

Diana Košir,* Tomaž Erjavec**

Korpusna analiza pripovednega sloga in jezikovne norme v starejši verski periodiki

IZVLEČEK

V prispevku je predstavljen postopek izdelave korpusa CVET, ki vsebuje besedila patra Hijacinta Repiča, objavljena v verski reviji Cvetje z vertov sv. Frančiška v obdobju 1881–1916. Korpus je uporabljen kot podlaga za jezikovno in stilistično analizo, opravljeno z orodjem noSketch Engine. Z analizo frekvenčnosti izbranih spremenljivk sta opisana besedišče patra Repiča in njegov pripovedni slog. Nazadnje je na primeru besed tipa bralec/bravec opazovan sinhrono-diahroni in normativni vidik starejšega slovenskega jezika besedil v korpusu.

Ključne besede: starejši slovenski jezik, verski tisk, korpusno jezikoslovje, korpusna stilistika, jezikovna norma

ABSTRACT

CORPUS ANALYSIS OF NARRATIVE STYLE AND LINGUISTIC NORM IN AN OLDER RELIGIOUS PERIODICAL

The article presents the process of creating the CVET corpus, which contains the texts of Father Hijacint Repič published in the religious journal Cvetje z vertov sv. Frančiška between 1881 and 1916. The corpus then serves as the basis for a linguistic and stylistic analysis, carried out with the noSketch Engine tool. Frequency analysis of selected variables is used to describe

* Asistentka, Inštitut za jezikoslovne študije, Znanstveno-raziskovalno središče Koper, Garibaldijeva 1, SI-6000 Koper, diana.kosir@zrs-kp.si; ORCID: 0009-0009-4428-9698

** Dr., strokovno-raziskovalni svetnik, Institut »Jožef Stefan«, Odsek za tehnologije znanja, Jamova 39, SI-1000 Ljubljana, tomaz.erjavec@ijs.si; ORCID: 0000-0002-1560-4099

Father Repič's vocabulary and narrative style. Finally, the synchronic-diachronic and normative aspects of the older Slovenian language in the texts of the corpus are considered using the example of words of the type bralec/bravec.

Keywords: *older Slovenian language, religious press, corpus linguistics, corpus stylistics, linguistic norm*

Uvod

Razvoj slovenskega knjižnega jezika do konca 19. stoletja je mogoče opazovati skozi tri pomembna obdobja, začenši z dobo utemeljevanja kranjske knjižne norme od Bohoriča do Kopitarja (1584–1808), ki sta ji sledila »puristično« obdobje s Kopitarjem, Miklošičem in Janežičem (1808–1854–1863) s pretenzijami po skupnem knjižnem jeziku in ukinjanju drugih deželnih knjižnih različic ter obdobje razvoja poklicnega slovanskega in slovenskega jezikoslovja (začetek Orožen označi z Miklošičevim nastopom profesure na dunajski univerzi v letih 1850–1880).¹

Od sredine 19. stoletja je slovenski prostor zaznamovalo obdobje narodne prebuje s čitalniškim in taborskim gibanjem, vzponom prosvetno-kulturnih in gospodarskih društvenih organizacij ter porastom slovenskega knjižnega in periodičnega tiska.² Zlasti na obrobju slovenskega kulturnega prostora je imel tisk zaradi jezikovnega elementa tudi narodno identifikacijsko in narodnopolitovno vlogo.³ Slovenski jezik je tako za Slovence pomenil enega od ključnih gradnikov osebne in nacionalne identitete ter narodne zavesti.⁴

V obdobju do prve svetovne vojne se je ob slovenskem gospodarskopoličnem in drugem periodičnem časopisju, denimo *Kmetijskih in rokodelskih novicah* (Ljubljana), *Slovenskem gospodarju* (Maribor), *Slovenskem narodu* (Maribor), *Slovincu* (Ljubljana), *Soči* (Gorica), *Gorici* (Gorica), *Primorskem listu* (Gorica), *Edinosti* (Trst), vzpostavila mreža literarnih revij, na primer *Slovenska bčela* (Celovec), *Slovenski glasnik* (Celovec), *Glasnik* (Maribor), *Zvon* (Dunaj), *Ljubljanski zvon* (Ljubljana), *Kres* (Celovec), *Dom in svet* (Ljubljana), *Slovenka* (Trst), strokovnih in stanovskih glasil, na primer *Časopis za zgodovino in narodopisje* (Maribor), *Učiteljski tovariš* (Ljubljana), ter verske periodike, denimo *Cerkveni glasbenik* (Ljubljana), *Zgodnja Danica* (Ljubljana), *Krščanski detoljub* (Ljubljana), *Angelček* (Ljubljana), *Družinski prijatelj* (Trst), *Svetilnik* (Trst), *Rimski katolik* (Gorica), *Drobtinice* (Gradec, Maribor, Ljubljana) in *Cvetje z vertov*

1 Martina Orožen, *Oblikovanje enotnega slovenskega knjižnega jezika v 19. stoletju* (Ljubljana: Filozofska fakulteta, Znanstveni inštitut Filozofske fakultete, 1996), 13–22.

2 Urška Perenič, *Empirično-sistemsko raziskovanje literature: Konceptualne podlage, teoretski modeli in uporabni primeri* (Ljubljana: Slavistično društvo Slovenije, 2010), 101–13, 174–80. Marijan Dovič, *Slovenski pisatelji: Razvoj vloge literarnega proizvajalca v slovenskem literarnem sistemu* (Ljubljana: Založba ZRC, ZRC SAZU, 2007), 119–27.

3 Urška Perenič, »Čitalništvo v perspektivi družbenogeografskih dejavnikov.« *Slavistična revija* 60, št. 3 (2012): 365–82.

4 Vesna Mikolič, »Povezanost narodne in jezikovne zavesti,« *Jezik in slovstvo* 45, št. 5 (2000): 180–82.

sv. *Frančiška* (Gorica, Kamnik).⁵ Uredniki (na primer Bleiweis, Einspieler, Janežič, Jurčič, Stritar, Levec, Sket, Slomšek, Škrabec idr.) so pogosto tudi lektorsko posegali v besedila skladno s svojimi nazori, zato lahko časopisje predstavlja relevanten vir za opazovanje razvoja slovenskega knjižnega jezika in postopnega sprejemanja t. i. »novih oblik«.

Jezikoslovec p. Stanislav Škrabec je svoje jezikoslovne razprave objavljale na platnicah revije *Cvetje z vertov sv. Frančiška* (v nadaljevanju *CFr*), uredniško pa je bdel nad jezikovno podobo besedil znotraj revije. O tem je p. Kunstelj zapisal: »Kar je on učil, je predvsem sam izvrševal. Zato je pa presedel dostikrat skoro po cele noči, da je preliil in predelal spise drugih po svojih zahtevah, zlasti pa, da je svoje lastne, dostikrat precej obširne sestavke, svojim nazorom primerno priredil.«⁶

Opis raziskovalnega gradiva in metod

Osrednje gradivo za raziskavo predstavlja manjši specializirani korpus starejšega slovenskega jezika CVET 1.0,⁷ ki vsebuje vse objave p. Hijacinta Repiča v reviji *CFr* v letih 1881–1916 (ur. Škrabec). Pričujoči članek je nastal na osnovi prispevka avtorjev na konferenci Jezikovne tehnologije in digitalna humanistika (2024),⁸ pri čemer gre za razširitev stilistične analize z dodanimi spremenljivkami za opis pripovednega sloga patra Repiča, kot jo ponuja korpusni pristop z uporabo orodij noSketch Engine in Sketch Engine. Dodan je tudi primer korpusne analize zgodovinskih leksikalnih oblik kot odraz udejanjanja Škrabčeve knjižne norme v praksi, ob primerjavi z drugimi referenčnimi jezikovnimi priročniki.

Izdelava in opis korpusa CVET 1.0

Priprava gradiva za korpus je zajemala izbor besedil (pridobljenih s strani dLib), pretvorbo v besedilne datoteke Word in kritični prepis (podrobnejši opis priprave gradiva za korpus prinaša konferenčni prispevek).⁹ Vsaka datoteka Word je bila označena z metapodatki, zbranimi v preglednici Excel (identifikator besedila – ime datoteke, avtor, naslov članka, mesto objave (leto, letnik, številka), URL izvornega mesta objave v dLib in številke strani, na katerih se članek pojavi).

5 Miša Šalamun, *Slovensko primorsko časopisje. Zgodovinski pregled in bibliografski opis* (Koper: Lipa, 1961), 25, 26. Dovič, *Slovenski pisatelji*, 120–23.

6 Bruno Korošak, P. Stanislav Škrabec, *frančiškan, v očeh sodobnikov* (Ljubljana: Založba Brat Frančišek, 2001), 19.

7 Diana Košir in Tomaž Erjavec, »Corpus of texts by Hijacint Repič CVET 1.0« (2024), *Slovenian language resource repository CLARIN.SI*, <http://hdl.handle.net/11356/1226>.

8 Diana Košir in Tomaž Erjavec, »Korpus CVET 1.0: izdelava, opis in analiza zbirke starejših besedil v verski periodiki«, prispevek objavljen na letni konferenci JT-DH 2024, *Conference on Language Technologies and Digital Humanities*, Ljubljana, Slovenija, 19.–20. 9. 2024, <https://doi.org/10.5281/zenodo.13936400>.

9 Ibidem.

Nato je bilo gradivo iz formatov Word in Excel pretvorjeno v XML-shemo, skladno s priporočili za kodiranje besedil TEI.¹⁰ Datoteke Word so bile pretvorjene v osnovni TEI z uporabo standardnih skript XSLT za pretvorbo v dokumente TEI in iz njih,¹¹ razpredelnica Excel z metapodatki o posameznih besedilih pa je bila shranjena kot datoteka TSV. Z namensko skripto so bile besedilne datoteke TEI povezane z metapodatki v eno datoteko TEI (vrhnji element <TEI>), ki vsebuje kolofon (element <teiHeader>) in besedilo (<text>). Začetek ene od teh datotek ilustrira Slika 1.

Slika 1: Primer začetka zapisa posameznega besedila v formatu TEI

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="CVET-1887_7_2_43-46" xml:lang="sl">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Vzroki in koristi terpljenja pobožnih duš. (1887) [CVET]</title>
        <author>Repič, Hijacint</author>
      </titleStmt>
      <editionStmt>
        <edition>1.0</edition>
      </editionStmt>
      <extent>
        <measure unit="words">1203</measure>
      </extent>
      <publicationStmt>
        <distributor>CLARIN.SI</distributor>
        <idno type="handle">http://hdl.handle.net/11356/2026</idno>
        <date when="2024-05-07">7. maj 2024</date>
      </publicationStmt>
      <sourceDesc>
        <bibl type="article">
          <author>Repič, Hijacint</author>
          <title level="a">Vzroki in koristi terpljenja pobožnih duš.</title>
          <bibl type="monogr">
            <title level="j">Cvetje z vertov sv. Frančiška</title>
            <biblScope unit="volume">7</biblScope>
            <biblScope unit="number">2</biblScope>
            <biblScope unit="page" from="11" to="14">11-14</biblScope>
            <date when="1887">1887</date>
            <idno type="URI" subtype="URN">URN:NBN:SI:DOC-HLXF9DN</idno>
          </bibl>
        </bibl>
      </sourceDesc>
    </fileDesc>
    ...
```

Vir: lastno delo

Korpus je formiran kot dokument XML s krovno datoteko (element <teiCorpus>), ki vsebuje kolofon korpusa in povezave (elementi XInclude) na besedilne datoteke korpusa.

Gradivo v starejši slovenščini vsebuje arhaične leksikalne oblike, zato je bilo zaradi lažje nadaljnje uporabe in jezikoslovnega označevanja zbirke avtomatsko posodobljeno z odprtokodnim orodjem za avtomatsko posodabljanje leksike cSMTiser,¹² ki je posodobilo 13.033 oziroma 7,4 odstotka besed. Izvirne besedne oblike s tem niso

10 TEI Consortium, eds., *Guidelines for Electronic Text Encoding and Interchange*, <http://www.tei-c.org/P5/>.

11 *GitHub - TEIC/Stylesheets: TEI XSL Stylesheets*, <https://github.com/TEIC/Stylesheets>.

12 Yves Scherrer in Nikola Ljubešić, »Automatic Normalisation of the Swiss German ArchiMob Corpus Using Character-Level Machine Translation,« prispevek objavljen na letni konferenci KONVENS 2016, *13th Conference on Natural Language Processing*, Bochum, Germany, 19.–21. 9. 2016.

bile izgubljene, pač pa so posodobljene oblike, kjer se razlikujejo od izvirnih, bile pripisane tem.

V postopku avtomatskega posodabljanja starejšega jezika so bile ustrezno posodobljene na primer besede, ki imajo polglasnik zapisan z »e« (npr. *miloserčnost* > *milosrčnost*, *serce* > *srce*, *smert* > *smrt*), končnica -vec > -lec (*bravec* > *bralec*), besede s premeno -t- > -d- v korenu (npr. *britkosti* > *bridkosti*), -z- > -g- v korenu (npr. *druzega*, *vbozih* > *drugega*, *ubogih*), izpad -i- v stariši > starši (mn.) in *kristijani* > *kristjani*, izpad -t- v *bogatstvo* > *bogastvo*, zapis v- > u- pri nekaterih glagolih (npr. *vmreti*, *vmerl* > *umreti*, *umrl*), oblika *aposteljnov* > *apostolov* in nekatere funkcijske besede: zaimsek *gdo* > *kdo*, veznik *temuč* > *temveč*, veznik in členek *toraj/torej* > *torej*, členek *vže* > *že*, predlog *sè* > *s*, *ž* > *z*. Nekatere starejše oblike besed so bile v postopku strojnega posodabljanja leksike spregledane (npr. predlog *mej* > *med*, *blager* > *blagor*). Te bodo skupaj z nekaterimi zaznanimi napačnimi posodobitvami (pregledno jih navaja konferenčni prispevek) popravljene v prihodnji verziji korpusa.

Nato je bil korpus jezikoslovno označen z odprtokodnim orodjem CLASSLA,¹³ s katerim so bile vsaki besedi dodane njena lema oziroma osnovna oblika, njene oblikoskladenjske lastnosti in skladdenjska razčlenitev povedi po sistemu Universal Dependencies za slovenski jezik.¹⁴

Korpus vsebuje 230 besedil oziroma 3228 odstavkov, 10.109 (avtomatsko določenih) povedi in 175.907 besed. Različica 1.0 je bila objavljena v repozitoriju CLARIN.SI pod odprto licenco Creative Commons, priznanje avtorstva (CC BY).¹⁵ Za prevzem je na voljo v štirih stisnjenih datotekah. Vsaka vsebuje direktorij, v njem pa datoteke za eno od variant korpusa (opis prinaša konferenčni prispevek). Repozitorijski vnos je povezan s konkordančniki CLARIN.SI noSketch Engine in KonText, kar omogoča poizvedbe in analize korpusa brez znanja programiranja – z različnimi vizualizacijami in možnostjo shranjevanja rezultatov poizvedb.

V nadaljevanju je korpus CVET 1.0 podlaga za a) analizo pripovednega sloga p. Hijacinta Repiča in b) komparativno analizo izbranih zgodovinskih oblik v CFr. Ob Škrabčevih načelih je preverjena norma v naslednjih normativnih priročnikih: Janežičevi *Slovenski slovnici za domačo in šolsko rabo* (1863), Pleteršnikovem *Slovensko-nemškem slovarju* (1894–1895), Levčevem *Slovenskem pravopisu* (1899) in Breznikovi *Slovenski slovnici za srednje šole* (1916). Prikazan je primer analize zgodovinske paradigme na glasoslovni ravni, na kateri je bilo tudi sicer zaznanih največ razlik v primerjavi s sodobno knjižno normo, ki pa jih je orodje za avtomatsko posodabljanje leksike v korpusu CVET v glavnini pravilno prepoznalo.

13 Nikola Ljubešić, Luka Terčon in Kaja Dobrovoljc, »CLASSLA-Stanza: The Next Step for Linguistic Processing of South Slavic Languages,« prispevek objavljen na konferenci JT-DH 2024, *Conference on Language Technologies and Digital Humanities*, Ljubljana, Slovenija, 19.–20. 9. 2024, <https://doi.org/10.5281/zenodo.13936406>.

14 Kaja Dobrovoljc, Tomaž Erjavec in Simon Krek, »The Universal Dependencies Treebank for Slovenian,« prispevek objavljen na konferenci *Proceedings of the 6th Workshop on Balto-Slavic Natural Language Processing*, Association for Computational Linguistics, 33–38, <https://doi.org/10.18653/v1/W17-1406>.

15 Diana Košir in Tomaž Erjavec, »Corpus of texts by Hijacint Repič CVET 1.0« (2024), *Slovenian language resource repository CLARIN.SI*, <http://hdl.handle.net/11356/1226>.

Stilometrija in korpusna stilistika

Na presečišču jezikoslovja in literarnih ved je literarna stilistika, opredeljena kot jezikoslovno preučevanje sloga oziroma namenske specifične rabe jezika kot osrednjega medija literature, ki nakazuje odnos med kreativnim dosežkom (učinkom) in jezikovno manifestacijo,¹⁶ na to pa lahko vplivajo različne nejezikovne spremenljivke, kot so žanr, avtor, zgodovinsko obdobje ipd.¹⁷ Pri stilistični analizi opazujemo sestavine sloga (angl. *features of style*), jezikovne elemente, ki v danem korpusu izstopajo in prispevajo k oblikovanju sloga.¹⁸ Slogovne označevalce (angl. *style markers*) Leech in Short za potrebe stilistične analize razvrščata v štiri kategorije: a) leksikalne prvine, b) slovnične prvine, c) retorična sredstva (figure in tropi) ter č) kohezija in kontekst.¹⁹

V okviru računalniške kvantitativne stilistike se ločita pristopa stilometrije in korpusne stilistike. Stilometrija je preštevalna metoda določanja stila in temelji na vnaprej izbranih markerjih, kot so najpogostejše besede, njihova pogostnost in razpršenost v danem korpusu, raznolikost (gostota) besedišča in *hapax legomena* oziroma *hapax dislegomena*, povprečna dolžina besed oziroma povedi, črkovni in besedni n-grami, pogostost najpogostejših besednih kolokacij itd.²⁰ Korpusna stilistika pa je razumljena kot uporaba teorij, modelov in metod stilistike pri korpusni analizi,²¹ pri čemer je poudarek na kvalitativni analizi kvantitativnih podatkov oziroma literarnovedni interpretaciji, kjer v raziskavo vstopijo naratološke teme, kulturno-družbeni in kognitivni vidiki literature.²²

Jasno razmejitev med obema naboroma metod in pristopov je težko določiti. Analize v članku združujejo pristope korpusnega jezikoslovja, stilometrije in korpusne stilistike, slednje zlasti pri interpretaciji izbranih prvin za zvrstno-žanrski opis besedil. Za analizo so bile izbrane štiri spremenljivke, povezane s frekvenčnostjo: najpogostejših 100 besed, najmanj pogoste besede, najpogostejši besedni 2-grami in ključne besede. Frekvenčni sezname so bili pridobljeni z orodjem noSketch Engine, s plačljivo različico Sketch Engine pa so bili analizirani besedni 2-grami. CLARIN.SI konkor-dančnik noSketch Engine omogoča funkcijo iskanja ključnih besed (*keywords*), pridobljenih glede na referenčni korpus PriLit 1.0,²³ ki vsebuje starejša slovenska besedila (rokopise, pridige, krajšo pripovedno prozo) od sredine 17. do sredine 19. stoletja.

16 Geoffrey N. Leech in Mick Short, *Style in Fiction: A Linguistic Introduction to English Fictional Prose* (Harlow: Pearson Education Limited, 2007), 11, 55.

17 Lesley Jeffries in Daniel McIntyre, *Stylistics* (Cambridge: Cambridge University Press, 2010), 1.

18 Leech in Short, *Style in Fiction*, 44, 56.

19 Ibid., 61–66.

20 David L. Hoover, »Frequent Word Sequences and Statistical Stylistics«, *Literary and Linguistic Computing* 17, št. 2 (2002). Jack Grieve, »Quantitative Authorship Attribution: An Evaluation of Techniques«, *Literary and Linguistic Computing* 22, št. 3 (2007). Andrejka Žejn, »Računalniško podprta stilometrična analiza pripovedne literature Janeza Ciglerja in Christoph Schmid v slovenščini«, *Fluminensia: časopis za filološka istraživanja* 32, št. 2 (2020).

21 Dan McIntyre in Brian Walker, *Corpus Stylistics: Theory and Practice* (Edinburgh: Edinburgh University Press, 2019), 15.

22 Berenike J. Herrmann, Arthur M. Jacobs in Andrew Piper, »Computational Stylistics«, v: Donald Kuiken in Arthur M. Jacobs, ur., *Handbook of Empirical Literary Studies* (Berlin, Boston: De Gruyter, 2021), 460–69.

23 Andrejka Žejn in Tomaž Erjavec, »The corpus of older Slovenian narrative prose PriLit 1.0« (2021), *Slovenian language resource repository CLARIN.SI*, <http://hdl.handle.net/11356/1319>.

Analiza opazovanih parametrov služi opisu leksike (slogovnega in normativnega vidika), vpogledu v raznolikost besedišča in tematskemu opisu korpusa, kar razodeva avtorjevo pomensko polje, njegov spoznavni in literarni svet.²⁴ Pri izboru slogovnih označevalcev in njihovi interpretaciji pa je do neke mere treba upoštevati tudi subjektivno presojo in interpretativni okvir raziskovalca, na kar sta opozorila tudi Leech in Short.²⁵

Korpusna analiza avtorjevega pripovednega sloga

Med najpogostejšimi 100 besedami (lemami) v korpusu CVET 1.0 so: a) glagoli *biti, imeti, moči, hoteti, reči, morati, priti, storiti, prositi, govoriti, videti, delati, vedeti, ljubiti*; b) samostalniki *Bog, svet, duša, človek, brat, Frančišek, gospod, dan, življenje, greh, srce* (izv. *serce*), *oče, volja, beseda, Jezus, milost, otrok, Kristus, križ, ljubezen, mati*, samostalniški zaimki *jaz, ti*; c) pridevniki *božji, dober, velik*, svojilni zaimki *moj, tvoj, svoj, njegov, naš*; č) prislovi *jako, dobro*; d) vezniki *da, in, ter, v, ako*, oziralni in vprašalni zaimki v vlogi veznika *kateri* (izv. *keteri*), *kakor* (izv. *kaker*), *kar, kaj, kako*; e) členki *tudi, naj, še, ne, le, samo*. Lema *nič* se pojavi kot samostalnik (*Oh saj gre vse sčasoma v nič!*), sam. zaimek (*Zato se ne veseli v ničemer drugem, kar je pod nebom!*) in prislov (*Keder se nič več ne želi, takrat se čuti pravo, popolno veselje*), zaimek *vsak* v pridevniški (*Priporočeval se jima je serčno vsako jutro in vsak večer*) in samostalniški rabi (*Ako je vsacemu dovoljeno storiti si iz sukna tako ali tako obleko, bo li Bog imel menj pravic do svojega?*).

Med glagoli se poleg tistih, ki so značilni za vsakdanje sporazumevalne okoliščine (pomožni glagol *biti, imeti, delati, dati*), pogosto pojavijo glagoli rekanja (*reči, prositi, govoriti, tudi praviti, odgovoriti, povedati*) in modalni glagoli (prvo število so pojavitve, drugo pa odstotek besedil v korpusu, kjer se lema pojavi): *moči* (597 = 75 %), *hoteti* (577 = 70,87 %) in *morati* (465 = 66,52 %). Kontekst rabe teh naklonskih glagolov je pričakovan za nabožno vzgojno literaturo, npr. *Naposled, ker brez milosti božje ne moreš nič dobrega storiti, prosi Boga, da ti bode milostljiv / Ali Bog te hoče poskušati s skušnjavami, je li res, da ga ljubiš ali ni / Nadalje se moramo vdati božji volji, ako imamo dušne ali telesne pogoške*. V primerih je opazno prehajanje med prvo- in drugoosebnimi glagolskimi oblikami oziroma načini nagovarjanja bralca. Če v kontekstu diskurza s prevladujočo vplivajnsko vlogo opazujemo različne pripovedne perspektive ('mi', 'ti' in 'vi') za glagol *morati*, ki je z vidika naklonskosti najbolj neposreden v izražanju nujnosti, ugotovimo, da se oblika *moramo* (1. os. mn.), kjer sta pripovedovalec in bralec del množinskega subjekta, pojavi 123-krat, oblika *morate* (2. os. mn.), kjer se pripovedovalec distancira od množinskega prejemnika in se postavi v superiorno vlogo, le 11-krat, oblika *moraš* (2. os. ed.), kjer gre v primerjavi s prejšnjima za najbolj neposredno obliko nagovarjanja, pa 56-krat. Med opazovanimi je najpogostejša

24 Vesna Mikolič, *Izrazi moči slovenskega jezika* (Koper: Annales ZRS, Ljubljana: Slovenska matica, 2020). Vesna Mikolič, *Ali bereš Cankarja?* (Ljubljana: Slovenska matica, 2022).

25 Leech in Short, *Style in Fiction*, 3, 34–36.

prvoosebna množinska oblika, kjer gre za približevanje sporočevalca naslovniku, razmeroma pogosto pa pripovedovalec rabi tudi bolj neposredno, edninsko obliko, ki deluje kot okrepitev sporočila (npr. *Ako hočeš torej uživati Boga, se moraš odpovedati grešnim posvetnim tolažbam*).

Za ugotovitev, ali raba naklonskega glagola *morati* v korpusu CVET statistično značilno izstopa, sta bila za primerjavo izbrana dva korpusa starejšega slovenskega jezika: korpus KDSP²⁶ s pripovedno prozo 19. stoletja in korpus Slovenska periodika (1771–1914).²⁷ V korpusu KDSP se glagol *morati* pojavi v 99,6 odstotka besedil in predstavlja 0,16 odstotka korpusa, v korpusu periodike je prisoten v 73,19 odstotka besedil in obsega 0,15 odstotka korpusa, v korpusu CVET pa se *morati* pojavi v 66,5 odstotka besedil, medtem ko vse pojavitve predstavljajo 0,22 odstotka korpusa. V korpusu periodike in v CVET je prvi podatek (odstotek besedil s pojavnico) pričakovano bistveno nižji kot v KDSP, ker vsak posamezni članek predstavlja svojo besedilno enoto. Je pa v korpusu CVET opazen najvišji odstotek pojavnosti glagola v razmerju do celotnega korpusa. Kvantitativna analiza je, upoštevajoč zvrstno raznolikost primerjalno gradivo, pokazala višjo prisotnost naklonskega glagola *morati* v pripovednem slogu patra Repiča v besedilih (polliterarne) zvrsti versko-vzgojnega periodičnega članka, kot se je to pokazalo za korpus literarnih besedil in periodične članke v glavnem necerkvene vsebine. Smiselno bi bilo zato opraviti dodatne kvalitativne analize kontekstov rabe, da bi se lahko utemeljilo razmerje med sporočanjso in vplivajnsko vlogo besedil ter piščevo izbiro posameznih naklonskih sredstev kot stilistično karakteristiko avtorja in literarne zvrsti oziroma žanra.

Religiozni diskurz tipično opredeljuje moralno-vrednostno razmerje dobro : slabo in med najpogostejšimi besedami v korpusu CVET so predvsem te, ki semantično sodijo v vrednostno oznako »dobro« (ljubiti, Bog, duša, srce, milost, ljubezen, dober, dobro).

Večina od prvih 100 besed se pojavi v več kot 60 odstotkih vseh besedil, kar pomeni, da je njihova zastopanost razpršena po celotnem korpusu.

Po pogostosti izstopa raba zaimka *kateri*, ki se pojavi kar 1822-krat in v 90,87 odstotka besedil. Izrazita je vezniška raba, kjer uvaja podredje z oziralnim odvisnikom, v današnji knjižni slovenščini bi ga nadomestil zaimek *ki* (npr. *Ali ti, moj Jezus, kateremu je vse odkrito, in kateri si vstvaril vse v meri in številu ...*). Tudi sicer hiter pregled gradiva pokaže, da so v Repičevih tekstih pogoste dolge povedi s podredji, kar potrjuje statistični podatek za povprečno dolžino povedi (izračunan na podlagi števila besed in povedi v korpusu), ki znaša 17 besed (npr. *Kar se tiče pa vsacega od nas v tacih občnih nesrečah, pomislimo sè zaupanjem na Boga, kateri zna število naših las, in katerih nam niti eden ne pade z glave brez njegove presvete volje, kar pomeni, da se nam čisto nič ne more pripetiti, kar Bog neče ali ne dopusti.*).

26 Lucija Mandić in Tomaž Erjavec, »Corpus of longer narrative Slovenian prose KDSP 1.0« (2023), *Slovenian language resource repository CLARIN.SI*, <http://hdl.handle.net/11356/1823>.

27 Filip Dobranič, Bojan Evkoski in Nikola Ljubešić, »Corpus of Slovenian periodicals (1771–1914) sPeriodika 1.0« (2023), *Slovenian language resource repository CLARIN.SI*, <http://hdl.handle.net/11356/1881>.

Med načinovnimi prislovi po frekvenčnosti izstopa leksem *jako* (342-krat; ob enajstih pojavitvah sopomenke *zelo*; npr. *To prepričanje nas jako tolaži, ker vemo, da nas Bog more tako lahko poklicati k sebi, ko smo v sreči in zdravju, kaker ko smo v občnih nesrečah*). Glede na konkordance in metapodatke o izvirnih besedilih, ki jih posreduje korpus, je mogoče ugotoviti, da se starejša oblika 'jako' pojavlja po vsem gradivu (v 57,83 odstotka besedil), novejša oblika 'zelo' pa ob njej nekoliko bolj konsistentno po letu 1913, kar kaže na utrjevanje nove oblike v knjižni normi. Ker se prislov *jako*, kot bo razvidno v nadaljevanju, pojavi tudi med ključnimi besedami glede na referenčni korpus starejših pripovednih besedil, ga je mogoče označiti kot značilno potezo patro-vega pripovednega sloga.

Med najmanj pogostimi besedami (ena do dve pojavitvi, korpus takih lem navaja 4656) so nepričakovane tvorjenke: manjšalnice (človeček, stvarca), kletvina (ob kletev), pekočina, bogatin (ob bogataš in posamostaljenem pridevniku bogati (mn.)), tolažilo (ob tolažba), zaveržek, pomoček, kroatizem škatulje (mn.), nasladnost (ob naslada), ostrašen (ob prestrašen), modifikacija glagolskega vida (peljavati, zmaščevati, oveseliti), onomatopoetični glagol zberbrati (= na hitro, nerazločno oziroma nepremišljeno izreči, zmoliti); deležnik na -č: kretajoč (se), pazeč, deležnik na -ši: znebiti, zbudit; členek znabiti, izvirni besedni zvezi čudapoln mir in prvak apostolov, zvezdoznanka, zgoja (ob vzgoja), zgojitelj, razjasnjenje, život (ob telo), nesposobnost, neizmernost, gotovost, enakoličnost itd. Nepričakovane oziroma z vidika rabe redke dvojnice lahko razumemo kot odraz tedaj še nepoenotene pisne norme (npr. vzgoja ali zgoja), nekatere glagolske oblike in izvirne tvorjenke, ki lahko glede na dani kontekst okrepijo pomen, pa kažejo na bogato besedišče in piščevo ustvarjalno rabo jezika. Skozi uporabo ekspresivne, slogovno zaznamovane leksike se pisec hkrati lahko čustveno razodeva, izraža naklonjenost oziroma nenaklonjenost.

Tretja opazovana spremenljivka so besedni 2-grami (najpogostejše zveze dveh lem). Namen je pridobiti vpogled v najpogostejše in ustaljene ter redke in nepričakovane besedne zveze, ki tematsko (še bolj ilustrativno kot posamezne besede) označujejo besedilni korpus, z vidika pripovednega sloga pa je mogoče v danih kontekstih opazovati njihovo semantično plat in estetsko funkcijo (dejanski in preneseni pomen, učinek v besedilu). Izluščene so bile zveze 'samostalnik + pridevnik' in 'samostalnik + samostalnik', pri čemer so bila izločena osebna lastna imena. V analizi besednih zvez so zaradi večje preglednosti navedene posodobljene leksikalne oblike.

Med najpogostejšimi besednimi zvezami so samostalniške zveze s pridevnikom *božji/a/e* (volja, milost, dobrota, ljubezen, milosrčnost, previdnost, usmiljenje, čast, blagoslov, sodba, naredba, roka, sin, služabnik, mož, prijatelj, svetnik; miza, služba, pot, dar, beseda, zapoved, pomoč, martra), zveze z lastnostnim pridevnikom *velik/a/o* (ki se v korpusu v osnovni in stopnjevani obliki pojavi kar 727-krat) za izražanje visoke stopnje (milost, veselje, ljubezen, svetost, zaupanje, dobrota, ponižnost, milosrčnost, slava, sreča, spoštovanje, žalost; grešnik, greh, dolžnost, tolažba, čudež, ubožstvo, čudodelnik, čednost, imenitnost, prijatelj, skušnjava, sladkost, popolnost, razloček, sočutje, strah), za izražanje skladnosti z zaželenim oziroma resničnim so pogoste zveze

s *pravi/a/o* (pobožnost, gorečnost, veselje; prijatelj, ponižnost, služabnik, hrana, vera), enako učinkuje tudi zveza *neovrgljiva resnica*. Podobno velja za besedne zveze z drugimi pridevniki za stopnjevanje navzgor: besedotvorni pridevniški presežniki z obrazilom (presveta volja, presveto ime, presladko ime, presladke solze, presladka beseda, velikansko dejanje) in pomenski presežniki²⁸ (*večen/a/o* življenje, slava, modrost, plačilo, krona, resnica, ljubezen, pogubljenje, ogenj) – v krščanskem diskurzu so pozitivni atributi vezani na Boga in posmrtno življenje v nebesih, negativni pa na pekel; *nepopisljiv/a/o* (veselje, radost, žalost, ginjenost, bridkost, sladkost, želja), *neskončna dobrota*, *neizmerna* (dobrota, ljubezen), *neomejeno zaupanje*, *izredna milost*, *imeniten stan*, *dragocen zaklad*, *strašna bolečina*. Pogoste so zveze s pomensko nasprotnimi pridevniki *svet(i)/a/o* (obhajilo, maša, cerkev, zakrament, pismo, razpelo, volja, življenje, apostol, mož) in *pobožen(/-ni)/na/no* (človek, kristjan, redovnik, ljudstvo, starš, mož, žena, kmet; molitev, pesem, misel, duša) ter *posveten/na/no* (čast, slava, veselje, naslada, razveseljevanje, veselica, človek, nečimrnost, tolažba) in *grešen(/-ni)/na/no* (življenje, naslada, natura, priložnost, veselica). Izluščene so bile še nekatere frekventne besedne zveze, ki kažejo na subjektivno vrednotenje (*otročje zaupanje* (= pozitivno, pristno), *neprecejena nedolžnost*, *bridko trpljenje*, *prazna hvala*, *dober* (kristjan, človek, oče; glas, uspeh), *uboga žena*, *lepa čednost*, *sladke besede*, *goreča molitev*, *goreča pobožnost*, *stanovitna molitev*, ekspresivne besedne zveze (*napuh življenja*, *poželenje oči*, *peklenska hudoba*, *znamenje mlačnosti*, *dejanje čednosti*, *dušna bolezen*, *dušna suhota*), biblijske in izvirne metafore (*dušni pastir*, *božji poslanec*, *prvak apostolov*, *solzna dolina*, *smrtna ura*, *smrtna postelja*, *strupena kača*, *goreča peč* (= trpljenje); *nesrečno spanje* (= brezbožno življenje), *tečajna zvezda* (= Bog), *kukec nečimrnosti* (črv = greh), druge besedne zveze, ki preraščajo v retorične trope, na primer pretiravanje ali hiperbola (*morje bridkosti*, *brezno nesreč*, *brezno ponižnosti*) in bistroumni nesmisel ali oksimoron (*neusmiljeno usmiljenje*, *radovoljna raztresenost*).

Potem so bile v fokusnem korpusu CVET analizirane ključne besede skozi primerjavo z izbranim referenčnim korpusom starejše pripovedne proze PriLit, ki vsebuje starejša besedila od sredine 17. do sredine 19. stoletja in je zato relevanten vir za opazovanje razvoja tem do obdobja, ki ga zajema korpus CVET. Poleg tega sta bila korpusa podobno procesirana in posodobljena z enakim orodjem, kar omogoča večjo primerljivost gradiva.

Med 100 ključnimi besedami se za korpus CVET pojavijo naslednje leme:

- samostalniki: *Frančišek*, *Monald* (izv. tudi *Monaljd*),²⁹ zemljepisna imena *Asiz* (= Assisi), *Koper*, *Padova*; *redovnik*, *frančiškan*, *sobrat*, *tretjerednik*, *zavetnica*, *voditelj*, *častilec* (izv. *častivec*), *posvetnjak*, *vodilo*, *miloščina*, *milosrčnost* (izv. *miloserčnost*), *sočutje*, *blager* (= blagor; trikrat samostalniška raba v pomenu sreča, blagoslov), *način* (prim. *na ta(k)/en/pervi/drugi/vsak način*), *sredstvo*, *oblika* (prim. *živeti po obliki sv. evangelija* = *živeti skladno z nauki in zgledi iz evangelija*), *kesanje*, *naslada*, *pogrešek*, *vdanost*, izglagolska samostalnika *zatajevanje*, *občevanje*;

²⁸ Mikolič, *Izrazi moči*, S7, S8.

²⁹ Krajšava izv. se v članku rabi za 'izvirno', torej tako, kot je zapisano v neposodobljenem korpusu besedil.

- pridevniki: *redoven, blažen, brezmadežen, blagoslovljen, presladek, izreden, nepopisljiv*, (Marija) *Porcijunkuljska* (= nanaša se na cerkev Marije Angelske pri Assisiju, t. i. Porcijunkulo), *vsakovrsten* (izv. *vsakoversten*), *brezštevilen*;
- glagoli: *pridigati, oznanjevati, občevati* (= biti v stiku, sporazumevati se), *spolnjevati* (redko *izpolnjevati*), *spodbujati, rabiti, zanemarjati, zoperstavljati se, vničiti* (= uničiti);
- prislovi: *jako, kesneje* (= kasneje), *naposled*;
- vezniki: *koliker* (= kolikor), *čiger* (= čigar), *keteri* (= kateri), *mariveč* (= marveč);
- členki: *nikaker* (= nikakor), *seveda* (pisano tudi *se ve da*), *potemtakem, blager* (= blagor; 26-krat členkovna raba).

Izkazalo se je, da so bile poleg frekvenčno izstopajočih kot ključne besede prepoznane tudi besede, ki niso bile posodobljene ali pa so v postopku posodabljanja pridobile napačno obliko.

Ugotoviti je mogoče, da polnopomenske besede nedvomno pripadajo religijskemu (krščanskemu) diskurzu, nekatere med njimi še podrobneje usmerijo na red sv. Frančiška Asiškega (npr. *Frančišek, Asiz, frančiškan, vodilo, Porcijunkuljska, redovnik, redoven, sobrat, tretjerednik, voditelj*) in na lokalno okolje, od koder je pisec p. Repič prihajal (*bl. Monald Koprski, Koper*).

Med glagolskimi ključnimi besedami so pogosti glagoli rekanja, vezani na širjenje krščanske vere (liturgijo, pastoralo), ob tem pa tudi nekaj takšnih, ki v danih kontekstih pomenijo dejanja kršenja krščanskih načel.

Pridevniki in pomenski presežniki *blažen, brezmadežen, blagoslovljen, presladek, izreden, nepopisljiv, vsakovrsten, brezštevilen* ter členek *blagor* imajo ob sopojavitvi s pomensko pozitivnimi samostalniki močan pozitiven naboj, izražajo odobravanje in občudovanje koga/česa. V vsakem primeru pa z vidika intenzitete jezika omenjene jezikovne prvine okrepijo pomen ubesedenega.³⁰ V smer krepitve argumenta vodijo tudi pogosta raba prislova *jako* (redko *zelo*) in pomensko nasprotna si členka *nikaker* in *seveda* – slednji skupaj s členkom *potemtakem* nastopa tudi v vlogi diskurznega označevalca, ko usmerja potek diskurza z vrednotenjem, potrjevanjem in sklepanjem ter tako vpliva na bralčevo recepcijo besedila.

Primerjalna analiza jezikovne norme

V nadaljevanju je predstavljen primer korpusne analize za izbrano glasoslovno paradigmo, pri kateri je na prelomu 19. v 20. stoletje prihajalo do razlik med jezikoslovci, in sicer zapis besedotvorne končnice *-lec/-vec* v samostalniki tipa *bralec*.³¹ Orisana je norma v tedaj pomembnejših jezikoslovnih priročnikih, Škrabčeva stališča pa so preverjena v korpusnem gradivu.

³⁰ Mikolič, *Izrazi moči*, 56, 66, 67.

³¹ Razhajanja so se sicer nadaljevala tudi še v 20. stoletje, npr. v polemikah ob izdaji *Slovenskega pravopisa* 1962.

V zvezi s samostalniško končnico *-vec/-lec* v poimenovanju delavnih ali delujočih oseb Janežič v *Slovenski slovnici* (1863) navaja, da se ta tvorijo tako, da se med osnovo in pripono *-ec* pri odprtih zlogih doda črka *v* (npr. *bravec*, *pevec*, *pivec*, *delavec*, *igravec*, *morivec*, *pisavec*, *poslušavec*, *svetovavka*). Ob tem dodaja, da nekateri pisatelji (slovenski in srbohrvaški) takšne samostalnike tvorijo iz preteklega deležnika in pišejo *delalec*, *igralec*, *poslušalec*, *svetovalec*, *tkalec*, *vladalec*.³²

Škrabec se je v razpravah vprašanja lotil prek opazovanja zgodovinskega razvoja oblik in v težnji po enotni razlagi prišel do naslednjih ugotovitev: a) nekatere besede s pripono *-lec* [*-lac*] so se razvile iz poimenovanj za orodje, navadno s samostalniško končnico *-lo* (npr. *tkalec*, *rilec*, *motovilec*); b) nekatere besede s pripono *-vec* [*-vəc*] so nastale iz glagolskega pridevnika na *-av(en)* (npr. *smrkavec*, *delavec*, *igravec*)³³; c) besede za delujočo oziroma delavno osebo lahko imajo pripono *z l* ali *v*, s tem da prva pomeni tistega, ki je glagolsko dejanje že izvršil/prestal (nastala iz deležnika na *-l*, npr. *pogorelec*), druga pa tistega, ki ga vrši ali more vršiti; č) mogoče so tudi dvojnice v primerih, kjer gre za pomensko razločevanje (pripono *z v* naj bi imela poimenovanja, izpeljana iz glagolskih pridevnikov, »ki zaznamujejo nagnjenje, navado, službo ali lastnost osebe, ki kaj stanovitno dela«,³⁴ pripono *z l* pa poimenovanja oseb, ki so kaj dovršile; npr. *morivec* in *morilec*, *prebivavec*, ko je mišljen kraj, in *prebivalec*, ko se nanaša na prebivanje).³⁵ Stramljič Breznik poudarja, da je Škrabec s takšno utemeljitvijo dopuščanja obstoja obeh oblik, tvorjenih iz iste podlage, uveljavljal razlikovalna načela, ki jih je obenem kritiziral pri sodobnikih (npr. pri Perušku).³⁶ V korpusu CVET je mogoče najti naslednja primera rabe: *Ako pa vi stariši to dosežete, namreč pravo kerščansko življenje vaših otrok, srečni vi, ker ste dosegli konec zgoje, ketera je v pravem pomenu besede: skerb, da otroci postanejo nebeški prebivavci; Ne živi kaker da bi bil prebivalec tega sveta*. V korpusnem gradivu semantično razlikovanje med oblikama sicer ni dosledno upoštevano, pojavljata se denimo obe besedni zvezi, *nebeški prebivavec/prebivalec*, kar kaže na ohlapno utemeljitev pomenskega razlikovanja obeh oblik (po letu 1898 je v rabi le oblika na *-vec*).

V korpusu CVET se pojavijo naslednje oblike:

- na *-vec*: *bravec*, *pisavec*, *poslušavec*, *obrekovavec*, *opravljavec*, *klepetavec*, *častivec*, *hvalivec*, *ljubivec*, *zapeljivec*, *lovec*, *tergovavec*, *kupčevavec*, *pohujšljivec*, *molivec*, *lažnjivec*, *delavec*, *izdajavec*, *zaničevavec*, *vdovec*, *slepivec*, *godernjavec*, *togotljivec*, *delivec*, *hinavec*, *zapeljivec*, *vbijavec*, *premišljavec*, *ražaljivec*, *skušnjavec*, *spolnjavec*, *podiravec*, in ženske oblike *nasledovavka*, *napovedovavka*, *hinavka*;
- na *-lec*: *obiskovalec*, *izgojevalec*, *odgojevalec*, *tekalec*, *prilizovalec*, *oporekovalec*, *gledalec*, *spričevalec*, *posvečevalec*;

32 Anton Janežič, *Slovenska slovnica za domačo in šolsko rabo* (Celovec: Janez Leon, 1863), 119.

33 Stanislav Škrabec, *Jezikoslovna dela 1* (Nova Gorica: Frančiškanski samostan Kostanjevica, 1994), 46.

34 Irena Stramljič Breznik, »Škrabčeva obravnava priponskih obrazil (a/i) v/l (əc),« v: Jože Toporišič, ur., *Škrabčeva misel II* (Nova Gorica: Frančiškanski samostan Kostanjevica, 1997), 197.

35 Stanislav Škrabec, *Jezikoslovna dela 2* (Nova Gorica: Frančiškanski samostan Kostanjevica, 1994), 425, 426.

36 Stramljič Breznik, »Škrabčeva obravnava,« 197.

- dvojne oblike:³⁷ *zatajevavec* (1)/*zatajevalec* (1), *posredovavec* (2)/*posredovalec* (1), *prebivavec* (5)/*prebivalec* (3), *svetovavec* (1)/*svetovalec* (2), *spoznavavec* (4)/*spoznavalec* (3), *posnemovavec* (2)/*posnemovalec* (4), *zmagovavec* (1)/*zmagalec* (1).

V kasnejši razpravi Škrabec ugotavlja, da je prava končnica poimenovanj delavnih oseb *-vec* in ne *-lec*,³⁸ kar potrjuje tudi znatno večji obseg oblik na *-vec*, med dvojnicami pa je prevlada tvorjenk z *-vec* opazna po letu 1900. Škrabec je verjel, da lahko le normativni priročnik, kot je bil Pleteršnikov slovar, naredi konec jezikovni pravdi med zapisovanjem *bravec*/*bralec*.³⁹

Pri tem jezikoslovnem vprašanju se je pokazal Škrabčev vpliv na Levca, ki se je odločil za prepoved zapisovanja besednih oblik tipa *bralec* z namenom, da bi ustavil »elkanje«.⁴⁰ V *Slovenskem pravopisu* (1899) pri tvorjenju samostalnikov s priponami navaja, da samostalniki na *-ec* pomenijo delujoče osebe, med korenem in pripono pa stoji soglasnik *-v-* (npr. *bravec*, *igravec*, *svetovavec*, *morivec*, *pisavec*; kot napačne, čeravno v obči rabi, navaja oblike: *delalec*, *igralec*, *morilec*, *pisalec*, *poslušalec*, *svetovalec* itd.). Kot pravilne oblike s pripono *-lec* pa navaja nekatere tvorjenke iz preteklega deležnika (npr. *tkalec*, *pogorelec*, *črešnjejelec*, *gnilec*, *umrlec*, *otelec*, *pismoznalec*, *prišlec*, *rilec*, *vrtelec*, *vžgalec* idr.).⁴¹

Tudi Pleteršnik (po Janežiču in Škrabcu) v svojem slovarju (1894–95) navaja oblike s pripono *-vec* (npr. *bravec*, *poslušavec*, *igravec*, *morivec*, *prebivavec*, *delavec*), enako kot predhodniki pa oblike tipa *tkalec* (*motovilec*, *pogorelec* itd.).

Breznik v *Slovenski slovnici* (1916) zmedo med zapisovanjem *l* ali *v* v poimenovanih delujočih oseb pojasnjuje z enako izreko pripone *-vac/-lác* (od 17. stoletja dalje). Deloma povzema Škrabčeve ugotovitve in dopušča dvojnično rabo v primerih, kjer so se iz prvotnih končnic *-(a/i)vec* v novejši pisavi pojavile *-(a/i)lec* (npr. *bravec*/*bralec*, *poslušavec*/*poslušalec*, *igravec*/*igralec*, *šivavec*/*šivalec*, *zaničevavec*/*zaničevalec*) oziroma kjer je *l* v korenu (npr. *volivec*/*volilec*, *delivec*/*delilec*, *ponavljavec*/*ponavljalec*). Prvotni *v* pa je (kot pri Škrabčevi ugotovitvi o izpeljavi iz glagolskega pridevnika) ostal v besedah *brivec*, *pivec*, *delavec*, *kimavec*, *pevec*, *tajivec*. Samo *-lec* se piše za osebe, ki so kaj storile ali trpele (*umrlec*, *prišlec*, *osamelec*, *prebivalec*) in v izpeljankah iz samostalnikov na *-lo* (*motovilec*, *rilec*, *tkalec* itd.).⁴²

37 V oklepaju je za primerjavo zapisano število pojavnic v korpusu.

38 Stanislav Škrabec, *Jezikoslovna dela* 3 (Nova Gorica: Frančiškanski samostan Kostanjevica, 1995), 132, 133.

39 Ibid., 84–86.

40 Helena Dobrovoljc, *Pravopisje na Slovenskem* (Ljubljana: Založba ZRC, ZRC SAZU, 2004), 49.

41 Fran Levec, *Slovenski pravopis* (Dunaj: Cesarska kraljeva zaloga šolskih knjig, 1899), 54.

42 Anton Breznik, *Slovenska slovnica za srednje šole* (Celovec: Družba sv. Mohorja, 1916), 31.

Razprava in sklepni razmislek

V korpusu CVET so zbrana vsa besedila patra Hijacinta Repiča, ki so izvirni in prevodni zapisi po tujejezičnih predlogah (vir avtor konsistentno navaja v samem besedilu oziroma opombah). Prikazana korpusna analiza je bila usmerjena v raziskovanje avtorjevega pripovednega sloga. Uporabljene so bile funkcije, ki jih ponuja prostodostopni konkordančnik noSketch Engine na CLARIN.SI: seznam najpogostejših in najmanj pogostih besed (lem) in seznam ključnih besed (lem) glede na izbrani referenčni korpus ter analiza najpogostejših besednih zvez (2-gramov) z orodjem Sketch Engine. Analiza frekvenčnosti besedja je podala nekatere pričakovane rezultate, kot so samostalniki in pridevniki ter besedne zveze, značilni za religiozni diskurz, ter glagoli, vezani na vsakdanje praktičnosporazumevalne okoliščine, pri čemer so po pogostosti izstopali modalni glagoli. Z vidika naklonskosti bi bilo smiselno primerjati različne verske pripovednike, kjer bi primerjava jezikoslovnih in diskurzivnih razsežnosti, ki kažejo na razmerje med sporočanjso in vplivanjsko vlogo besedil in na sporočevalčev odnos do vsebine (vzgojno-moralne problematike) ter naslovnika, lahko podala širši vpogled v razvoj krščanske misli.

Analiza najpogostejših samostalnikov in pridevnikov potrjuje vsebinski okvir besedil, skozi kontekste rabe posamezne leksike pa je mogoče dostopati do vseh pojavit v gradivu. Pri Repiču se pojavijo nekatere besedne zveze, kjer imajo posamezne sestavine ekspresiven, dobeseden oziroma prenesen pomen (ustaljene ali izvirne metafore, prisposode): *otročje zaupanje, neprecenjena nedolžnost, prazna hvala, lepa čednost; napuh življenja, poželenje oči, peklenska hudoba, znamenje mlačnosti, dejanje čednosti, dušna bolezen, dušna suhota; prvak apostolov, solzna dolina, smrtna ura, smrtna postelja, strupena kača, goreča peč, nesrečno spanje, tečajna zvezda, kucec nečimrnosti; morje bridkosti, brezno nesreč, brezno ponižnosti; neusmiljeno usmiljenje, radovoljna raztresenost.*

Analiza najpogostejših besednih 2-gramov je pokazala, da med zvezami 'pridevnik + samostalnik' prevladujejo: (1) pridevniki, ki označujejo veliko oziroma presežno lastnost samostalnika (*velik, velikanski, presladek, presvet; večen, nepopisljiv, neskončen, izreden* itd.) ali lastnost, ki je v skladu z zaželenim/resničnim (*pravi, neovrgljiv, dober, lep*); (2) pomensko diametralno nasprotni pridevniki (*svet, pobožen – posveten, grešen*), ki v okviru religioznega diskurza pridobijo simbolni pomen pravih/dobrih in napačnih/slabih življenjskih zgledov, nazorov in izbir ter tako nastopajo v vzgojni funkciji.

Opazovani najpogostejši prislovi, zaimki, vezniki in členki zaznamujejo tipično skladnjo (za Repičev slog so značilne dolge povedi z veliko podredji, pogosta je raba veznika *kateri*) oziroma so označevalci intenzitete, ki bodisi krepijo (npr. *jako, zelo; vsak; seveda, nikakor*) ali šibijo (*le, samo*) končni pomen in argumentativno moč sporočila.⁴³ Šele po primerjalni analizi z drugimi pripovedniki pa bi bilo mogoče sklepati, ali gre tudi za specifične lastnosti Repičevega sloga.

Primer primerjalne normativne analize v članku je osvetlil različne poglede tedanjih jezikoslovcev na zapisovanje samostalniške pripone *-vec/-lec*.

43 Mikolič, *Izrazi moči*, 79–82.

Ugotovljeno je bilo, da je imel Škrabec v jezikoslovnih razpravah tendenco jasno definirati in semantično razmejiti rabo ene in druge končnice v primeru dvojnic, kar se je v praksi izkazalo za 'prisiljeno' in se potrjuje tudi na korpusnem gradivu. Je pa njegovo zagovarjanje zapisa besed tipa *bravec* vplivalo na to, da je tudi Breznik v slovnici še dopuščal dvojnično rabo.

Korpus CVET 1.0 je prva takšna jezikovna zbirka besedil Škrabčevega verskega glasila *Cvetje z vertov sv. Frančiška*, velja pa omeniti, da je bilo besedje *Cvetja* sporadično zajeto že v Pleteršnikovem *Slovensko-nemškem slovarju* (1894–1895). Raziskava je pokazala, kako lahko korpusno gradivo, označeno z metapodatki o avtorstvu in objavi (leto, letnik, zvezek), služi za komparativne sinhrono-diahrone raziskave normiranja slovenskega jezika na prelomu 19. v 20. stoletje v periodičnem tisku in za raziskave udejanjanja knjižne norme, ki jo je urednik Škrabec podrobno predstavil v svojih jezikoslovnih znanstvenih razpravah na platnicah, v sami vsebini *Cvetja*. Hkrati korpus ponuja tematski vpogled v vsebino verske revije (versko-moralni nauki, krščanska vzgoja, hagiografija itd.). V nasprotju s primerljivim verskim periodičnim tiskom, denimo Slomškovimi *Drobtinicami*, ki so pričele izhajati pol stoletja prej in v podobnem formatu kot *Cvetje*, je vsebina slednjega (razen jezikoslovnih razprav na platnicah) še neraziskana. Korpus objav patra Hijacinta Repiča bi lahko z nadgradnjo postal podkorpus obsežnejšega korpusa celotnih izdaj revije *Cvetje z vertov sv. Frančiška*.

Zahvala

Članek je nastal v okviru raziskovalnega programa P5-0409 *Razsežnosti slovenstva med lokalnim in globalnim v začetku tretjega tisočletja*, ki ga financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS) iz državnega proračuna.

Viri in literatura

Literatura

- Dobrovoljc, Helena. *Pravopisje na Slovenskem*. Ljubljana: Založba ZRC, ZRC SAZU, 2004.
- Dobrovoljc, Kaja, Tomaž Erjavec in Simon Krek. »The Universal Dependencies Treebank for Slovenian.« V: *Proceedings of the 6th Workshop on Balto-Slavic Natural Language Processing*. Association for Computational Linguistics, 33–38. <https://doi.org/10.18653/v1/W17-1406>.
- Dovič, Marijan. *Slovenski pisatelj. Razvoj vloge literarnega proizvajalca v slovenskem literarnem sistemu*. Ljubljana: Založba ZRC, ZRC SAZU, 2007. <https://doi.org/10.3986/9789610503200>.
- Grieve, Jack. »Quantitative Authorship Attribution: An Evaluation of Techniques.« *Literary and Linguistic Computing* 22, št. 3 (2007): 251–70. <https://doi.org/10.1093/lc/fqm020>.

- Herrmann, Berenike J., Arthur M. Jacobs in Andrew Piper. »Computational Stylistics.« V: *Handbook of Empirical Literary Studies*, uredila Donald Kuiken in Arthur M. Jacobs, 451–86. Berlin, Boston: De Gruyter, 2021. <https://doi.org/10.1515/9783110645958>.
- Hoover, David L. »Frequent Word Sequences and Statistical Stylistics.« *Literary and Linguistic Computing* 17, št. 2 (2002): 157–80. <https://doi.org/10.1093/lc/17.2.157>.
- Jeffries, Lesley in Daniel McIntyre. *Stylistics*. Cambridge: Cambridge University Press, 2010. <https://doi.org/10.1017/CBO9780511762949>.
- Korošak, Bruno. *P. Stanislav Škrabec, frančiškan, v očeh sodobnikov*. Ljubljana: Založba Brat Frančišek, 2001.
- Košir, Diana in Tomaž Erjavec. »Korpus CVET 1.0: izdelava, opis in analiza zbirke starejših besedil v verski periodiki.« Prispevek objavljen na letni konferenci JT-DH 2024, *Conference on Language Technologies and Digital Humanities*, Ljubljana, Slovenija, 19.–20. 9. 2024. <https://doi.org/10.5281/zenodo.13936400>.
- Leech, Geoffrey N. in Mick Short. *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. 2nd ed. Harlow: Pearson Education Limited, 2007. Pridobljeno 10. 5. 2024. <https://sv-etc.nl/styleinfiction.pdf>.
- Ljubešić, Nikola, Luka Terčon in Kaja Dobrovoljc. »CLASSLA-Stanza: The Next Step for Linguistic Processing of South Slavic Languages.« JT-DH 2024, *Conference on Language Technologies and Digital Humanities*, Ljubljana, Slovenija, 19.–20. 9. 2024. <https://doi.org/10.5281/zenodo.13936406>.
- McIntyre, Dan in Brian Walker. *Corpus Stylistics: Theory and Practice*. Edinburgh: Edinburgh University Press, 2019.
- Mikolič, Vesna. »Povezanost narodne in jezikovne zavesti.« *Jezik in slovstvo* 45, št. 5 (2000): 173–86. Pridobljeno 15. 5. 2024. URN:NBN:SI:DOC-SBD79SLS.
- Mikolič, Vesna. *Izrazi moči slovenskega jezika*. Koper: Annales ZRS; Ljubljana: Slovenska matica, 2020.
- Mikolič, Vesna. *Ali bereš Cankarja?*. Ljubljana: Slovenska matica, 2022.
- Orožen, Martina. *Oblikovanje enotnega slovenskega knjižnega jezika v 19. stoletju*. Ljubljana: Filozofska fakulteta, Znanstveni inštitut Filozofske fakultete, 1996.
- Perenič, Urška. *Empirično-sistemsko raziskovanje literature: Konceptualne podlage, teoretski modeli in uporabni primeri*. Ljubljana: Slavistično društvo Slovenije, 2010. Pridobljeno 10. 5. 2024. <https://zdsds.si/tiskovina/562/>.
- Perenič, Urška. »Čitalništvo v perspektivi družbenogeografskih dejavnikov.« *Slavistična revija* 60, št. 3 (2012): 365–82. Pridobljeno 15. 5. 2024. https://srl.si/ojs/srl/article/view/COBISS_ID_50413154.
- Scherrer, Yves in Nikola Ljubešić. »Automatic Normalisation of the Swiss German ArchiMob Corpus Using Character-Level Machine Translation.« Prispevek objavljen na letni konferenci KONVENS 2016, *13th Conference on Natural Language Processing*, Bochum, Germany, 19.–21. 9. 2016. Pridobljeno 15. 10. 2025. <https://archive-ouverte.unige.ch/unige:90846>.
- Stramljč Breznik, Irena. »Škrabčeva obravnava priponskih obrazil (a/i) v/1 (əc).« V: *Škrabčeva misel II*, uredil Jože Toporišič, 193–200. Nova Gorica: Frančiškanski samostan Kostanjevica, 1997.
- Šalamun, Miša. *Slovensko primorsko časopisje. Zgodovinski pregled in bibliografski opis*. Koper: Lipa, 1961.
- Žejn, Andrejka. »Računalniško podprta stilometrična analiza pripovedne literature Janeza Ciglerja in Christoph Schmidta v slovenščini.« *Fluminensia: časopis za filološka istraživanja* 32, št. 2 (2020): 137–58. Pridobljeno 30. 4. 2024. <https://doi.org/10.31820/f.32.2.5>.

Spletni viri

- Dobranič, Filip, Bojan Evkoski in Nikola Ljubešič. »Corpus of Slovenian periodicals (1771–1914) sPeriodika 1.0« (2023). *Slovenian language resource repository CLARIN.SI*. <http://hdl.handle.net/11356/1881>.
- Košir, Diana in Tomaž Erjavec. »Corpus of texts by Hijacint Repič in 'Cvetje z vertov sv. Frančiška' CVET 1.0« (2024). *Slovenian language resource repository CLARIN.SI*. <http://hdl.handle.net/11356/1226>.
- Mandič, Lucija in Tomaž Erjavec. »Corpus of longer narrative Slovenian prose KDSP 1.0« (2023) *Slovenian language resource repository CLARIN.SI*. <http://hdl.handle.net/11356/1823>.
- Pleteršnik, Maks. *Slovensko-nemški slovar*, 1894–1895. Spletna izdaja. Ljubljana: ZRC SAZU, 2010. Pridobljeno 2. 1. 2023. <http://bos.zrc-sazu.si/pletersnik.html>.
- Žejn, Andrejka in Tomaž Erjavec. »The corpus of older Slovenian narrative prose PriLit 1.0« (2021). *Slovenian language resource repository CLARIN.SI*. <http://hdl.handle.net/11356/1319>.

Tiskani viri

- Breznik, Anton. *Slovenska slovnica za srednje šole*. Celovec: Družba sv. Mohorja, 1916. <https://www.fran.si/slovnice-in-pravopisi/35/1916-breznik>.
- Janežič, Anton. *Slovenska slovnica za domačo in šolsko rabo*. Celovec: Janez Leon, 1863. <https://www.fran.si/slovnice-in-pravopisi/26/18631854-janezic>.
- Levec, Fran. *Slovenski pravopis*. Dunaj: Cesarska kraljeva zaloga šolskih knjig, 1899. <https://fran.si/slovnice-in-pravopisi/34/1899-levec>.
- Škrabec, Stanislav. *Jezikoslovna dela* 1. Nova Gorica: Frančiškanski samostan Kostanjevica, 1994.
- Škrabec, Stanislav. *Jezikoslovna dela* 2. Nova Gorica: Frančiškanski samostan Kostanjevica, 1994.
- Škrabec, Stanislav. *Jezikoslovna dela* 3. Nova Gorica: Frančiškanski samostan Kostanjevica, 1995.

Diana Košir, Tomaž Erjavec

CORPUS ANALYSIS OF NARRATIVE STYLE AND LINGUISTIC NORM IN AN OLDER RELIGIOUS PERIODICAL

SUMMARY

The paper presents the process of creating and linguistically tagging the CVET 1.0 corpus, which contains the texts of Father Hijacint Repič in the older Slovenian language, published in the religious journal *Cvetje z vertov sv. Frančiška* between 1881 and 1916. The texts were obtained in PDF format from the dLib portal, edited in Microsoft Word and then converted to TEI. Older words were automatically updated using an open-source normalisation tool that facilitates corpus search and further analysis of the material. The updated texts were then automatically linguistically annotated according to the Universal Dependencies Formalism for Slovenian. The TEI-encoded versions were converted into various formats and published under an open licence in the CLARIN.SI repository.

The second part of the paper presents an example of corpus analysis of the author's narrative style, using NoSketch Engine and Sketch Engine, based on four frequency variables: the most frequent 100 words, the least frequent words, the most frequent word-2-grams, and keywords. In addition to the expected lexis, the analysis revealed the frequent use of inflectional verbs (especially *morati*), noun-adjective phrases in which the adjective expresses a great or superlative quality of the noun (*velik, velikan-ski, presladek, presvet; večen, nepopisljiv, neskončen, izreden* etc.), and adjectives with diametrically opposed meanings (*svet, pobožen – posveten, grešen*), which in the context of religious discourse take on the symbolic meaning of right/good and wrong/bad examples of life, views and decisions, and thus have an educational function. The long sentence structure with subordinating conjunctions (e.g. *kateri*) has also emerged as a characteristic feature of Repič's style. The study has demonstrated how corpus material provided with bibliographic metadata can be used for comparative synchronic-diachronic studies of Škrabec's linguistic norm (e.g. word type *bralec/bravec*).