

SPLOŠNE INFORMACIJE

Naslov zbirke podatkov: »**Prvaki slovenstva**« v Istri: jezikovna, pragmatična in stilometrična analiza pisne zapuščine patra Hijacinta Repiča ter narodotvorna vloga koprskega samostana sv. Ane : raziskovalni podatki (stilometrija)

Avtor: Diana Košir

ORCID: <https://orcid.org/0009-0009-4428-9698>

Ustanova: Znanstveno-raziskovalno središče Koper, Inštitut za jezikoslovje

E-mail: diana.kosir@zrs-kp.si

Trajanje raziskave: 2021–2025

Financiranje:

ARIS P5-0409 Razsežnosti slovenstva med lokalnim in globalnim v začetku tretjega tisočletja

Ključne besede: zgodovinska sociolingvistika, pragmatika, stilometrija, pripisovanje avtorstva, versko pripovedništvo, pridiga, nabožni članek, starejši slovenski jezik, korpusno jezikoslovje

MOŽNOSTI DOSTOPA

Licenca: CC BY 4.0, Creative Commons Priznanje avtorstva 4.0 Mednarodna

SEZNAM DATOTEK

- **PreberiMe.pdf:** datoteka s splošnimi ter metodološkimi informacijami in opisom podatkov
- **Korpusi za stilometrijo.zip:** komprimirana datoteka, ki vsebuje štiri mape z različnimi korpusi besedil (mape »Korpus 1–4«, v vsaki so besedilne datoteke (txt) z imenom avtorja in naslovom dela)
- **Avtorji in besedila_metapodatki.xlsx:** Excelov dokument z metapodatki o obravnavanih besedilih
- **Posodobljena leksika.txt:** seznam normalizirane leksike

SPLOŠNE INFORMACIJE

Jezik: slovenščina (starejši slovenski jezik, pisava gajica)

Vir podatkov: glej tabelo Excel

Tip besedila: religiozna besedila; leksika

Način zbiranja in urejanja podatkov: ročni; postopek avtomatskega posodabljanja leksike z odprtokodnim orodjem cSMTiser je opravil dr. Tomaž Erjavec

Povzetek:

V okviru doktorske raziskave¹ je bila celovito opisana in interpretirana pisna zapuščina patra Hijacinta Repiča (1863–1918), s posebnim poudarkom na njegovem jeziku in pripovednem slogu. V ta namen so bile opravljene sociolingvistična analiza, primerjalna analiza jezikovne norme ter pragmatična, stilistična in stilometrična analiza, ki so bile podprte s korpusom in uporabo računalniških orodij. Za potrebe raziskave je bil na podlagi omenjenih metod in pristopov oblikovan in preizkušen izvirni interdisciplinarni model preverbe avtorstva starejših besedil.

Objavljeni podatki so rezultat priprave gradiva za stilometrično analizo s programom Stylo, ki je bila v raziskavi opravljena z namenom preverbe morebitnega Repičevega avtorstva slovenskih rokopisnih pridig, najdenih v arhivskem fondu samostana sv. Ane v Kopru.

Osnova za stilometrično analizo je bil nelematiziran korpus »golihi tekstov« (txt). Na podlagi kritičnih prepisov pridig je bil oblikovan manjši korpus »pridige«. Primerjalno gradivo z deli izbranih avtorjev (Hijacint Repič, Anton M. Slomšek, primorska pridigarja Filip Jakob Kaffol in Matija Vertovec, baročni pridigar Janez Svetokriški, Marko Pohlin, Franc Veriti, prevodi Benjamina Franklina in Christopha Schmida) je bilo v primernem formatu (.pdf, .txt) pridobljeno na portalu dLib oz. v obstoječih elektronskih korpusih CVET (Košir in Erjavec 2024), PriLit (Žejn in Erjavec 2021) in eZISS: *Tri pridige o jeziku* (Faganel, Ogrin in Erjavec 2003).

Na podlagi korpusa CVET, ki vsebuje vse Repičeve objave v reviji *Cvetje z vrtov sv. Frančiška*, je bil izdelan manjši korpus z 51 patrovimi izvirnimi besedili, ki ustrezajo žanru nabožnega, hagiografskega in verskovzgojnega članka. Izbrana besedila drugih avtorjev so bila pridige, hagiografije ter poučna versko-vzgojna in moralistična proza.

Opravljenih je bilo 6 testov z različnimi korpusi:

- Test 1 in 2: korpus 1 - posamezni avtorji in njihova dela so samostojne enote (Repičevi članki združeni v 9 datotek glede na letnice objav v *CFr*)

¹ Doktorska disertacija je nastala pod mentorstvom dr. Krištofa Jacka Kozaka in akad. dr. Jožice Škofic na študijski smeri Slovenistika Fakultete za humanistične študije Univerze na Primorskem ter pod raziskovalnim mentorstvom dr. Vesne Mikolič na Inštitutu za jezikoslovne študije ZRS Koper.

- Test 3 in 4: korpus 2 - enaka dela in avtorji; vsak avtor predstavlja samostojno enoto (vsa njegova dela so združena v eno txt datoteko)
- Test 5: korpus 3 - enaka dela in avtorji; vsak avtor predstavlja samostojno enoto (združena dela) + korpus anonimnih pridig
- Test 6: korpus 4 - enak kot v testu 5, a s posodobljeno leksiko

Jezikovno posodobitev (normalizacijo) korpusa z orodjem cSMTiser za avtomatsko posodabljanje besedil, prilagojenim za starejši slovenski jezik, je opravil dr. Tomaž Erjavec z Instituta »Jožef Stefan« v Ljubljani.

Glavno omejitev raziskave je predstavljal omejen izbor besedil (optimalen rezultat bi podala analiza *vseh* besedil *vseh* mogočih avtorjev) in postopek procesiranja gradiva za računalniško analizo, ki ne poteka avtomatizirano (preverjanje pravilnosti zapisa besedil v postopku digitalizacije in kritični prepis potekata ročno) in je zato podvržen napakam. V izogib enostranskim analizam, ki bi podpirale uvodno hipotezo o Repičevem avtorstvu, je bilo opravljenih več testov z različnimi metodami in ob različnih spremenljivkah in parametrih. Stilometrične analize so pokazale največjo algoritemsko izmerjeno stilistično bližino (podobnost) med pridigami in Repičevimi izvirnimi teksti.

Raziskava je potekala v okviru raziskovalnega programa »Razsežnosti slovenstva med lokalnim in globalnim v začetku tretjega tisočletja« (ARIS P5-0409).

Opis raziskovalnih podatkov:

Podatki za raziskavo so bili pridobljeni iz dveh korpusov na repozitoriju CLARIN.SI:

- Košir, Diana and Erjavec, Tomaž, 2024, *Corpus of texts by Hijacint Repič in "Cvetje z vertov sv. Frančiška" CVET 1.0*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1226>,
- Žejn, Andrejka and Erjavec, Tomaž, 2021, *The corpus of older Slovenian narrative prose PriLit 1.0*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1319>;

na portalu Elektronske znanstvenokritične izdaje slovenskega slovstva:

- Faganel, Jože, Ogrin, Matija in Tomaž Erjavec, 2003, *Anton Martin Slomšek: Tri pridige o jeziku*, Inštitut za slovensko literaturo in literarne vede ZRC SAZU, ISBN 961-6500-64-3, <https://nl.ijs.si/e-zrc/slomsek/>;

na portalu Digitalna knjižnica Slovenije:

- Kaffol, Jakob Filip, 1853: *Domači ogovori po nedelskih evangeljih za verne ljudi na deželi*. Celovec: Društvo sv. Mohorja, <https://www.dlib.si/details/URN:NBN:SI:DOC-CFCQCDLM>;

- Slomšek, Anton Martin, Lendovšek, Mihael (urednik), 1899: *Slomškove pridige osnovane*. Maribor: Katoliško tiskovno društvo,
<https://www.dlib.si/details/URN:NBN:SI:DOC-WLVBCSJ7>;
- Vertovec, Matija, 1850: *Shodni ogovori*. Ljubljana: Jožef Blaznik,
<https://www.dlib.si/details/URN:NBN:SI:DOC-YBEAVJWK>.

Stilometrična analiza je bila izpeljana z računalniškim orodjem Stylo v programskem jeziku R. Teksti, pridobljeni iz zgornjih virov, so bili v ta namen pretvorjeni v besedilne datoteke (txt). Besedilna datoteka z anonimnimi pridigami je nastala po kritičnih prepisih v disertaciji.