrstne in obsegovne različnosti ter specializiranosti aktualnih slovenskih korpusov se tudi rezultati poizvedb pričakovano razlikujejo, kar je – kot bo razvidno iz nadaljevanja prispevka – relevantno pri vseh tipih poizvedb, ki temeljijo na presojah pogostnosti, predvsem pa pri poizvedbah, ki vplivajo na leksikografsko socialno- in funkcijskozvrstno vrednotenje oz. kategoriziranje izbranih jezikovnih prvin.

Korpusne poizvedbe iste jezikovne prvine v različnih korpusih naj bi odsevale sorazmerno različnost v odvisnosti od različnih gradivskih zaledij. V korpusnem jezikoslovju obstaja precej primerjalnih korpusnih raziskav, raziskovalci so opozarjali predvsem na probleme pri vzporejanju korpusov z različnim obsegom (Górski, 2007: 119–124; Sharoff, 2010) ali na obsegovne razlike med poizvedbami po različnih korpusih (Möhrs idr., 2017). Precej več študij je bilo usmerjenih v utemeljenost jezikovnotehnoških analiz oz. poudarjanje prednosti računalniško podprte interpretacije jezikovnih vzorcev v različnih ali velikih korpusih, ki omogočajo primerjavo

med uresničitvami v zvrstno različnih besedilih, hkrati pa nakazujejo tudi na novosti (Biber, 1998).

Pričujoča raziskava se približuje tematski primerjavi korpusov Gigafida in slWaC (Logar Berginc in Ljubešič, 2013), ki razkriva vrzeli pri primerjavi lem v korpusu Gigafida kot korpusu tiskanih oz. sekundarno digitaliziranih besedil glede na korpus spletnih besedil slWaC. Te seveda nakazujejo tudi potencialne nadaljnje razlike pri slovničnih, pravopisnih in leksikografskih poizvedbah.

2. Namen članka

V prispevku želimo prikazati postopek preverbe treh relevantnih jezikoslovnih pojavov, za katere sodobna jezikovna praksa izkazuje nekoliko drugačno vedenje od tradicionalnega, in sicer v najbolj aktualnih korpusih slovenskega jezika. S tem bo za razliko od dosedanjih študij prikazano, kako izbira korpusa vpliva na poizvedbo in koliko poizvedb je potrebnih za relevantnost podatka o aktualni rabi. Pri tem bo upoštevano dejstvo, da je referenčni korpus Gigafida, v katerem so zadnja besedila iz leta 2011, za raziskave novejšje leksike (npr. *brexiteh*, *fintech*, *slimacid* ...) in premikov v jezikovni praksi zadnjega desetletja vse bolj neuporaben.

V poglavju »Predstavitev korpusov« so orisane temeljne značilnosti korpusov in njihove razlikovalne lastnosti, zaradi katerih je mogoče pričakovati razlike pri rezultatih. V poglavju »Predstavitev preverb« so predstavljena tri področja leksikološko-normativne obravnave, za katera predvidevamo, da bi pregled njihove rabe v naštetih korpusih pokazal na bistvena razhajanja, ki vplivajo na vrednotenje in kategorizacijo jezikovnih prvin. Gre za področja socialne stratifikacije, normativnosti in empiričnih pristopov pri ugotavljanju specifik izrazne ravnine jezika, ki so se v zadnjem času korenito spremenila.

S primerjavo poizvedb (poglavje »Raziskava«) po izbranih gradivskih korpusih želimo pokazati na potrebo po integraciji orodij ter na interpretativne možnosti, ki jih omogoča raznolikost korpusov.

3. Predstavitev korpusov

Pri izbranih preverbah, ki jih predstavljamo v nadaljevanju, bomo izvedli poizvedbe po naslednjih korpusih.¹

3.1. Referenčni korpus Gigafida

Referenčni korpus Gigafida (v1.1 DeDup – referenčni, dedupliciran) z več kot 918 milijoni pojavnic vključuje besedila iz tiskanih medijev ter le 15 odstotkov spletnih besedil. Besedila segajo časovno v obdobje 1991–2011, kot navajata Logar Berginc in Ljubešič (2013), pa je večina pojavnic iz obdobja po letu 2000.

¹ Do vseh korpusov smo dostopali s pomočjo orodja NoSketch Engine, <https://www.clarin.si/noske/>, od koder so povzeti tudi podatki o številu pojavnic.

3.2. Korpus slWaC

Korpus slWaC (v.2.1) vsebuje 749 milijonov pojavnic s spletnih strani, ki nosijo domeno .si (Ljubešič, Erjavec, 2014). Pregledana je bila različica iz leta 2014, ki časovno pokriva obdobje do junija 2014 (Erjavec, Ljubešič, Logar, 2015).

3.3. Korpus KAS

Korpus KAS (v0.2) je korpus z več kot 771 pojavnicami iz zaključnih akademskih del (diplomske, magistrske, doktorske naloge) iz repozitorijev slovenskih univerz iz obdobja 2000–2015. Ker večina slovenskih akademskih ustanov od kandidatov, diplomantov, magistrandov in doktorandov zahteva lektorirana besedila, so to jezikovno korigirana in zgoščena strokovna oz. znanstvena besedila (Erjavec idr. 2016).

3.4. Korpus Janes

Korpus Janes (v.1.0) vsebuje približno 189 milijonov pojavnic oz. 13.000.000 besedil slovenske računalniško posredovane komunikacije (*computer mediated communication* – CMC). Besedila, vključena v korpus, so bila objavljena v obdobju 2001–2017, razlikujejo pa se glede na to, katero obliko CMC zajamemo (najstarejši so forumski viri, najbolj sveži pa so tviti) (Fišer, Erjavec, Ljubešič, 2016).

Prikazati želimo, da bi pri pravopisnih in leksikografskih poizvedbah, ki temeljijo na ovrednotenju obstajajočih različic jezikovne rabe, integracija korpusov s skupnim iskalnim vmesnikom in možnost statistične primerjave rezultatov olajšala poizvedbe in poleg tega ponujala možnost vsebinske analize.

4. Predstavitev preverb

Razlike pri poizvedbah po različnih korpusih bodo prikazane ob:

- (1) prikazu izjem pri **preglaševanju** končniškega in priponskega *o v e* pri samostalnikih moškega spola z izglasnim [c] (npr. *Leibniz* – *Leibnizov*, *Leibnizev*, *Leibničev*);
- (2) odpravljanju zastarelih **stilističnih** zapovedi (*prek/preko*, *upravljati z/s čim*);
- (3) utemeljevanju sprememb pri **funkcijsko-socialnozvrstni kategorizaciji** besedja v pomenskorazlagalnih slovarjih (*tenisač*, *mentalen* – *duševen*).

Vse nakazane zadrege so izbrane kot aktualna vprašanja slovenske normativistike, saj se mdr. izkazujejo v vprašanih jezikovnih uporabnikih, zastavljenih v Jezikovni svetovalnici Inštituta za slovenski jezik Frana Ramovša pri ZRC SAZU, ki jo uporabljamo kot vir za ugotavljanje perečih oz. aktualnih uporabniških zadreg.

4.1. Končniško in obrazilno preglraševanje

V slovenščini naj bi samoglasniško preglraševanje oziroma prilikovanje samoglasnikov za funkcijsko mehkiimi soglasniki dosledno uresničevali pri vseh samostalnikihi moškega spola z izglasnim [c, j, č, ž, š, dž], pa tudi pri tistih, pri katerih navedenemu izglasju sledi *o* (*Srečo, Mišo, Braco*). Enako velja za samostalnike srednjega spola (*sonce* – tudi v imenovalniku), za pridevniško sklanjatev srednjega spola in za nedoločniško glagolsko pripono *-ova-* nasproti *-eva-* (*kupovati – bojevati, kupčevati*) (Toporišič, ⁴2004: 265–266).

V *Slovenskem pravopisu* 2001 (dalje SP 2001) je navedeno slovnično določilo uveljavljeno kot normativna rešitev z eno samo izjemo, in sicer z določilom: »Za *ts* pišemo samo *-ov*: *Smuts – Smutsov*.« (SP 2001: § 957) Že v slovarskem delu istega pravopisa opazimo razhajanje s pravili ob zgledu *Keats* s pisno dvojnico *Keatsev* in *Keatsov* in z enotno fonetično uresničitvijo [kíčev-].² Drugi pravopisni odmik od realne prakse (in pravil) pa je povezan s preoblikovanjem izglasja imen, ki tvorijo pridevniško obliko, npr. *Clausewitz – Clausewitzev/Clausewičev*, pri čemer dvojnične oblike *Clausewičev* (§ 957) v rabi ne zasledimo.

Opisana glasovna posebnost se zlasti pri nekaterih tujih imenih moškega spola (*Franz, Leibniz*) in tudi domačih vzdevkih (*Rac*) ter imenih z izglasnim *o* (*Joco, Braco*) ne uresničuje.³

Pri preverbi rabe se bomo osredinili na tuja imena z veččrkovnim izglasjem, pri katerih naj bi se (v normativnih priročnikih) preglas uresničeval, a raba izkazuje izjeme.

Sodobno usmerjanje jezikovne rabe, ki se odraža v novem pojmovanju normativnosti, ne sledi zgolj sistemskim možnostim, temveč upošteva tudi načela **izročila** (tradicije) in **jezikovne rabe**:

(1) Glede na **izročilo** v normativnih priročnikih do SP 2001 preglraševanje ni bilo nikoli izpeljano brezizjemno. Tudi določilo v *Slovenski slovnici* (⁴2004: 196) Jožeta Toporišiča je glede preglasa zadržano: »če se končuje na *c*, se ta navadno premenjuje s *č* in tudi dobiva *-ev*, če pa se ne premeni (navadno pri prevzetih besedah), pa *-ev* ni obvezen, zato *stric – stričev* proti *Horac – Horačev* in *Horacev/Horacov*«. Kljub temu je pravopisni slovar (SP 2001) dvojnice iz slovnice ukinil: *Horác – Horačev*.

(2) Opazovanje **rabe** pravi, da na neupoštevanje kodifikacije zagotovo vpliva opazna razlika med knjižnim in govorjenim jezikom, pri katerem se premena zadnjejezičnega *o* v sprednjejezični sredinski samoglasnik *e* v slovenskih regijah od govora do govora razlikuje (Smole 1997).

² Pravilo bi zahtevalo nekaj dopolnil, npr. opozorilo glede neenotnega uresničevanja pisnih izglasnih sklopov *-ts* in *-ds* kot [c] oz. kot [ts], saj SP 2001 take rešitve uveljavlja, npr. pri *Massachusetts* [mesečujsets] in *Yeats* -a [jêjts] nasproti *shórts* [šorc].

³ Kot je razvidno iz raznovrstnih vprašanj v Jezikovni svetovalnici, imajo uporabniki ob njej občutek negotovosti, npr.: Kako je prav: »Francova« ali »Frančeva« sestra? (Dobrovoljc, 2015); Pregibanje imen dveh avstrijskih politikov: »Kurz« in »Van der Bellen« (Dobrovoljc, Lengar Verovnik, 2018).

4.2. Normativna in stilistična napotila na (obliko)skladenjskem področju

Z vse pogostejšo rabo skladenjsko prevzetih in stilistično zaznamovanih zgradb – tudi v novih skladenjskih položajih – prihaja do pomenskih in kategorialnih sprememb ter premikov, ki spreminjajo normativna načela, če slednja pojmujejo kot ovrednotenje jezikovnih pojavov glede na rabo v knjižnem jeziku.

4.2.1. upravljati kaj oz. upravljati s čim

Med dolgo »vzdrževanimi« stilističnimi napotki je zavračanje vezave glagola *upravljati* s predmetom v orodniku (*upravljati s čim*) in iskanje primernejših vzporednic (npr. *upravljati krmilo → usmerjati*). Vezava glagola *upravljati* z orodnikom (*upravljati s čim*) pa se vse od SP 1962 uvršča med skladenjske zgradbe, ki so prepovedane oz. od SP 2001 odsvetovane za rabo v knjižnem jeziku.⁴ Nekoliko blažje omejitve prinaša *Slovar slovenskega knjižnega jezika* (dalje SSKJ), ki vezavo s tožilnikom »rehabilitira«, saj namesto »nepravilne« zveze *upravljati s skladom* usmerja k vezavi s tožilnikom (*upravljati sklad*). V drugi izdaji SSKJ (dalje SSKJ2 2014) sestavljavci ne govorijo več o nepravilnosti vezave z orodnikom, le usmerjajo k prednostni vezavi s tožilnikom.

4.2.2. prek/preko/po

Tradicionalna stilistika, ki je upoštevana v SSKJ, je rabo predloga *prek* (*preko*) v pomenu 'izražanje sredstva, posrednika' zavračala, uporabnike pa so navadno napotili k predlogu *po*: *javnost je obvestil prek radia – javnost je obvestil po radiu*. V SSKJ se neposrednemu normativnemu vrednotenju sicer izognejo z uporabo oznake »neustaljeno« (neustalj.); ta zaznamuje jezikovno prvino, ki se »kljub dosedanjim prepovedim dosti uporablja« (SSKJ, Uvod § 157).

Ne navadnost te slovarske oznake izvira iz normativnega ovinka: avtorji slovarja so ugotovili, da se jezikovna prvina razmeroma pogosto uporablja, pri tem pa ni pojasnjen (niti v uvodnih poglavjih niti v spremljajoči literaturi) značaj te prepovedi, torej kdo in kje⁵ obravnavano prvino prepoveduje. Kot nadrejena sopenka predlogu *prek* je predlagan predlog *po*:

prek [...] neustalj. za izražanje sredstva, posrednika;
po: pismo so poslali prek kurirja; prek radia vplivati na javno mnenje (SSKJ)

SP 2001 to omejitev stopnjuje in rabo predloga *prek* v tem pomenu označuje kot prepovedano, svetuje pa nadomestilo s predlogom *po*:

⁴ V SP 1962 je zgradba *upravljati z/s čim* označena s praznim krožcem, kar pomeni, da gre za »večjo besedno in slogovno spako in najhujši nebodigatreba«, *upravljati kaj* pa je označena kot »nepotrebna ali nelepa ali v nasprotju z duhom slovenskega jezika«.

⁵ Vprašanje, ki se zastavlja, je povezano predvsem s provenienco prepovedi: ali gre za pravopis kot uradni kodifikacijski priročnik ali za individualne presoje jezikoslovcev (Gradišnik, pisci brusov, tj. normativnih priročnikov s prepovedovalno črno-belo logiko, npr. Sršen, ipd.)?

prek [...] poslati pismo •~ kurirja *po kurirju* (SP 2001)

V SSKJ2 (2014) je normativna oznaka »neustaljeno« odstranjena, slovarski prikaz še vedno usmerja na predlog *po*, zgledi so nespremenjeni:

prek [...] *za izražanje sredstva, posrednika; po*²:
pismo so poslali prek kurirja; prek radia vplivati na javno mnenje (SSKJ2)

Vendar pa pregled rabe priča, da narašča pogostnost predloga *prek* v zgradbah, ki opisujejo dejavnosti, povezane s spletnimi oz. internetnimi storitvami, npr. *prodaja prek spleta*. V teh kontekstih se pojavljajo zgledi kot *pogovarjati se prek Skypa; sporočiti prek interneta*, ki še niso obremenjeni s normativnostjo, uporabniki pa jih pojmujejo kot nevtralne (Dobrovoljc, 2014), zato je take primere smiselno preučiti na novo.

4.3. Socialnozvrstna in funkcijskozvrstna kategorizacija

Leksika v slovenskem jeziku se nenehno in tudi zelo opazno spreminja z vidika pogostnosti, medtem ko njeno stilno prevrednotenje opazujemo redkeje, morda še najbolj očitno v primerih, ko v jeziku soobstajata pomensko prekrivni besedi in se sčasoma začneta specializirati za omejeno zvrstno in stilno uporabo.⁶

4.3.1. *tenisač* (šport. žarg.)

Tako je npr. beseda *tenisač* v prvi in drugi izdaji SSKJ veljala za *žargonsko*, danes pa jo obravnavamo kot nevtralnno. Ž argonizmi, ki v SSKJ2 (Uvod § 142) označujejo »strokovno pogovorno leksiko, ki **ni znana širšemu krogu**, temveč predvsem poznavalcem področja« (poud. a.), se danes že zelo razširjeno uporabljajo v knjižnem jeziku, celo v strokovnih besedilih.⁷ Malenkostno omiljena je omejitev žargonske rabe v pravopisnem slovarju (SP 2001: § 1060), saj je oznaka »žargonsko« opredeljena kot »jezikovna prvina, ki se v strokovnem izrazju uporablja **namesto navadnega strokovnega poimenovanja**« (poud. a.). Zdi se, da je ta omejitev za besede, kot sta npr. *tenisač* in *radijec*, preozka.

4.3.2. *duševen* – *mentalen*

Spreminja se tudi funkcijskozvrstna pripadnost leksike, ki jo v prispevku ponazarjamo pri sopomenskem paru *mentalen* – *duševen*. Medtem ko je prvi pridevnik (*duševen*) prisoten že v protestantskem besedišču (*Besedje*, 2014), drugega (*mentalen*) poznajo slovenski slovarji šele od prvega povojnega pravopisa (SP 1950) naprej. Prvi

slovarji pridevnika *mentalen* normativno ne odsvetujejo, so pa bolj naklonjeni tedaj bolj razširjeni sopomenski različici *miseln*, ki kasneje iz slovarskih napotil ob pridevniku *mentalen* izgine, ohranja pa se pri pojmovni samostalniški izpeljanki *mentalnost* – *miselnost*, *mentaliteta* (SSKJ2, 2014). Tudi najnovejši slovar sopomenk, *Sinonimni slovar slovenskega jezika* (Snoj idr., 2016), pridevnika *miseln* ne navaja med sopomenkami pridevnikov *duševen* in *mentalen*.

Izhodišče za korpusno primerjavo je torej opredelitev v SSKJ2 (2014). Pridevnik *mentalen* (*mentalni*) v tem slovarju opredeljujejo kot izraz, rabljen »zlasti v leposlovnem ali znanstvenem jeziku«, in zanj predlagajo nevtralnno sopomenko *duševen* (*duševni*). Vendar pa pogostnost uporabe pridevnika *mentalen* nakazuje, da ga v nekaterih kontekstih ni mogoče nadomestiti s slovarsko nadrejenim pridevnikom.

5. Raziskava in interpretacija

5.1. Oblikoslovno-besedotvorno uresničevanje preglasa pri izbranih tujih imenih⁸

5.1.1. *Leibniz* – *Franz* – *Fritz*

Lastno ime *Leibniz* po veljavni pravopisni normi tvori svojilni pridevnik na *-ov/-ev* s pisnima oblikama *Leibničev* in *Leibnizev*, ki se govorno uresničujeta kot [lájbničev-], v rabi pa zasledimo tudi nestandardno možnost *Leibnizov*, torej nepreglašeno obliko, za katero predvidevamo, da se bo pojavljala le v korpusih z nelektoriranimi besedili. Primerjalno poizvedbo smo izvedli tudi za imeni *Franz* in *Fritz*, pri katerih smo odkrivali le razmerje med pridevniki s pisno nespremenjenim izglasjem, a z uresničenim končniškim preglasom. Izbira imen *Franz* in *Fritz* se namreč ni izkazala za najbolj primerno, saj podomačenih oblik *Frančev* in *Fričev* v korpusih ni bilo mogoče ločiti od enakopisnih pridevnikov na *-ov/-ev* iz lastnih imen *Franc* in *Fric*. Za izvedbo primerjave smo se odločili, ker smo presodili, da je relevantno tudi razmerje med obema obraziloma, npr. *Franzov* – *Franzev*.⁹

Besedotvorje

Pravopisna prednostna oblika svojilnega pridevnika *Leibničev* in neprednostna dvojica *Leibnizev* se v rabi potrjujeta samo v korpusu Gigafida v nizkem odstotku. Nestandardna tvorba *Leibnizov*, ki po rabi izstopa kot najpogostejša (77 %), se pojavlja v treh korpusih. (Izjema je korpus Janes.)

⁶ O podobnih vprašanjih je ob parih *stegno* – *bedro*, *čik* – *ogorek* ipd. razpravljal že Tomo Korošec (*Pet minut za boljši jezik*, 1972), ko tovrstne raziskave še niso bile empirično podkrepjene.

⁷ Npr.: *Jeseni 1941 so nekateri slovenski športniki iz Ljubljane (atleti, plavalci, kajakaši, tenisača in pozimi umetnostna drsalka) nastopili na različnih tekmovanjih v Italiji, peščica pa jih je nadaljevala športno pot v italijanskih klubih. (Zanimanje za sport je prodrlo med Slovenci že v široke sloje*, 2005).

⁸ Za iskanje smo uporabili pogoje npr. [lemma_lc="Franzev" & tag="P.*"] in [lemma="Franz" & lc="Franzem" & tag="S.e.*"],

s katerimi smo želeli zmanjšati šum in pri oblikoslovju zajeti le orodniško obliko ednine. Najdene rezultate smo vedno primerjali tudi s preprostim iskanjem. Razlog za tovrstno poizvedbo je v tem, da imena niso vedno zapisana z veliko začetnico (napaka v lematizaciji morda?).

⁹ Naslov zgodbe o dečku Francu avstrijske pisateljice Christine Nöstlinger se v tiskanih različicah glasi *Francove zgodbe*, medtem ko jih v elektronskih virih najdemo lektorirane, npr. v Wikipediji (https://sl.wikipedia.org/wiki/Christine_Nöstlinger), v *Franceve zgodbe*, ne najdemo pa oblike *Frančeve zgodbe*.

Lema	Pridevnik	GF	KAS	slWaC	Janes	Σ
Leibniz	<i>Leibnizev</i>	7	6	32	0	45
	<i>Leibnizov</i>	217	35	42	4	298
Franz	<i>Leibničev</i>	7	0	0	0	7
	<i>Franzev</i>	33	9	8	0	50
Fritz	<i>Franzov</i>	13	23	3	1	50
	<i>Fritzev</i>	29	3	45	1	78
	<i>Fritzov</i>	35	9	40	2	86

Tabela 1: Pogostnostna primerjava besedotvorne variantnosti imen *Leibniz*, *Franz* in *Fritz* v štirih izbranih korpusih

Pri imenih *Fritz* in *Franz* bi najbolj zvesto približanje normi knjižnega jezika pričakovali v besedilih korpusa akademskih besedil, ki naj bi bila tudi v največjem odstotku lektorirana, a je v tem korpusu največje odstopanje od edine oblike, ki jo pravopis dopušča: *Franzev* in *Fritzev*. Najbolj se normativnim pričakovanjem približata Gigafida (*Franzev*, toda *Fritzov*!) in korpus spletnih besedil slWaC, v katerem se raba nagiba k preglašnim oblikam. (Pojavitev v korpusu Janes je bilo zelo malo ali nič.)

Oblikoslovje

Oblika za orodnik ednine s preglasom (***Leibnizem***) se potrjuje v korpusih Gigafida, KAS in slWaC, a ima zelo malo pojavitev, nestandardna oblika brez preglasa (*Leibnizom*) pa je pogostejša in se pojavi v korpusih KAS, slWaC in Janes.

Pri orodniku ednine sta obliki s preglasom (*Franzem*, *Fritzem*) pogostejši, a tudi nepreglašeni obliki (*Franzom*, *Fritzom*) sta zastopani: v korpusu slWaC imata skoraj pol toliko pojavitev kot preglašeni obliki. V korpusu Janes ni nobene od njiju.

Lema	Pridevnik	GF	KAS	slWaC	Janes	Σ
Leibniz	<i>Leibnizem</i>	185	0	8	0	193
	<i>Leibnizom</i>	217	2	10	1	230
Franz	<i>Franzem</i>	216	28	77	0	221
	<i>Franzom</i>	72	13	31	2	116
Fritz	<i>Fritzem</i>	90	9	56	0	155
	<i>Fritzom</i>	34	5	36	0	75

Tabela 2: Pogostnostna primerjava oblikoslovne variantnosti imen *Leibniz*, *Franz* in *Fritz* v štirih izbranih korpusih

5.1.2. *Keats* – *Yeats*

Imeni *Keats* in *Yeats* smo izbrali zaradi očitne zadrege pravopiscev, ki so v SP 2001 pri imenu *Yeats* predvideli obliko pridevnika *Yeatsov*, za *Keats* pa obliki *Keatsev* in *Keatsov*; obe se govorno uresničujeta kot [kičev-]. Podobno

¹⁰ Gre za zgled v jezikoslovnem delu, natančneje v končnem poročilu projekta Sporazumevanje v slovenskem jeziku, Kazalnik 17 – Slogovni priročnik, kjer so nanizane empirično pridobljene normativne zadrege uporabnikov:

je tudi pri orodniških oblikah, saj imenu *Yeats* pripišejo orodnik *Yeatsom*, imenu *Keats* pa pisno dvojnico: *s Keatsem/Keatsom* z enotno govorno uresničitvijo [s kičem]. Zanimalo nas je, kako se neenotna kodifikacija odraža v rabi oz. v besedilih korpusov 17 let po izidu pravopisa.

Besedotvorje

Lema	Pridevnik	GF	KAS	slWaC	Janes	Σ
Yeats	<i>Yeatsov</i>	0	0	0	0	0
	<i>Yeatsov</i>	15	1	13	0	39
Keats	<i>Keatsev</i>	0	0	0	0	0
	<i>Keatsov</i>	16	1	6	0	23

Tabela 3: Pogostnostna primerjava besedotvorne variantnosti imen *Yeats* in *Keats* v štirih izbranih korpusih

Korpusi Gigafida, KAS in slWaC (v korpusu Janes ni zadetkov) za svojilni pridevnik na *-ov/-ev* iz imen *Yeats* in *Keats* izkazujejo prevlado nepreglašene oblike (*Yeatsov*, *Keatsov*), ki pa je v nasprotju z imeni *Leibniz*, *Franz* ... tudi pravopisno dopuščena, pri *Yeats* celo edina.

Oblikoslovje

Tudi pri oblikoslovnem pregibanju v orodniku prevladujejo nepreglašene oblike (*Yeatsom*, *Keatsom*). Edini primer preglašene oblike (*Yeatsem*) najdemo v korpusu KAS.¹⁰

Lema	Orodnik	GF	KAS	slWaC	Janes	Σ
Yeats	<i>Yeatsem</i>	0	1	0	0	1
	<i>Yeatsom</i>	4	1	2	0	6
Keats	<i>Keatsem</i>	0	0	0	0	0
	<i>Keatsom</i>	3	1	4	0	8

Tabela 4: Pogostnostna primerjava oblikoslovne variantnosti imen *Yeats* in *Keats* v štirih izbranih korpusih

Imeni *Yeats* in *Keats* se torej približujeta pravopisnemu določilu v členu 957, ki določa, da za izglasnim sklopom »ts« pišemo samo *-ov*: *Smuts – Smutsov*«, in oddaljujeta od napotil pravopisnega slovarja za ime *Keats*. Vendar bi za nedvoumno napotilo v kodifikacijskem priročniku potrebovali predvsem podatek o tem, kako zapisa *Keatsov* in *Keatsev* izgovorno uresničevati

5.2. Normativna napotila na (obliko)skladenjskem področju

5.2.1. *upravljati kaj* oz. *upravljati s čim*

V izbranih korpusih smo preverili pojavljanje zvez *upravljati kaj* in *upravljati s čim*.¹¹ Ugotovili smo, da sta v

http://projekt.slovenscina.eu/Media/Kazalniki/Kazalnik17/Kazalnik_17_Slogovni_prirocnik_SSJ.pdf

¹¹ Iskalna pogoja: [lemma="upravljati"] [tag="S...t"] in [lemma="upravljati"] [lemma="z"] [tag="S...o"].

Gigafidi zvezi približno enako pogosti, medtem ko je v drugih treh korpusih pogostejša zveza z orodnikom, ki je bila v normativno-slogovnih priročnikih druge polovice 20. stoletja odsvetovana.

	GF	KAS	slWaC	Janes
<i>upravljati kaj</i>	2597	4258	2230	276
<i>upravljati s čim</i>	2668	6084	3248	497

Tabela 5: Prikaz različne vezljivosti glagola *upravljati* v štirih izbranih korpusih

Teško bi trdili, da je glagol *upravljati* v orodniški vezavi drugače kolokabilen kot v tožilniški, saj se v vseh korpusih najpogosteje pojavljajo tudi sicer prekrivne besedne zveze:

- s tožilnikom: *upravljati družbo*, *upravljati državo*, *upravljati premoženje*;
- z orodnikom: *upravljati z denarjem*, *upravljati z državo*, *upravljati s premoženjem*.

Navedeno je potrdil tudi primerjalni pregled najpogostejših lem, ki se pojavljajo v tožilniku in orodniku v več kot dveh korpusih.¹²

<i>upravljati</i>			
Tožilnik	Korpusi	Orodnik	Korpusi
<i>družbo</i>	GF, KAS, SW, J	<i>s premoženjem</i>	GF, KAS, SW, J
<i>državo</i>	GF, KAS, SW, J	<i>z denarjem</i>	GF, KAS, SW, J
<i>premoženje</i>	GF, KAS, SW, J	<i>z državo</i>	GF, KAS, SW, J
<i>delovanje</i>	GF, KAS, SW	<i>s časom</i>	KAS, SW, J
<i>podjetje</i>	GF, KAS, SW	<i>s sistemom</i>	GF, KAS, SW
<i>sredstva</i>	GF, KAS, SW	<i>s sredstvi</i>	GF, KAS, SW
<i>vozilo</i>	GF, KAS, SW	<i>z nepremičninami</i>	GF, KAS, SW

Tabela 6: Primerjava najpogostejših samostalniških lem ob glagolu *upravljati*, ki se pojavljajo v več kot dveh korpusih.

Čeprav bi pri orodniški vezavi pričakovali več samostalnikov, ki lahko nastopajo v vlogi sredstva (*upravljati z denarjem*, *s sredstvi*), pri tožilniški vezavi pa samostalnice, ki imajo abstrakten ali pojmovni pomen neke skupnosti (*družba*, *država*, *podjetje*), se je po razvrstitvi lem izkazalo, da je razdelitev preveč razpršena, da bi jo lahko posplošili. V obeh sklonih se namreč pojavljajo samostalniki iz obeh zgoraj navedenih skupin: *dokument*, *država*, *elektrarna*, *informacija*, *infrastruktura*, *podatek*, *podjetje*, *premoženje*, *sklad*, *smučišče*, *sprememba*, *sredstva*, *tveganja*, *viri*, *vozilo*, *znanje*.

¹² Upoštevali smo 40 najpogostejših kolokacij.

¹³ Iskali smo po lemah *prek*, *preko* in *po*.

Raziskava vezave glagola *upravljati* s tožilnikom in orodnikom ter primerjava vseh štirih korpusov sta bili zamudni, saj iskalni mehanizem ne omogoča primerjave rezultatov različnih korpusov, zato smo primerjave lem izvedli ročno. Z vidika presoje ustreznosti slovarskih napotil tudi gradivska raziskava pritrjuje mnenju, izraženemu v Jezikovni svetovalnici (Vranjek Ošlak, 2017), saj je orodniška vezava kljub večdesetletnemu prepovedovanju pogostejša in med uporabniki bolj razširjena, ne glede na pomenski oz. kolokacijski okvir.

5.2.2. *prek/preko/po*

Pregled izbranih korpusov je pokazal,¹³ da pri predlogih *prek* in *preko* v prvih 40 samostalniških kolokacijah v vseh štirih korpusih močno prevladuje pomen 'sredstvo, posrednik', npr.: *prek/preko interneta*, *prek/preko elektronske pošte*, *prek aplikacije*, *prek/preko javnih medijev*, *prek/preko mobilnega telefona*, *prek/preko omrežja*, *prek/preko računalnika*, *prek/preko vmesnika*, *prek/preko spletne strani*, *prek/preko študentskega servisa* itd.

V drugih pomenih ('več kot', 'čez') nastopata ta dva predloga v zvezah, kot so npr. *prek/preko meje*, *preko noči*, *preko zime*, *preko mostu*, *preko dneva*, *prek 200 metrov* itd.

Za primerjavo smo preverili še 40 najpogostejših kolokacij predloga *po*, ki naj bi bil primerna oz. normativno priporočljiva zamenjava za predloga *prek* in *preko*. V pomenu 'sredstvo, posrednik' nastopa ta predlog v le treh kolokacijah, in sicer: *po telefonu*, *po pošti* in *po cesti*.

Očitno uporabniki kljub normativnim napotkom v slovarjih (tako SP 2001 kot obeh izdajah SSKJ) uporabljajo predloga *prek* in *preko* v pomenu 'sredstvo, posrednik' in ju le izjemoma zamenjujejo s predlogom *po*. Zlasti se nagibajo k rabi predloga *prek* v zvezah s samostalniki, ki označujejo sodobne načine komunikacije, npr. *internet*, *elektronska pošta*, *aplikacije*, *mobilni telefon*, *omrežje*, *vmesnik*.

5.3. Stilno prevrednotenje

5.3.1. *tenisač* (šport, žarg.)

Pri preverjanju socialnozvrstnih oznak je korpusna poizvedba bistvenega pomena. Samostalnik *tenisač* je kot športni žargonski izraz (tako ga opredeljujejo vsi sodobni normativni slovarji) poenobesedena različica »nevtralne« zveze *teniški igralci*. Vendar pa pregled prakse po korpusih pokaže, da uporabniki tudi samostalnik *tenisač* dojemajo kot nevtralen. Navajamo primer rabe iz korpusov KAS (diplomsko delo) in Janes (primer iz spletne klepetalnice):¹⁴

- *Iz načina preživljanja njegovega prostega časa in izbora vrste športa, lahko precej razberemo o njegovih značajskih karakteristikah. Igralci košarke in nogometa so verjetno bolj timsko naravnani ljudje kot tenisači.* (vir: Ana Jakelič: *Vpliv zunanje podobe na poslovno uspešnost menedžerjev*);

¹⁴ Navajamo le nekaj zgledov izmed mnogih.

- *Izmerjen Dopplerjev efekt ustvarja resen dvom o vedno enaki svetlobni hitrosti na ponoru. Če se vrnem na opisovan primer **tenisačev** z ogledali, mirujoči **tenisač**, kjer koli na poti od izvora do ogledala prestreže foton, opazi vedno enako valovno dolžino.* (vir: *Kvarkadabra*, Vprašanja & odgovori, Vprašanja za Einsteina).

5.3.2. duševen – mentalen¹⁵

Pogostnostna primerjava sopomenk *duševen* in *mental* v različnih korpusih kaže, da je pridevnik *duševen* skoraj trikrat pogostejši od pridevnika *mental*.

	GF	KAS	SIWaC	Janes	Σ
<i>mental</i>	5751	13.211	11.218	2405	32.585
<i>duševen</i>	31.679	61.832	25.723	3037	122.271

Tabela 7: Prikaz pogostnosti pridevniškega para *mental* – *duševen* v štirih izbranih korpusih

Pregled rabe v referenčnem korpusu Gigafida je pokazal, da je frekvenca pojavljanja pridevnikov po besedilnih tipih zelo podobna (čeprav sta pogostnostno različna – v razmerju 85 : 15 %). Oba sta najbolj značilna za besedilni tip **tisk/knjižno/strokovno**, najbolj pa se njuna relativna frekvenca razlikuje pri **internetnih besedilih** (zanje je bolj tipičen pridevnik *mental*) in tipu **tisk/drugo** (pogostejši je pridevnik *duševen*). To je deloma v skladu z opisom pridevnika *mental* v SSKJ, saj je tudi pridevnik *duševen* pogost v strokovnih besedilih, hkrati pa se pridevnik *mental* pojavlja tudi na internetu, kjer prevladuje praktičnosporazumevalna zvrst.¹⁶

Pregled in primerjava kolokacij v vseh štirih korpusih sta pokazala, da oba pridevnika nastopata v naslednjih besednih zvezah:¹⁷

- vsi štirje korpusi: *mentalno/duševno zdravje*, *mentalna/duševna zaostalost*, *mentalno/duševno stanje*;
- samo Gigafida: *mentalna/duševna prizadetost*, *mentalna/duševna motnja*;
- samo Janes: *mentalna/duševna higiena*, *mentalni/duševni bolnik*.

Pridevnik *mental* se v vseh **štirih** korpusih povezuje v besedne zveze *mentalna higiena*, *mentalna sposobnost*, *mentalno stanje*, *mentalna zaostalost* in *mentalno zdravje*.

V **treh** korpusih najdemo besedni zvezi *mentalna kondicija* (Gigafida, KAS in SIWaC) in *mentalna lenoba* (Gigafida, SIWaC in Janes).

V **dveh** korpusih pa so pogoste naslednje besedne zveze: *mentalni zemljevid*, *mentalna izčrpanost* (Gigafida in KAS); *mentalna retardacija*, *mentalna reprezentacija* (KAS in SIWaC); *mentalna stimulacija* (Gigafida in

SIWaC); *mentalni domet* (Gigafida in Janes); *mentalni napor* (SIWaC in Janes).

Zanimive besedne zveze, ki se pojavljajo samo v enem od korpusov, so npr. še:

- *mentalna sirota*, *mentalni horizont*, *mentalni slepec*, *mentalna prizadetost*, *mentalni revček* (Gigafida);
- *mentalni trening*, *mentalna predstava*, *mentalni leksikon* (KAS);
- *mentalna jasnost* (SIWaC);
- *mentalni invalid*, *mentalna pohabljenost*, *mentalna okužba* (Janes).

Pridevnik *duševen* se v vseh **štirih** korpusih povezuje v besedne zveze *duševna bolečina*, *duševna bolezen*, *duševni bolnik*, *duševna celovitost*, *duševna motnja*, *duševno počutje*, *duševna prizadetost*, *duševno ravnovesje*, *duševni razvoj*, *duševno stanje*, *duševna stiska*, *duševna zaostalost* in *duševno zdravje*.

V **treh** korpusih se pojavita besedni zvezi *duševni mir* in *duševna muka* (mn.) (Gigafida, SIWaC in Janes).

V **dveh** korpusih najdemo besedne zveze *duševno trpljenje* (Gigafida in SIWaC); *duševna obremenitev* (Gigafida in Janes); *duševna zmogljivost*, *duševna sprostitve* (KAS in SIWaC).

Zanimive besedne zveze, ki se pojavijo le v **enem** od korpusov, so npr. še: *duševni pretres* (Gigafida); *duševna manjrazvitost*, *duševna integriteta* (KAS); *duševni aparat* (Janes).

Iz analize je mogoče razbrati, da je pridevnik *duševen* pogostejši v besednih zvezah, ki se pojavijo v vseh štirih korpusih. To je lahko znak večje razširjenosti in funkcijskozvrstne nevtralnosti tega pridevnika v nasprotju s pridevnikom *mental*. Vendar pa tudi pridevnik *mental* v sodobni jezikovni praksi ni omejen zgolj na leposlovna in znanstvena besedila, kot so izkazovale leksikalne raziskave v preteklosti, temveč opažamo pojavljanje v različnih zvrsteh, npr. v korpusu spletnih besedil SIWaC in v korpusu jezika družbenih omrežij oz. nestandardnega jezika Janes (razmerje v odstotkih je med *mental* in *duševen* tudi glede na korpus GF precej bolj izenačeno, 56 : 44 %). Prav nestandardni jezik predstavlja za sodobno normativistiko enega od prvih detektorjev napovedujočih se pomenskih premikov.

Naj izpostavimo še očitne pomenske prenose oz. čustveno zaznamovanost pridevnika *mental* v besednih zvezah korpusa nestandardne slovenščine Janes (*mentalni invalid*, *mentalna pohabljenost*, *mentalna okužba*), ki kažejo na primerno ustaljenost pridevnika v nevtralnem pomenu, da lahko razvija pomenske odmike.

6. Sklep

Glavni izziv pri korpusno podprtih raziskavah je za pisce jezikovnih priročnikov, slovnice in slovarjev v izbiri

¹⁵ Iskali smo po lemah *mental* in *duševen*.

¹⁶ Takšna primerjava po besedilnih tipih za druge korpusne ni bila niti mogoča niti smiselna, saj ima le Gigafida enakovredno zastopnost različnih besedilnih tipov.

¹⁷ Upoštevali smo prvih 20 kolokacij, in sicer samo zveze pridevnika v levem prilastku in samostalnika v jedru.

orodja, ki lahko potrdi intuitivne domneve o spremembah jezikovnih zgradb, oblik, tvorb ... ali ki odkrije še neopisane jezikovne vzorce. Opis sodobne slovenščine temelji na slovarskih in slovničnih raziskavah leksike iz 70. in 80. let ter normativnem opisu iz 90. let, poskusi izkoriščanja korpusa kot gradivske osnove pa so se izkazali kot uspešni pri pripravi leksikalne baze za slovenščino, ki je predstavljena v monografiji *Leksikografski opis slovenščine v digitalnem okolju* (Gantar, 2015). Tudi prenova druge izdaje SSKJ (2014) je potekala s pomočjo referenčnega korpusa Gigafida, ki časovno zajema obdobje do leta 2010.

Pri interpretaciji sodobne jezikovne prakse Gigafida ne zadostuje več kot edini vir, še zlasti za novejša besedja ne (npr. *brexit*, *finteh*, *slimacid* ...), novi korpusi akademske, nestandardne in spletne slovenščine pa so dobrodošla dopolnila, ki ponujajo nove rezultate, a so poizvedbe zelo zamudne, saj moramo postopke ponavljati za vsak korpus posebej.

6.1. Povezovanje korpusov

V prispevku smo zato želeli pokazati na prednosti, ki bi jih imel enoten iskalni vmesnik (še zlasti, če bi ta ponudil tudi primerjave deležev med obsegovno različnimi korpusi). Po vseh korpusih je sicer mogoče iskati z orodjem Sketch Engine, ni pa možnosti primerjave rezultatov iskanj (moramo jih izvoziti v Excell ali kako drugo obliko). Primerjava rezultatov iskanja po vsakem korpusu posebej je zamudna, zaradi količine podatkov je tudi več možnosti za napake. Primerjava rezultatov iskanj znotraj skupnega vmesnika bi možnost napake zmanjšala, hitrost iskanja in primerjave pa bi se bistveno povečala. Dobrodošla bi bila tudi možnost iskanja po velikem združenem korpusu (sestavljene iz vseh korpusov, ki so zdaj na voljo), če bi omogočal ohranitev podatkov iz izvornih »podkorpusov«.

6.2. Pomen rezultatov poizvedb

Pri prikazu korpusne preverbe stanja smo se zato odločili za tri poizvedbe, ki so nujne in jezikoslovno pereče, hkrati pa do končnih rezultatov ni mogoče priti brez ročnega dela. Pri treh primerih jezikovne neustaljenosti smo ugotovili naslednje.

Na **oblikoslovno-besedotvornem področju** je mogoče zaznati odstopanja od zapovedanih normativnih vzorcev enakomerno v vseh korpusih (*Franzov : Franzev*), razlike so le v korpusu Janes, v katerem nekaterih lem ni bilo mogoče preveriti. Izrazito sistemsko utemeljene normativne zapovedi, ki očitno niso bile preverjene z rabo (*Keatsev, Yeatsev*), kljub določilom v pravopisu tudi po 20 letih v rabi niso prisotne – ne glede na tip korpusa.

Na **področju (obliko)skladnje** smo raziskovali tožilniško in orodniško vezavo pri glagolu *upravljati* in rabo predloga *prek/preko* za izražanje sredstva. Tu se stanje v Gigafidi temeljito razlikuje od drugih korpusov predvsem pri vezavi, saj je razmerje med dvojnicama (*upravljati kaj : upravljati s čim*) v Gigafidi enakomerno, medtem ko nepriporočljiva orodniška vezava prednjači v vseh drugih treh korpusih. Pri rabi predloga *preko/prek* razlike med korpusi niso izrazite.

Tudi na področju **stilnega prevrednotenja** (presoja žargonskosti pri samostalniku *tenisač* in raba pridevniškega para *mentalen – duševen*) se največje razlike izkazujejo med korpusoma Gigafida in Janes, pri čemer v slednjem zasledimo inovativne pomenske prenose, ki jih drugi korpusi ne izkazujejo, kar kaže na pomensko produktivnost prevzetega pridevnika *mentalen* v nestandardnem jeziku oz. na ustvarjalnost govorcev v nestandardnih zvrsteh.

Preverbe kažejo predvsem na to, da so razlike med rabo (oz. stanjem v korpusih) in priročniki manj razvidne na ravni sistemskega normiranja oz. vrednotenja jezikovnih prvin kot pri stilni in zvrstni kategorizaciji leksike ter stilističnih napotkih, ki koreninijo v jezikoslovju druge polovice 20. stoletja, kjer je bila v slovenščini očitno narejena »tiha« revolucija, s katero se sociolingvistika še ni spoprijela.

7. Literatura

Besedje slovenskega knjižnega jezika 16. stoletja. 2014. www.fran.si. Dostop 3. 4. 2018.

Douglas Biber, Susan Conrad in Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Helena Dobrovoljc. 2014. Raba predloga »prek« v pomenu 'izražanje sredstva'. *Jezikovna svetovalnica*, <https://svetovalnica.zrc-sazu.si/topic/9/raba-predloga-prek-v-pomenu-izrazanje-sredstva>. Dostop 9. 4. 2018.

Helena Dobrovoljc. 2015. Kako je prav: »Francova« ali »Frančeva« sestra?. *Jezikovna svetovalnica*, <https://svetovalnica.zrc-sazu.si/topic/789/kako-je-prav-francova-ali-francjeva-sestra>. Dostop 9. 4. 2018.

Helena Dobrovoljc in Tina Lengar Verovnik. 2018. Pregibanje imen dveh avstrijskih politikov: »Kurz« in »Van der Bellen«. *Jezikovna svetovalnica*, <https://svetovalnica.zrc-sazu.si/topic/2604/pregibanje-imen-dveh-avstrijskih-politikov-kurz-in-van-der-bellen>. Dostop 9. 4. 2018.

Tomaž Erjavec, Darja Fišer, Nikola Ljubešić, Nataša Logar in Milan Ojsteršek. 2016. Slovenska znanstvena besedila: prototipni korpus in načrt analiz. V: Erjavec, Tomaž (ur.), Fišer, Darja (ur.). *Zbornik konference Jezikovne tehnologije in digitalna humanistika*. Znanstvena založba Filozofske fakultete. 58–64.

Tomaž Erjavec, Nikola Ljubešić in Nataša Logar Berginc. 2015. The slWaC corpus of the Slovene Web. *Informatica: an international journal of computing and informatics*, št. 1, 35–42.

Darja Fišer, Tomaž Erjavec in Nikola Ljubešić. 2016. JANES v0.4: Korpus slovenskih spletnih uporabniških vsebin. *Slovenščina 2.0*, 4 (2): 67–99.

Polona Gantar. 2015. *Leksikografski opis slovenščine v digitalnem okolju*. Ljubljana: Filozofska fakulteta.

Rafał Górski. 2007. Representativeness of a written part of a Polish general-reference corpus. Primary notes. V: *Corpus linguistics, computer tools, and applications - state of the art*: PALC.

Nataša Logar in Nikola Ljubešić. 2013. Gigafida in slWaC: tematska primerjava. *Slovenščina 2.0*. 78–110.

Christine Möhrs, Meike Meliss in Dolores Batinić. 2017. LeGeDe – Towards a Corpus-based Lexical Resource of Spoken German. V: *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference*. Ur. Iztok Kosem. Carole Tiberius. Miloš Jakubiček. Jelena Kallas. Simon Krek. Leiden, the Netherlands, 19–21 September 2017. 281–298.

NoSketch Engine. URL: <https://www.clarin.si/noske/>. Dostop 23. 3. 2018 – 14. 4. 2018.

Serge Sharoff. 2010. Analysing Similarities and Differences between Corpora. V: T. Erjavec, J. Žganec Gros (ur.): *Zbornik Sedme konference Jezikovne tehnologije*. 5–11.

SSKJ = *Slovar slovenskega knjižnega jezika*. Prva knjiga A–H (1970); druga knjiga I–Na (1975); tretja knjiga Ne–Pren (1979); četrta knjiga Preo–Š (1985); peta knjiga T–Ž (1991) z dodatki od A–Š. Ljubljana: SAZU – Državna založba Slovenije.

SSKJ2 = *Slovar slovenskega knjižnega jezika*. 2014. Druga, dopolnjena in deloma prenovljena izdaja. Ljubljana: Založba ZRC, ZRC SAZU.

SP 1950 = *Slovenski pravopis*. 1950. Ljubljana: SAZU – Državna založba Slovenije.

SP 1962 = *Slovenski pravopis*. 1962. Ljubljana: SAZU – Državna založba Slovenije.

SP 2001 = *Slovenski pravopis*. 2001. Ljubljana: Založba ZRC.

Vera Smole. 1997. Sovplivanje samoglasnikov in soglasnikov v vzhodnodolenjskih govorih. *Jezikoslovni zapiski*. 167–174.

Jerica Snoj idr. 2016. *Sinonimni slovar slovenskega jezika*. Ljubljana: Založba ZRC.

Jože Toporišič. 2004. *Slovenska slovnica*. Četrta izdaja. Maribor: Založba Obzorja.

Urška Vranjek Ošlak. 2017. Nekoč nepravilno, danes dovoljeno: upravljati s skladom in upravljati sklad. *Jezikovna svetovalnica*. URL: <https://svetovalnica.zrc-sazu.si/topic/2666/nekoč-nepravilno-danes-dovoljeno-upravljati-s-skladom-in-upravljati-sklad>. Dostop 3. 4. 2018.